Application of Binary Logistic Regression to Identify Determinants of Non-Performing Loans

Putri Nilam Cayo[(1)], Ngudiantoro[(2)], Irmeilyana[(3)]

Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Sriwijaya

Ogan Ilir, 30662, Indonesia

e-mail: nilam@unsri.ac.id[(1)], ngudiantoro@unsri.ac.id[(2)], irmeilyana@unsri.ac.id[(3)]

ABSTRAK

Penelitian ini bertujuan untuk mengidentifikasi faktor-faktor yang memengaruhi terjadinya kredit macet serta mengembangkan model berbasis karakteristik nasabah untuk mengestimasi probabilitas kredit macet pada bank. Berbeda dengan penelitian sebelumnya yang umumnya berfokus pada faktor internal perbankan atau kondisi makroekonomi, penelitian ini menggunakan karakteristik individu nasabah dan informasi kredit yang diperoleh dari data pengajuan kredit. Penelitian dilakukan menggunakan data dari 100 nasabah pada salah satu bank di Kota Palembang. Variabel prediktor yang diteliti meliputi usia, jumlah tanggungan keluarga, total pendapatan keluarga, jenis pekerjaan, tingkat pendidikan, jumlah dana yang dipinjam, jangka waktu pembayaran angsuran, dan jumlah angsuran bulanan. Metode yang digunakan adalah analisis regresi logistik biner dengan prosedur Backward Elimination untuk mengidentifikasi faktor-faktor yang berpengaruh signifikan terhadap kredit macet serta mengestimasi probabilitas terjadinya kredit macet. Hasil penelitian menunjukkan bahwa jenis pekerjaan, jumlah dana yang dipinjam, dan jangka waktu pembayaran angsuran berpengaruh signifikan terhadap terjadinya kredit macet. Model akhir yang diperoleh menghasilkan akurasi klasifikasi sebesar 81% dan Area Under the Receiver Operating Characteristic Curve (AUC) sebesar 0,858, yang menunjukkan kemampuan prediksi yang baik. Model regresi logistik biner yang diperoleh dapat digunakan untuk mengestimasi probabilitas terjadinya kredit macet serta membantu pihak bank dalam mengidentifikasi potensi risiko kredit dan mendukung proses pengambilan keputusan dalam evaluasi kredit.

Kata kunci: Regresi Logistik Biner, Backward Elimination, Kredit Macet, Penilaian Risiko Kredit, Prediksi Kredit Macet.

ABSTRACT

This study aimed to identify the factors associated with non-performing loans (NPLs) in banks and develop a borrower-level model to estimate the probability of loan default. Unlike previous studies that mainly focus on bank-level financial indicators or macroeconomic factors, this study utilizes borrower characteristics and loan information obtained from credit application data at a commercial bank in Palembang, Indonesia. The study used data from 100 borrowers, with predictor variables including age, number of family dependents, total household income, occupation, educational attainment, loan amount, loan term, and monthly installment amount. Binary logistic regression with backward elimination was applied to identify significant predictors of NPLs and to estimate the probability of loan default. The results showed that occupation, loan amount, and loan term significantly influenced the occurrence of non-performing loans. The final model achieved a classification accuracy of 81% and an area under the receiver operating characteristic curve (AUC) of 0.858, indicating good predictive performance. The obtained binary logistic regression model can be used to estimate the probability of non-performing loans and assist banks in identifying potential credit risks during the credit evaluation process.

Keywords: *Binary Logistic Regression, Backward Elimination, Non-Performing Loan, Credit Risk Assessment, Default Prediction*

INTRODUCTION

Credit refers to the provision of goods, services, or funds by a creditor to a borrower based on trust, with an obligation to repay the amount at an agreed maturity date [1]. Banks are among the financial institutions that play a significant role in providing credit services to the public. Various factors motivate individuals to apply for credit with banking institutions. According to Sirait et al. [2], credit decisions are influenced by several factors. First, interest rates affect customers' credit decisions by determining monthly installments and the overall financial burden during the loan period. Second, service quality, including the ease of the credit application process, staff competence, and after-sales support, influences customers' choice of bank for credit applications. Third, the efficiency and speed of the credit approval and disbursement process also play an important role, as lengthy and complicated procedures may reduce customer interest in applying for credit.

Many borrowers, for various reasons, are unable to fulfill their loan repayment obligations. This condition subsequently leads to the emergence of non-performing loans (NPLs) within banks. NPLs have negative impacts at both the micro level, affecting individual banks, and the macro level, affecting the banking system and the national economy. Dalmen [3] concluded that high NPL rates are directly related to a decline in banks' capital adequacy ratios, which can threaten their survival and the stability of the financial system. Furthermore, Sintha et al. [4] stated that NPLs harm both debtors and creditors and can even affect a country's economy.

Various measures are needed to reduce and prevent loan defaults in response to the increasingly critical NPL problem. Information on the factors that influence the emergence of NPLs is essential for policy formulation and decision-making to mitigate credit risk. Therefore, further empirical research is necessary to identify and analyze the factors contributing to NPLs.

Recent empirical studies in Indonesia have employed statistical analysis to identify the determinants of non-performing loans in the banking sector. Laksono and Setyawan [5] examine the determinants of non-performing loans in Indonesian conventional commercial banks using panel data regression. The results indicate that bank-specific factors, including the Capital Adequacy Ratio (CAR), Loan-to-Deposit Ratio (LDR), operational efficiency (BOPO), and bank size, significantly influence the level of non-performing loans. Meanwhile, Fahmi et al. [6] apply multiple linear regression to examine the influence of macroeconomic variables and demonstrate that increases in interest rates significantly contribute to higher NPL ratios in Indonesian banks, whereas inflation shows no significant effect. These findings suggest that credit defaults in Indonesia are strongly influenced by statistically significant internal banking indicators and selected macroeconomic factors. Overall, previous studies have investigated non-performing loans using various statistical approaches, with different sets of predictor variables identified according to the research objectives and data characteristics.

Although previous studies have provided valuable insights into the determinants of non-performing loans, they have generally explained credit risk from the perspective of bank-level financial indicators or macroeconomic conditions. In contrast, this study focuses on borrower-level characteristics to estimate the probability that an individual borrower will experience non-performing loans. This borrower-level approach provides an interpretable statistical model that can support individual credit assessment and lending decisions. Binary logistic regression is an appropriate method for modeling loan default because it estimates the probability of a binary outcome and provides interpretable parameter estimates for credit risk assessment. Furthermore, binary logistic regression has been widely applied in credit risk modeling due to its transparency, ease of interpretation, and practical applicability in banking decision-making [7]. The practical application of this method was also demonstrated by Pramesti et al. [8], who used binary logistic regression for credit scoring to classify borrowers based on their credit risk. Based on these considerations, the present study employs binary logistic regression to identify the significant determinants of non-performing loans and to estimate the probability that a borrower will experience NPLs.

To address this research gap, the present study offers several contributions to the existing literature. Specifically, it utilizes borrower-level data obtained from a commercial bank in Palembang, Indonesia, with predictor variables derived from the bank's credit application requirements. In addition, it evaluates the predictive performance of the binary logistic regression model using accuracy, sensitivity, specificity, and the area under the receiver operating characteristic (ROC) curve. The findings are expected to provide empirical evidence for regional banking institutions and support more effective credit risk assessment and lending decisions.

METHOD

Data and Variables

This study utilized secondary data obtained from a banking institution in Palembang, South Sumatra. The study population consisted of 1.147 customers. The minimum sample size was determined using the Isaac and Michael sampling formula [9]. Based on the sample size calculation, a minimum of 50 observations was required. Since this study compared customers with non-performing and performing loans, 50 observations from each group were selected, resulting in a total sample of 100 customer records. Before analysis, the dataset was checked for missing values. No missing values were identified; therefore, all observations were included in the analysis.

The response variable was credit status, while the predictor variables included age (X_1), number of family dependents (X_2), total household income (X_3), occupation (X_4), educational attainment (X_5), loan amount (X_6), loan term (X_7), and monthly installment amount (X_8). These predictor variables were selected because they represent customer demographics and loan characteristics collected during the bank's credit application process and have been reported in previous studies as potential determinants of non-performing loans.

Before data processing and analysis, continuous predictor variables, namely age (X_1), total household income (X_3), loan amount (X_6), and monthly installment amount (X_8), were categorized using quartile-based cut-off points. Meanwhile, the loan term variable (X_7) was categorized according to the bank's existing classification system, while the remaining variables were already recorded as categorical variables. Although categorizing continuous variables may reduce statistical information, quartile-based cut-off points provide data-driven thresholds for categorizing continuous variables [10]. Because several continuous variables in this study were not evenly distributed, quartile-based cut-off points were used to obtain more balanced categories. In addition, some adjacent categories were combined to ensure sufficient observations in each category for binary logistic regression analysis. The final categorization of all variables is presented in Table 1.

Table 1. Categorization of Variables

Variable	Notation	Measurement Scale	Categorization
Credit Status	Y	Nominal	0: Performing Loan 1: Non-Performing Loan
Age	X_1	Ordinal	1: 21 – 31 years old 2: 32 – 42 years old 3: 43 – 53 years old 4: 54 – 64 years old
Number of Family Dependents	X_2	Nominal	1: ≤ 2 dependents 2: 3 dependents 3: 4 dependents 4: ≥ 5 dependents
Total Household Income	X_3	Ordinal	1: $1 \leq X_3 < 9$ million 2: ≥ 9 million
Occupation	X_4	Nominal	1: Civil Servant/State-Owned Enterprise Employee 2: Others (Private Employee, Entrepreneur, or Trader)
Educational Attainment	X_5	Ordinal	1: Junior/Senior High School 2: Undergraduate
Loan Amount	X_6	Ordinal	1: $20 \leq X_6 < 203$ million 2: ≥ 203 million
Loan Term	X_7	Ordinal	1: 1 – 5 years 2: 6 – 10 years 3: 11 – 15 years
Monthly Installment Amount	X_8	Ordinal	1: $0.4 \leq X_8 < 4.8$ million 2: ≥ 4.8 million

Data Analysis Procedure

Data were analyzed using binary logistic regression. The analysis consisted of model formation, simultaneous significance testing, partial significance testing, and best model selection. Before model fitting, multicollinearity among the predictor variables was assessed to ensure that none was substantial.

Multicollinearity Diagnostics

Before fitting the binary logistic regression model, multicollinearity among the predictor variables was assessed using the Generalized Variance Inflation Factor (GVIF), which extends the conventional Variance Inflation Factor (VIF) to accommodate categorical predictors with multiple degrees of freedom [11]. Since several predictor variables were categorical with multiple levels, the adjusted GVIF $\left(GVIF^{1/(2Df)}\right)$ was used to evaluate multicollinearity. Adjusted GVIF values below 2 indicate that multicollinearity is not a concern [12]. Recent studies continue to recommend GVIF for diagnosing multicollinearity in regression models involving categorical predictors [13].

Binary Logistic Regression

Binary logistic regression is a statistical model used to analyze the relationship between one or more factors and a dichotomous (binary) variable [14]. In binary logistic regression, the response variable has two categories: $Y = 1$ indicates the "success" condition, and $Y = 0$ indicates the "failure" condition. In general, the logistic regression probability model involving several predictor variables can be formulated as follows:

$$\pi(x) = \frac{exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)}{1 + exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)} \quad (1)$$

In this case:

$\pi(x)$: probability of the occurrence of the event
β_0 : intercept parameter
$\beta_1, \beta_2, \ldots, \beta_k$: regression coefficients for each predictor variable
$x_1, x_2, \ldots, x_k$: predictor variables
k : number of predictor variables

Simultaneous Significance Test

The simultaneous significance test is conducted to determine whether the predictor variables jointly affect the response variable [15]. The simultaneous testing of β_i parameters is evaluated by using the Likelihood Ratio Test with the following hypotheses:

H_0 : $\beta_1 = \beta_2 = \cdots = \beta_k = 0$
(means there is no simultaneous influence of the predictor variables on the response variable)
H_1 : There is at least one $\beta_i \neq 0$; $i = 1,2,\ldots,k$
(means there are predictor variables that influence the response variable)

The test statistic used is the G-test, defined as follows:

$$G = -2\ ln\left(\frac{L_0}{L_k}\right) \quad (2)$$

where:

G : likelihood ratio test statistic
L_0 : likelihood value under the null model
L_1 : likelihood value under the fitted model

The null hypothesis is rejected if $G > \chi^2_{(\alpha,k)}$.

Partial Significance Test

The partial significance test is conducted to determine which predictor variable significantly affects the response variable [16]. The partial testing of β_i parameters is evaluated by using the Wald Test with the following hypotheses:

H_0 : $\beta_i = 0$; $i = 1,2,\ldots,k$
(means the i-th predictor variable does not significantly influence the response variable)
H_1 : $\beta_i \neq 0$; $i = 1,2,\ldots,k$
(means the i-th predictor variable significantly influences the response variable)

The test statistic is defined as follows:

$$W_i = \left(\frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}\right)^2 \quad (3)$$

where:

W_i : Wald test statistic
$\hat{\beta}_i$: estimated regression coefficient of the i-th predictor variable
$SE(\hat{\beta}_i)$: standard error of the estimated coefficient

The null hypothesis is rejected if $W_i > \chi^2_{(\alpha,1)}$.

Best Model Selection

According to Bursac et al. [17], three well-documented variable selection procedures for binary logistic regression are Forward Selection, Backward Elimination, and Stepwise Selection. These procedures are used to identify relevant predictor variables and obtain a simpler and more interpretable regression model. Since the objective of this study was to identify significant determinants of non-performing loans rather than to compare alternative variable selection techniques, the variable selection procedure was based on these established binary logistic regression procedures. In this study, the Backward Elimination procedure was selected because all predictor variables were derived from the bank's credit application requirements and were therefore considered theoretically relevant. Consequently, all predictor variables were initially included in the model. Subsequently, the predictor variable with the highest p-value was removed from the model. This process was repeated until all remaining predictor variables were statistically significant at the 5% significance level [18].

Model Performance Evaluation

After obtaining the final binary logistic regression model via backward elimination, its predictive performance was evaluated. Model performance was evaluated using a confusion matrix based on accuracy, sensitivity, and specificity, which are commonly used to assess the classification performance of binary prediction models [19]. In addition, the Receiver Operating Characteristic (ROC) curve and the Area Under the ROC Curve (AUC) were used to evaluate the model's discriminative ability [20]. Higher AUC values indicate better discriminative performance and a greater ability of the model to distinguish between performing and non-performing loans.

RESULT AND DISCUSSION

The analysis began with a descriptive overview of the research data, followed by diagnostics for multicollinearity, binary logistic regression modeling, and interpretation of the results.

Description of Data

The data used in this study were from the customers of one of the banking institutions in Palembang. The percentage distribution of predictor variable categories for customers with performing and non-performing credit status is presented in Table 2.

Table 2. Percentage Distribution of Predictor Variable Categories by Credit Status

Predictor Variable	Categories with the Highest Frequency	Frequency Percentage	
		Performing Loan	Non-Performing Loan
Age	32 – 42 years old	55.36%	44.64%
Number of Family Dependents	3 dependents	48.89%	51.11%
Total Household Income	$1 \leq X_3 < 9$ million	47.76%	52.24%
Occupation	Civil Servant	70.83%	29.17%
Educational Attainment	Undergraduate	52.18%	47.82%
Loan Amount	$20 \leq X_6 < 203$ million	40.28%	59.72%
Loan Term	6 – 10 years	51.79%	48.21%
Monthly Installment Amount	$0.4 \leq X_8 < 4.8$ million	47.67%	52.33%

Based on Table 2, customers aged 32–42 years, with three family dependents, and earning household incomes between 1 and 9 million rupiah represented the dominant categories in the dataset. In terms of occupation, the highest proportions were observed among civil servants. Most customers also had an undergraduate degree, borrowed between 20 and 203 million rupiah, selected loan terms of 6–10 years, and paid monthly installments ranging from 0.4 to 4.8 million rupiah. Overall, customers in these dominant categories tended to have a good credit status.

Multicollinearity Assessment Using GVIF

Before fitting the binary logistic regression model, multicollinearity among the predictor variables was examined using the adjusted Generalized Variance Inflation Factor $\left(GVIF^{1/(2Df)}\right)$. The results are presented in Table 3.

Table 3. Generalized Variance Inflation Factor (GVIF) Values for Predictor Variables

Predictor Variable	GVIF	Df	$GVIF^{1/(2Df)}$
Age (X_1)	1.871	3	1.110
Number of Family Dependents (X_2)	1.924	3	1.115
Total Household Income (X_3)	2.262	1	1.504
Occupation (X_4)	1.197	1	1.094
Educational Attainment (X_5)	1.184	1	1.088
Loan Amount (X_6)	2.279	1	1.510
Loan Term (X_7)	1.551	2	1.116
Monthly Installment Amount (X_8)	1.867	1	1.366

All predictor variables had adjusted GVIF values ranging from 1.088 to 1.510, which were well below the recommended threshold of 2. These results indicate that no multicollinearity was detected among the predictor variables. Therefore, all predictor variables were retained for the subsequent binary logistic regression analysis.

Binary Logistic Regression Model

Binary logistic regression was conducted to determine which predictor variables affect credit status. The best model was obtained at the sixth iteration using the Backward Elimination method.

From the model selection, variables X_4, X_6, and X_7 were found to be significant. The final model obtained is presented in Table 4.

Table 4. Estimated Parameters of The Final Model

Variable	$\hat{\beta}$	Significance	$Exp(\hat{\beta})$
X_4	1.202	0.020	3.327
X_6	-1.938	0.001	0.144
X_7	-1.887	0.000	0.152
Constant	4.597	0.002	99.188

Based on Table 4, the best binary logistic regression model obtained for estimating the probability of non-performing credit is given by:

$$\pi(x) = \frac{exp(4.597 + 1.202X_4 - 1.938X_6 - 1.887X_7)}{1 + exp(4.597 + 1.202X_4 - 1.938X_6 - 1.887X_7)} \quad (4)$$

The obtained model was further evaluated to determine its statistical significance. The evaluation was conducted through simultaneous and partial significance tests. The simultaneous significance test was evaluated using Equation (2). The results of the simultaneous significance test are shown in Table 5.

Table 5. Result of The Simultaneous Significance Test

G-test Statistic	Significance	$\chi^2_{(0.05;3)}$
38.605	0,000	7.815

Table 5 shows that the G value exceeds the critical chi-square value (38.605 > 7.815), and the p-value is less than the significance level (0.000 < 0.05). Therefore, the null hypothesis is rejected. This indicates that at the 95% confidence level, at least one predictor variable significantly affects the response variable. To identify the predictor variables that significantly affect the response variable, the analysis was continued using the partial significance test. The partial significance test was evaluated using Equation (3), and the results are shown in Table 4. Based on Table 4, all significant predictor variables have *p-values* smaller than the significance level. These results are consistent with the simultaneous significance test.

Based on the selected best model, some variables had both positive and negative coefficients. Those coefficients were named as parameter values of the variable. A positive parameter value indicates that the predictor variable increases the likelihood of the event, whereas a negative parameter value indicates that it decreases the likelihood [21]. According to Table 4, the parameter value for variable X_4 was positive $(\hat{\beta} = 1.202)$, indicating a positive effect on the probability of non-performing loans. In contrast, the parameter values for variables X_6 and X_7 were negative ($\hat{\beta} = -1.938$ and $\hat{\beta} = -1.887$, respectively), indicating a negative effect on the probability of non-performing loans. The variable X_4 in this model was nominal, with 2 categories: Civil Servant/State-Owned Enterprise Employee (category 1) and Others (category 2). Thus, the higher the bank's occupation category, the greater the likelihood of a non-performing loan. Meanwhile, for variables X_6 and X_7, the lower the loan amount and the shorter the loan term, the greater the chance of a non-performing loan at the bank.

These findings can be explained from several perspectives. First, occupation significantly influences the probability of non-performing loans because employment type reflects the stability of borrowers' income. Borrowers employed as civil servants or employees of state-owned enterprises generally receive regular and predictable monthly salaries, whereas borrowers in the private sector or running their own businesses often experience greater income fluctuations. Such income uncertainty may reduce borrowers' repayment capacity, thereby increasing the likelihood of loan default. Similarly, Syaleh [22] identified occupation as a significant determinant of non-performing loans, supporting the present finding that employment characteristics play an important role in borrowers' credit performance. Therefore, occupation should be considered an important factor in credit assessment, as it reflects a borrower's financial stability and repayment capacity.

The loan amount was also found to significantly influence the probability of non-performing loans. Borrowers who received loans of less than IDR 203 million were more likely to experience non-performing loans than those who received larger loans. A possible explanation for this result is that the approved loan amount may indirectly reflect the bank's assessment of a borrower's overall creditworthiness. Borrowers who qualify for larger loans are generally considered to have lower credit risk because they satisfy the bank's lending requirements and are assessed as having adequate repayment capacity. Furthermore, larger loans typically undergo more comprehensive credit assessments, including evaluations of borrowers' financial conditions and collateral before loan approval, further reducing the likelihood that high-risk borrowers receive large loans. This interpretation is consistent with the findings of Nier et al. [23], who reported that borrowers with larger loans were less likely to default, suggesting that loan size may also serve as an indicator of borrowers' overall creditworthiness. Therefore, loan amount should be considered an important factor in credit assessment because it reflects borrowers' financial capacity and overall creditworthiness.

Another significant predictor identified in this study was loan term. Borrowers with shorter loan terms were more likely to experience non-performing loans than those with longer loan terms. This finding indicates that the repayment period may influence borrowers' ability to manage their financial obligations. Shorter loan terms provide less flexibility for borrowers to adjust their household cash flow because loan obligations must be fulfilled within a relatively limited period. Consequently, borrowers may have greater difficulty maintaining repayment when facing temporary financial pressures or unexpected changes in income conditions. In contrast, longer loan terms allow repayment obligations to be distributed over a longer period, providing borrowers with greater flexibility in managing their financial resources and maintaining regular payments. This result is consistent with the findings of Amaral and Iquiapaza [24], who identified loan term as one of the significant loan characteristics affecting non-performing loans in a development bank portfolio. Therefore, loan term should be considered an important factor in credit assessment because it reflects the flexibility available to borrowers in managing repayment obligations and sustaining loan performance.

From a practical perspective, these findings suggest that banks should not rely solely on the size of the approved loan when assessing credit risk. Borrowers requesting relatively small loans and shorter repayment periods may also require careful evaluation, as these characteristics were associated with higher probabilities of non-performing loans in this study. Incorporating occupation together with loan amount and loan term into credit assessment may improve borrower risk profiling and support more effective lending decisions while reducing the potential for non-performing loans.

Predictive Performance of The Final Model

After obtaining the final binary logistic regression model, its predictive performance was evaluated using a confusion matrix and the receiver operating characteristic (ROC) curve. The predictive performance measures are summarized in Table 6, while the corresponding ROC curve is presented in Figure 1.

Table 6. Predictive Performance of The Final Model

Performance Measure	Value
Accuracy	81.0%
Sensitivity	74.0%
Specificity	88.0%
Area Under the ROC Curve (AUC)	0.858

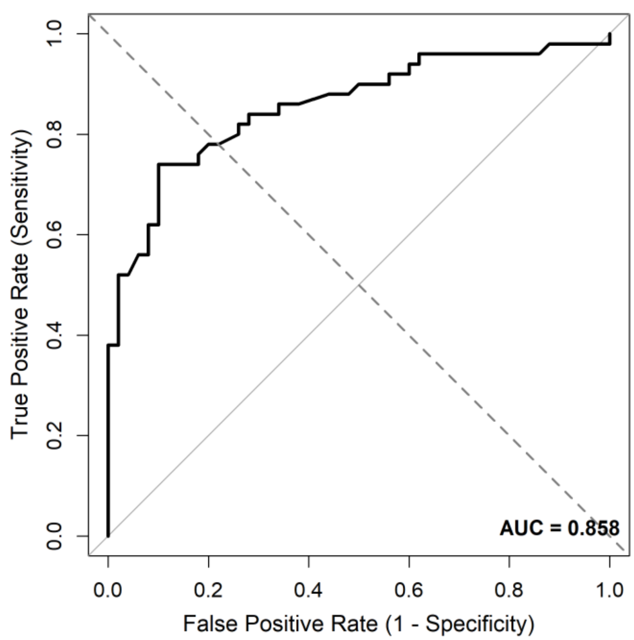

Figure 1. ROC Curve of The Final Model

As shown in Table 6, the model achieved an accuracy of 81.0%, indicating that it correctly classified most observations. The sensitivity and specificity were 74.0% and 88.0%, respectively, suggesting better performance in identifying performing loans than non-performing loans. Furthermore, the model achieved an AUC of 0.858, indicating good discriminative ability in distinguishing between performing and non-performing loans. As illustrated in Figure 1, the ROC curve lies well above the diagonal reference line, further confirming the model's good classification performance. Overall, these results demonstrate that the proposed model has satisfactory predictive performance for estimating loan default probability.

Illustrative Prediction Example

To illustrate the model's application, the probability of a non-performing loan at the bank can be calculated. For example, a customer who works as a non-civil servant/state-owned enterprise employee (category 2) applies for a loan amount ranging from 20 to 203 million rupiah (category 1) and pays the installment over 1–5 years (category 1). Based on those characteristics, the estimated probability is obtained as follows:

$$\pi(x) = \frac{exp(4{,}597 + 1{,}202(2) - 1{,}938(1) - 1{,}887(1))}{1 + exp(4{,}597 + 1{,}202(2) - 1{,}938(1) - 1{,}887(1))}$$

$$\pi(x) = \frac{exp(3{,}176)}{1 + exp(3{,}176)}$$

$$\pi(x) = 0.96$$

The value above indicates that a customer who works as a non-civil servant/state-owned enterprise employee, applies for a loan amount ranging from 20 to 203 million rupiah, and pays the installment over 1–5 years, has a probability of approximately 0.96 (96%) of experiencing non-performing credit. Meanwhile, the probability of having a performing credit status is approximately 0.04 (4%).

CONCLUSION

This study aimed to identify the factors affecting non-performing loans using binary logistic regression. The results showed that occupation, loan amount, and loan term significantly influenced the probability of non-performing loans. The final model achieved satisfactory predictive performance, with an accuracy of 81.0% and an AUC of 0.858. The findings suggest that banks should consider occupation, loan amount, and loan term as important factors in credit risk assessment. Incorporating these variables into credit evaluation procedures may support more effective borrower screening by helping identify borrowers with higher default probabilities, and assist banks in developing more targeted credit risk management strategies to reduce the incidence of non-performing loans.

This study has several limitations. First, the analysis was conducted using data from a single bank, which may limit the generalizability of the findings. In addition, although the final binary logistic regression model demonstrated satisfactory predictive performance, it has not yet been evaluated on data from other financial institutions. Therefore, future studies are encouraged to use larger, more diverse datasets from multiple financial institutions and to compare the binary logistic regression model with machine learning approaches to further evaluate its predictive performance.

ACKNOWLEDGMENTS

The author would like to thank the banking institution in Palembang, South Sumatra, for providing the data used in this study. Appreciation is also extended to colleagues and peers who assisted and engaged in constructive discussions throughout the research process.

REFERENCE

[1] R. Saraswati and J. Jaurino, "Peran Analis Kredit dalam Pemberian Kredit pada PT. BPR Cahaya Khatulistiwa Sejahtera Peran Analis Kredit dalam Pemberian Kredit pada PT. BPR Cahaya Khatulistiwa Sejahtera," *J. Akuntansi, Audit. dan Investasi*, vol. 4, no. 1, pp. 11–16, 2024, doi: 10.54035/jaadi.v4i1.457.

[2] E. T. I. S. Sirait, R. C. Tumanggor, W. V. A. Samosir, J. P. Daeli, H. Siallagan, and C. Hasugian, "Faktor-Faktor yang Mempengaruhi Keputusan Nasabah untuk Mengambil Kredit di Bank : Studi Pustaka," *Innov. J. Soc. Sci. Res.*, vol. 5, no. 1, pp. 1992–2002, 2025, doi: doi.org/10.31004/innovative.v5i1.17822.

[3] V. Dalmen, "Pengaruh Kredit Macet terhadap Kestabilan Keuangan Perbankan di Indonesia," *J. Penelit. Multidisiplin Nusant.*, vol. 6, no. 1, pp. 1–13, 2025.

[4] L. Sintha, F. Tobing, E. Siregar, and R. Juniasti, "Solusi Penanganan Kredit Bermasalah Dampak Pandemi Covid 19 pada Perbankan," *J. Ekobis Ekon. Bisnis Manaj.*, vol. 13, no. 1, pp. 15–24, 2023, doi: doi.org/10.37932/j.e.v13i1.646.

[5] J. D. Laksono and I. R. Setyawan, "Faktor Penentu Non-Performing Loan pada Bank Umum Konvensional di Indonesia," *J. Manajerial dan Kewirausahaan*, vol. I, no. 3, pp. 506–514, 2019, doi: doi.org/10.24912/jmk.v1i3.5362.

[6] H. F. Fahmi, N. R. Habibbah, G. P. Salsabil, D. Deviyanti, S. Nabila, and S. S. Maesaroh, "Analisis Pengaruh Suku Bunga dan Inflasi Terhadap Tingkat Kredit Macet di Perbankan Indonesia," *J. Ekon. Manajemen, Akunt. dan Keuang.*, vol. 6, no. 3, pp. 1–11, 2025, doi: doi.org/10.53697/emak.v6i3.2532.

[7] I. N. Barasa, S. W. Wanyonyi, and M. M. Kololi, "Application of Logistic Regression in Enhancing Digital Credit Risk Management in Commercial Banks," *Asian J. Probab. Stat.*, vol. 27, no. 2, pp. 13–26, 2025, doi: doi.org/10.9734/ajpas/2025/v27i2710 Open.

[8] S. H. Pramesti, Indahwati, and U. D. Syafitri, "Analisis Regresi Logistik dan Cart untuk Credit Scoring dengan Penanganan Kelas Tak Seimbang," *Xplore J. Stat.*, vol. 11, no. 3, pp. 226–237, 2022, doi: doi.org/10.29244/xplore.v11i3.1015.

[9] I. Dewi, H. Khotimah, R. Nasution, Zulpan, T. Hariyati, and A. Rusli, "Teori Populasi dan Pengambilan Sampel," *J. Literasiologi*, vol. 14, no. 2, pp. 1–9, 2023, doi: doi.org/10.47783/literasiologi.v9i4.

[10] D. G. Altman and J. M. Bland, "Quartiles, Quintiles, Centiles, and Other Quantiles," *BMJ*, vol. 309, 1994, doi: doi.org/10.1136/bmj.309.6960.996.

[11] J. Fox and G. Monette, "Generalized Collinearity Diagnostics," *J. Am. Stat. Assoc.*, vol. 87, no. 417, pp. 178–183, 1992, doi: doi.org/10.1080/01621459.1992.10475190.

[12] J. K. Harris, "Primer on binary logistic regression," *Fam. Med. Community Heal.*, vol. 9, no. 1, pp. 1–7, 2021, doi: doi.org/10.1136/fmch-2021-001290.

[13] M. P. Gaonkar, A. N. Liwo, and T. M. Beasley, "Computing Generalized Collinearity Diagnostics for Categorical Variables Using Multivariate Regression," *Gen. Linear Model J.*, vol. 47, no. 2, pp. 16–28, 2023, doi: doi.org/10.31523/glmj.047002.002.

[14] L. P. S. Pratiwi and I. M. P. P. Wijaya, "Binary Logistic Regression Analysis on the Spread of Dengue Fever in Bali Province," *J Stat.*, vol. 17, no. 1, pp. 614–623, 2024, doi: doi.org/10.36456/jstat.vol17.no1.a8955.

[15] R. Crisdianto, M. F. A. Rakafalih, A. Saputri, L. Sopiansyah, A. Fairuzindah, and D. S. Rini, "Analysis of the Timeliness of Graduation of FMIPA College KIP Students at Bengkulu University Using Binary Logistic Regression," *J Stat.*, vol. 16, no. 2, pp. 574–584, 2023, doi: doi.org/10.36456/jstat.vol16.no2.a8233.

[16] S. A. R. Manaf, Erfiani, Indahwati, A. Fitrianto, and R. Amelia, "Faktor – Faktor yang Memengaruhi Permasalahan Stunting di Jawa Barat Menggunakan Regresi Logistik Biner,"

J Stat., vol. 15, no. 2, pp. 265–274, 2022, doi: doi.org/10.36456/jstat.vol15.no2.a5654.

[17] Z. Bursac, C. H. Gauss, D. K. Williams, and D. W. Hosmer, "Purposeful Selection of Variables in Logistic Regression," *Source Code Biol. Med.*, vol. 3, no. 17, pp. 1–8, 2008, doi: doi.org/10.1186/1751-0473-3-17.

[18] M. A. Aprihartha, S. P. Azzahro, and R. Aziza, "Pemilihan Model Regresi Linear Berganda Terbaik untuk Menentukan Faktor-Faktor Penyebab Kasus Balita Gizi Buruk di Jawa Tengah," *J. EurekaMatika*, vol. 13, no. 1, pp. 35–46, 2025.

[19] M. R. F. Nur and S. I. Oktora, "Analisis Kurva ROC pada Model Logit dalam Pemodelan Determinan Lansia Bekerja di Kawasan Timur Indonesia," *Indones. J. Stat. Its Appl.*, vol. 4, no. 1, pp. 116–135, 2020, doi: doi.org/10.29244/ijsa.v4i1.524.

[20] B. Kochański, "The Shape of an ROC Curve in The Evaluation of Credit Scoring Models," *Stat. Transit.*, vol. 25, no. 2, pp. 205–218, 2024, doi: doi.org/10.59170/stattrans-2024-022.

[21] O. Achilonu, N. O. Echendu, and G. Chidumwa, "Making Sense of Regression Models in Clinical Research: A Guide to Interpreting Beta Coefficients and Odds Ratios," *Wits J. Clin. Med.*, vol. 7, no. 2, pp. 99–103, 2025, doi: doi.org/10.18772/26180197.2025.v7n2a12.

[22] H. Syaleh, "Analisis Faktor-Faktor yang Mempengaruhi Kredit Macet pada PT. BPR Dharma Pejuang Empatlima di Kabupaten Lima Puluh Kota," *J. Econ. Bus. Account.*, vol. 1, no. 2, pp. 153–166, 2018, doi: doi.org/10.31539/costing.v1i2.128.

[23] E. Nier, R. Popa, M. Shamloo, and L. Voinea, "Debt Service and Default : Calibrating Macroprudential Policy Using Micro Data," *IMF Work. Pap.*, vol. 19, no. 182, pp. 1–45, 2019, doi: doi.org/10.5089/9781513509099.001.

[24] G. H. D. O. Amaral and R. A. Iquiapaza, "Determinants of Non-Performing Loans and Loan Recovery of a Development Bank Portfolio," *BASE – Rev. Adm. e Contab. da Unisinos*, vol. 17, no. 3, pp. 483–519, 2020, doi: doi.org/10.4013/base.2020.173.05.