9710-29016-1-RV_TurnitIn_.docx

by Edi Widodo

Credit Risk Modeling Based on Geographical Location: A Case Study of Savings and Loan Cooperatives

Edi Widodo[1*], Ahmad Rifa'i[2], Pratama Angga Buana[3]

ABSTRACT

The aim of this study is to examine how geographical location affects the credit risk faced by savings and loan cooperatives. Using a quantitative approach, this research will develop a credit risk model that considers geographical variables, measured by the Human Development Index (HDI). The initial stage of the research involves classifying the credit dataset according to the categories determined by Bank Indonesia. The data cleansing process resulted in attributes such as credit ceiling, HDI, and credit category. Analysis was conducted using Chi-Square, and Logistic Regression methods. The Chi-Square analysis results showed a statistically significant relationship between credit ceiling, HDI, and credit category (p-value < 0.05). The Logistic Regression models demonstrated high accuracy in classifying the data, with Logistic Regression achieving 89.71%. In conclusion, credit ceiling and HDI have a significant influence on credit category, with the Logistic Regression model data classification. This study provides valuable insights into how credit ceiling and HDI influence credit categories, which can be used to make better decisions related to public policy, development planning, and social interventions

1. INTRODUCTION

Every entity involved in financial activities, especially savings and loan cooperatives (KSP), faces significant challenges related to non-performing loans (NPL). Delays in payment by KSP can lead to a decline in trust from the banking sector, as these cooperatives use bank credit facilities to grow their business[1]. KSPs play a central role in supporting economic development and community welfare, making credit management increasingly important [2].

Various methods for internal control to minimize NPL risk, such as the 5C Credit analysis [3][4][5], Profile Matching Analysis [6], and the Elimination and Choice Translation Reality method [7], have been implemented. However, according to data from the Financial Services Authority (OJK), the NPL rate in the People's Credit Bank (BPR) sector as of December 1, 2023, was 9.7%, amounting to 13.89 trillion Rupiah out of 194.98 trillion Rupiah in total loan disbursements[8].

The 5C and 7P principles remain crucial in the loan granting process, assessing prospective borrowers' feasibility[3][9]. Character, influenced by local culture, is one significant aspect [10].

Research by Djuarni and Ratnasari revealed that financial institutions did not properly implement the 5C analysis for MSMEs. Rahmasari's research focused on the effectiveness of the 5C and 7P principles from an operational management perspective using the McKinsey 7S Framework Model [9]. The strategic management process for lending begins with environmental analysis to determine the target market and assess credit risk, reflected in the NPL ratio, and ends with evaluating the 5C and 7P principles at PT Bank Mandatory's Directorate of Commercial Banking.

Lilis Sariani's research at BMT As'adiyah concluded that the 5C principles were well implemented, yet credit defaults still occurred, indicating the importance and limitations of these principles (Sariani, 2021). Research by Agus Sasmita and Nur Heri Cahyana used Association Rule Mining to identify borrower characteristics and predict creditworthiness. Their findings showed the Tree Weighted Itemset (WIT) method effectively predicted loan outcomes [1].

Prevention efforts, careful monitoring, and credit restructuring strategies are integral parts of risk management to reduce the negative impact of non-performing loans on financial institutions and the financial system as a whole [11].

In conclusion, the 5C and 7P principles, while standard for assessing creditworthiness, still result in significant non-performing loans, whether poorly or properly applied. This research also examines the impact of geographical location on credit risk in the KSP Sumber Rejeki, addressing how geographical factors influence credit risk and how integrating these elements can improve credit risk management and policy development.

RESEARCH METHODS

1. Data Collections

This research utilizes two data sources, primary and secondary. Primary data consists of creditor profile data obtained from research partners, spanning from 2018 to 2022. Meanwhile, secondary data utilizes the "Semarang dalam Angka" data from 2018 to 2022 published by the Central Statistics Agency of Central Java.

2. Methodology

This study uses a correlational method based on quantitative data processed using Regression Analysis algorithms to identify patterns of similarity[12]. The goal is to predict the repayment ability of prospective borrowers when applying for credit, based on profile similarities to the generated model. In this approach, various borrower characteristics such as income, credit history, and employment status are analyzed to determine common traits among those who successfully repay their loans. By comparing these traits with those of new applicants, the model can estimate the likelihood that a new applicant will also repay their loan, thereby helping financial institutions make more informed lending decisions.

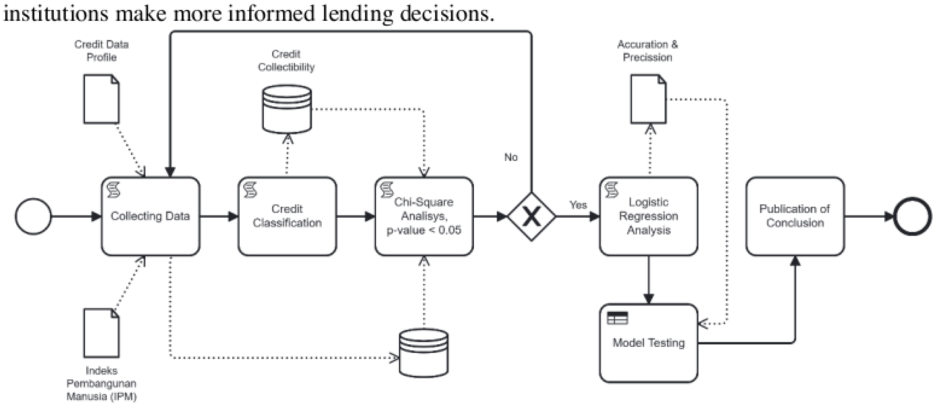

Figure 1 Research Methodology

The diagram illustrates the research methodology, outlining a sequence of steps and processes involved in the study.

RESULTS

In this section, the research stages and the results of each stage will be explained as follows:

– Collecting Data:
  The process starts with collecting data, which includes the Credit Data Profile and the Human Development Index (Indeks Pembangunan Manusia, IPM). We collected a total of 4,626 profile data records obtained from our partner's database, spanning the years 2018 to 2022. The Human

Development Index data was obtained from the Central Bureau of Statistics for the years 2018 to 2022. The sample credit data from our research partner, as shown in table 2, has intentionally omitted the name column to maintain the confidentiality of the debtors.

Table 1 Credit Data Table

Date	Name	Address	Ceiling	Day Late
14/06/2019	-	Karangsono 07/01 Mranggen	3,000,000	180
13/03/2018	-	Dukuhan 01/03 Sayung	1,000,000	1
26/08/2022	-	Tlogorejo 4/6 Karangawen	1,500,000	119
13/01/2020	-	Temuroso 03/06 Guntur	3,000,000	1
19/03/2022	-	Karangmalang 05/03 Gemulak	6,500,000	75
...	...	...	...	...

From the credit data in Table 1, we performed a cleaning process by separating the address column into village, sub-district, and district columns, and separating the year from the date column. This was done to link with the Human Development Index data in Table 3.

Table 2 Data Preparation

Year	Village	Sub-district	District	Ceiling	Day Late
2019	Karangsono	Mranggen	Demak	3,000,000	180
2018	Dukuhan	Sayung	Demak	1,000,000	1
2022	Sukodono	Bonang	Demak	1,500,000	119
2020	Temuroso	Guntur	Demak	3,000,000	1
2022	Gemulak	Sayung	Demak	6,500,000	5
	...	...	...	...	...

Table 3 Human Development Index (Indeks Pembangunan Manusia, IPM)

Subject	2022	2021	2020	2019	2018	2017
Sayung	72.12	72.33	71.98	71.63	71.02	70.17
Karangtengah	72.75	72.96	72.61	72.26	71.64	70.79
Gajah	70.85	71.05	70.71	70.37	69.77	68.94
Dempet	72.33	72.54	72.19	71.84	71.23	70.38
Wonosalam	73.91	74.13	73.77	73.41	72.79	71.92
Demak	74.76	74.97	74.61	74.25	73.62	72.74
Bonang	71.17	71.37	71.03	70.68	70.08	69.25
Wedung	70.43	70.63	70.29	69.95	69.36	68.53
Mijen	72.01	72.22	71.87	71.52	70.92	70.07
Karanganyar	71.38	71.58	71.24	70.89	70.29	69.45
Mranggen	74.12	74.34	73.98	73.62	73.00	72.13
Kebonagung	72.64	72.86	72.50	72.15	71.54	70.69
Guntur	72.22	72.43	72.08	71.73	71.12	70.28

Sources: The Central Statistics Agency

Table 4 Human Development Index Added to The Credit Data

Year	Village	Sub-district	District	Ceiling	Day Late	HDI
2019	Karangsono	Mranggen	Demak	3,000,000	180	73.62

2018	Dukuhan	Sayung	Demak	1,000,000	1	71.02
2022	Sukodono	Bonang	Demak	1,500,000	119	71.17
2020	Temuroso	Guntur	Demak	3,000,000	1	72.08
2022	Gemulak	Sayung	Demak	6,500,000	5	72.12
	...	...	...	...	...	...

– Credit Classification:
The collected data is then classified based on credit characteristics. The collected data was then classified according to credit criteria based on OJK regulations on credit classification, which are based on the total number of overdue loan installments. This step likely involves organizing the data into different categories or groups for further analysis.

Table 5 Credit Category Based on Day Late

Year	Village	Sub-district	District	Ceiling	Day Late	HDI	Category
2019	Karangsono	Mranggen	Demak	3,000,000	180	73.62	K5
2018	Dukuhan	Sayung	Demak	1,000,000	1	71.02	K1
2022	Sukodono	Bonang	Demak	1,500,000	119	71.17	K3
2020	Temuroso	Guntur	Demak	3,000,000	1	72.08	K1
2022	Gemulak	Sayung	Demak	6,500,000	5	72.12	K1
	...	...	...	...	...	...	

– Chi-Square Analysis:
Chi-square, or the chi-squared test, is a statistical test used to analyze categorical data. This test helps us determine whether there is a relationship between categorical variables or if the differences between the observed data and the expected data occur by chance. Chi-Square analysis is performed to determine the statistical significance of the classification. Specifically, the analysis checks if the p-value is less than 0.05. If the p-value is not less than 0.05, it suggests that there is no significant relationship, and the process might need to revisit the classification or collecting data steps [13].

$$x^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{1}$$

where O_i is the observed frequency and E_i is the expected frequency.

Based on the results of the Chi-Square analysis, it can be concluded that:

a. There is a significant relationship between category, ceiling, and HDI. The high Chi-Squared value (1180.65) indicates a substantial difference between the observed and expected data frequencies. The very small p-value (0.000), well below the commonly used threshold (0.05), suggests that these results are unlikely to have occurred by chance. The degrees of freedom (836) indicate the number of cells in the contingency table minus one.
b. This relationship is quite strong. The high Chi-Squared value indicates a significant difference between the category, ceiling, and HDI.

– Logistic Regression Analysis:

If the Chi-Square analysis indicates a significant relationship (p-value < 0.05), the process proceeds to Logistic Regression Analysis. This step involves using regression algorithms to model and predict the repayment ability of prospective borrowers based on their profiles.

$$P(Y = 1 \mid X) = \frac{1}{(1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n)})} \tag{2}$$

where $P(Y = 1 \mid X)$ is the probability of the desired event (Y=1) given the predictor X, β_0 is intercept, and $\beta_1 \beta_2 \ldots \beta_n$ regression coefficient for the independent variable $X_1 X_2 \ldots X_n$.

Based on the tests above, it can be concluded that the model performs quite well in classifying data into three categories (K1, K2, K3).

a. Accuracy: 0.8725490196078431
   The accuracy value indicates the overall percentage of correct model predictions. In this case, 87.25% of the Logistic Regression predictions for the testing data are correct. A high accuracy value shows that the model is generally good at classifying data.

b. Precision: 0.7945787225781401
   Precision measures the proportion of true positive predictions. In this context, a precision value of 0.795 indicates that 79.5% of the data classified as a certain category by the model truly belong to that category. A high precision value shows that the model rarely produces false positive predictions (e.g., classifying K2 data as K3).

c. Recall: 0.8284313725490197
   Recall measures the proportion of actual positive data correctly identified by the model. A recall value of 0.828 indicates that of all the data that actually belong to a certain category (e.g., K1), 82.8% are correctly classified by the model. A high recall value shows that the model can correctly identify most of the positive data.

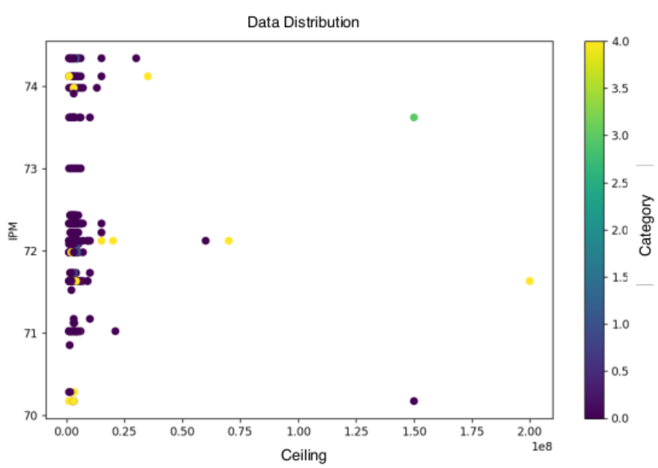

Figure 2 Data Distribution Visualization

– Model Testing:
The logistic regression model is then tested for accuracy and precision. This step ensures that the model's predictions are reliable and valid.

Testing with a ceiling of 1,000,000 and an IPM of 0.7 resulted in K1, which means the category is current, and Testing with a ceiling of 50,000,000 and an IPM of 0.7 resulted in K4, which means the category is non-performing.

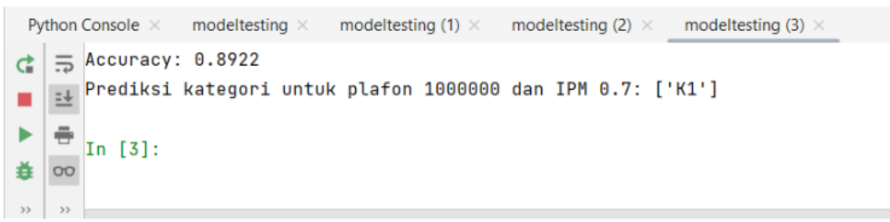

Figure 3 Testing Result for the ceiling = 1.000.000 and IPM = 70

Figure 4 Testing Result for ceiling = 50.000.000 and IPM = 70

DISCUSSION

The findings of this research highlight the significant interplay between ceiling limits, the Human Development Index (IPM), and loan categories, confirming the hypotheses that these variables are interrelated and influential. The high accuracy of Logistic Regression (87.25%) models in classifying loan data underscores the robustness of these methods in predictive analytics. The statistical significance of the relationship (p-value < 0.05) between the ceiling, IPM, and loan categories suggests that these factors are critical in determining loan performance. Notably, the slightly higher accuracy of the Logistic Regression model indicates its superior capability in handling this dataset. This research modifies existing theories by integrating IPM as a crucial variable in loan classification, which is often overlooked in traditional credit risk assessment models. The implications of these results are profound, suggesting that financial institutions can enhance their risk management strategies by incorporating these variables into their predictive models. The findings advocate for a more nuanced approach to credit assessment, potentially leading to more informed decision-making and improved financial stability.

CONCLUSIONS AND RECOMMENDATIONS

Based on the analysis presented earlier, the following conclusions can be drawn from this study:

1. Significant Relationship between Ceiling, IPM, and Category
   There is a statistically significant relationship between the ceiling, IPM, and category (p-value < 0.05). This indicates that these three variables are interrelated and can influence each other.
2. Accurate Classification Models
   Logistic Regression models show high accuracy in classifying data, achieving 89.71%. This demonstrates that models can be used to predict categories quite effectively.
3. Significant Impact of Ceiling and IPM
   Both the ceiling and IPM have a significant impact on the category. This indicates that these two variables can be used as determinants in predicting the category.
4. Logistic Regression is Slightly Better
   Overall, the Logistic Regression model is slightly better at classifying data compared to the Decision Tree. This is evident from its slightly higher accuracy.

From the results of the analysis that has been conducted, a programming language can be further developed to predict credit categories based on the credit application's loan amount (ceiling) and Human Development Index (Indeks Pembangunan Manusia, IPM). This can serve as a guide for granting credit to prospective borrowers.

Logistic Regression models that have been trained can be used to predict new categories based on the ceiling and IPM values. This can be beneficial for various purposes, such as:
- Targeted public policy determination
- Effective development planning
- Focused social interventions

It's equally important to note that predictive models are tools and should be used carefully and responsibly

9710-29016-1-RV_TurnitIn_.docx

ORIGINALITY REPORT

4%	1%	2%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	1%
2	J. N. Reddy. "Innovative Developments in Design and Manufacturing - Advanced Research in Virtual and Rapid Prototyping -- Proceedings of VRP4, Oct. 2009, Leiria, Portugal", CRC Press, 2019 Publication	1%
3	Anton J.J. Van Rompaey, Gerard Govers, Etienne Van Hecke, Kristine Jacobs. "The impacts of land use policy on the soil erosion risk: a case study in central Belgium", Agriculture, Ecosystems & Environment, 2001 Publication	1%
4	Submitted to Old Dominion University Student Paper	<1%
5	core.ac.uk Internet Source	<1%
6	ir-library.egerton.ac.ke Internet Source	<1%

7	Nurun Latifah, Ramaditia Dwiyansaputra, Gibran Satya Nugraha. "Multiclass Text Classification of Indonesian Short Message Service (SMS) Spam using Deep Learning Method and Easy Data Augmentation", MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, 2024 Publication	<1%
8	K. Krishnamoorthy. "Handbook of Statistical Distributions with Applications", Chapman and Hall/CRC, 2019 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		

9710-29016-1-RV_TurnitIn_.docx

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7