PARAMETER
JURNAL MATEMATIKA, STATISTIKA DAN TERAPANNYA

e-ISSN. 2829-6303
Vol. 04 No. 01, April 2025
Page 175-184

COMPARISON OF CLASSIFICATION MODELS USING SUPPORT VECTOR MACHINE (SVM) AND K-NEAREST NEIGHBOR (K-NN) METHODS IN NON-PERFORMING LOAN ANALYSIS CASE STUDY: PT. ADIRA FINANCE AMBON BRANCH

Charolina Putry Djami[1], M.S. Noya Van Delsen[2*], S. J. Latupeirisa[3]

[1 2,3] Department of Statistics, Fakulty of Science and Technology, Universitas Pattimura
Jl. Ir. M. Putuhena, Ambon, 97233, Maluku, Indonesia

E-mail Correspondence Author: *marlonnvd@gmail.com*

Abstract
SVM works by finding the best dividing line (or hyperplane) to separate two groups of data based on the maximum margin. k-NN classifies new data based on its similarity to previous, already-labeled data points. Non-performing loan analysis is a crucial aspect of credit risk assessment. This study compares the performance of the Support Vector Machine (SVM) and k-Nearest Neighbor (k-NN) classification methods in analyzing non-performing loans at PT. Adira Finance Ambon Branch. The dataset includes demographic and financial attributes, processed through normalization, data splitting, and evaluation using accuracy, precision, recall, and AUC metrics. The results show that SVM with a linear kernel performs best, achieving 97.83% accuracy and 95% AUC. Meanwhile, k-NN with k=5 attains 78.26% accuracy. Thus, SVM outperforms k-NN in classifying non-performing loans in this study.

Keywords: *Classification, k-Nearest Neighbor (k-NN), Non-performing Loans, Support Vector Machine (SVM)*

: https://doi.org/10.30598/parameterv4i1pp175-184

Submitted: Februari 2025
Accepted: April 2025

1. INTRODUCTION

Credit can be understood as the provision of funds or similar claims based on an agreement between a creditor and a debtor, where the debtor is required to repay the debt within a specific time period along with interest. This is in line with Indonesia's Banking Law No. 10 of 1998. In practice, lending activities always come with certain risks—one of the most common being credit default, which happens when a borrower can't fulfill their repayment obligations. Non-performing loans can have serious consequences for financial institutions, such as increasing the Non-Performing Loan (NPL) ratio, which in turn could disrupt the company's financial growth. Defaults usually occur when a borrower fails to pay back the loan as agreed, which can lead to legal consequences or financial losses [1]. The causes behind bad loans can come from either side. On the borrower's side, factors like poor financial planning, business failure, or broader economic conditions are often beyond their control. Meanwhile, lenders might also be at fault—especially when they're too lenient in analyzing creditworthiness. To minimize these risks, there needs to be a reliable credit assessment system in place to review a borrower's financial background before any loans are approved. At the moment, many cooperatives and financial institutions still rely on manual evaluations led by branch managers. However, not all of them have the background or experience of a credit analyst, which increases the risk of inaccurate assessments. That's why data-driven methods are becoming increasingly important to support decision-making, especially when it comes to identifying which customers are likely to default and which are not.

In the world of machine learning, there are various classification methods that can help with this [2]. Two of the most commonly used **are** Support Vector Machine (SVM) **and** k-Nearest Neighbor (k-NN) [3]. SVM works by finding the best dividing line (or hyperplane) to separate two groups of data based on the maximum margin [4]. On the other hand, k-NN classifies new data based on its similarity to previous, already-labeled data points [5]. Both methods have been widely applied in different kinds of classification problems, including in credit analysis. Some previous studies have shown SVM to outperform k-NN, while others found the opposite. For instance, in a study on classifying drinking water feasibility, SVM came out on top with an accuracy of 69.76% [6]. But in another study comparing SVM, k-NN, and Naïve Bayes for loan approval classification, k-NN and Naïve Bayes achieved better results (77%) compared to SVM (50%) [7].

This research aims to compare how well SVM and k-NN perform in classifying non-performing loans at PT. Adira Finance Ambon Branch. As one of the key financing companies in the area, Adira Finance also deals with loan defaults. So, having a more accurate way of evaluating borrower risk could be very helpful. The goal of this study is to offer insights into which classification method is most effective for tackling bad credit cases.

2. RESEARCH METHODOLOGY

2.1 Type Of Research

This study is a case study focusing on the issue of non-performing loans at PT. Adira Finance Ambon Branch. The goal is to find out which classification method SVM or k-NN gives the best results based on accuracy.

2.2 Time and Location

The research was conducted over a three-month period at the Department of Mathematics, Faculty of Science and Technology, Pattimura University in Ambon.

2.3 Tools and Materials

The data used in this research include various attributes such as Gender, Age, Marital status, Number of dependents, Housing Status, Occupation, Monthly income, Installment amount, Loan term and Loan application type. In total, there are 1,570 records of secondary data collected from PT. Adira Finance Ambon Branch for the year 2024. Data input and frame creation were done using Microsoft Excel, while the model analysis and evaluation were carried out in R programming language.

Table 1. Research Variables

Variable	Variable Symbol	Factors	Description
Response	y	Credit status	Categoric variable: 0 : Peforming loans 1: Non- Peforming loans
Prediktor	x_1	Gender	Categoric variable: 1 : Male 2 : Female
	x_2	Age (Years)	Numeric variable
	x_3	Marital status	Categoric variable: 1 : Married 2 : Unmarried
	x_4	Number of household dependents	Variable Numerik
	x_5	Housing status	Categoric variable: 1 : Own House 2 : Living with family 3 : Rented
	x_6	Occupation	Categoric variable: 1 : Goverment employee 2 : Private sector 3 : BUMN 4 : Entrepreneur
	x_7	Monthly income (Million Rupiah)	Numeric variable
	x_8	Installment Amount (Million Rupiah)	Numeric variable
	x_9	Loan term (Months)	Numeric variable
	x_{10}	Loan application type	Categoric variable: 0 : Motorcycle 1 : Car

2.4 Procedure

Here’s the step-by-step workflow used in this research:

1. Collect and input data.
2. Clean and transform the data.
3. Split the data into training and testing sets [8].
4. Apply the Support Vector Machine (SVM) classification method [9].
5. Apply the k-Nearest Neighbor (k-NN) classification method [10]
6. Evaluate each model’s performance based on classification metrics.

7. Compare SVM and k-NN based on evaluation results
8. Interpret the best-performing model

3. RESULTS AND DISCUSSION

3.1 Descriptive Statistics of the Variables

The overall distribution of loan status at PT. Adira Finance Ambon Branch is shown in **Figure 1**.

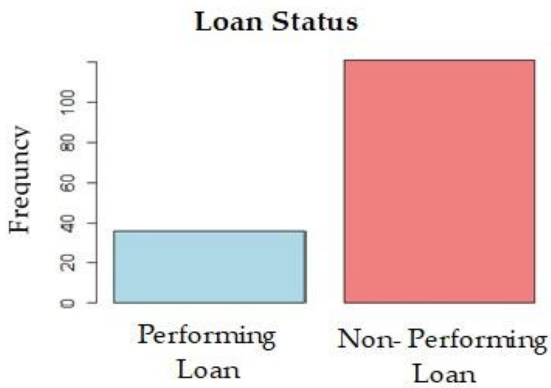

Figure 1. Non-Performing Loan Status

In this research credit status as response variable are categorized into two groups: performing loans (status = 0) and non-performing loans (status = 1). The dataset used in this research includes a total of 157 borrower records in PT Adira Finance Ambon Branch.

3.1.1 Descriptive Statistic of Numerical Predictor Variables

Based on **Table 2**, the descriptive statistics for the numerical predictor variables are as follows: variable X_2 (Age) shows that the youngest debtor is 19 years old, while the oldest is 56 years old. Variable X_4 (Number of Household Dependents) indicates that the highest number of dependents supported by a debtor is six. Variable X_7 (Income) reveals that the highest monthly income is IDR 10,000,000, while the lowest is IDR 1,323,000. Variable X_8 (Installment Amount) shows that the monthly installment ranges from IDR 870,000 to IDR 14,388,000. Lastly, variable X_9 (Installment Duration) shows that the longest installment term is 60 months, and the shortest is 12 months.

Table 1. Descriptive Statistic of Numerical Predictor Variables

Numerical Predictor Variables	Min	Max	Mean
X_2	19.00	56.00	37.18
X_4	0.00	6.00	2.815
X_7	1.32	10.00	5.07
X_8	8.70	14.38	1.70
X_9	12.00	60.00	25.92

3.1.2 Descriptive Statistic of Categorical Predictor Variables

a. Gender

Table 3 displays the distribution of variable X_1 (Gender), which is categorized into male and female. The data show that the majority of debtors, totaling 101 individuals, are male.

Table 2. Gender Variable

Variable	Categori	Number of debtors
X_1	Male	101
	Female	56
Total		157

b. Marital Status

Variable X_3 (Marital Status) is classified into two categories: married and unmarried. The results indicate that the majority of debtors 136 individuals are in the married category.

Table 3. Marital Status Variable

Variable	Category	Number of debtors
X_3	Married	136
	Unmarried	21
Total debtors		157

c. Housing status

Variable X_5 (Residential Status) includes three categories. The highest number of debtors, totaling 83 individuals, reside in their own homes.

Table 4. Housing Status Variable

Variable	Category	Number of debtors
X_5	Own house	83
	Living with family	40
	Rented	34
Total debtors		157

d. Occupation

Variable X_6 (Occupation) consists of four categories. The majority of debtors, totaling 65 individuals, are employed in the private sector.

Table 5. Occupation Varible

Variable	Category	Number of debtors
X_6	Goverment employee	32
	Private sector	65
	BUMN	13
	Entrepreneur	47
Total debtors		157

e. Loan Application Type

Variable X_{10} (Loan Application Type) consists of two categories: motorcycle and car loan applications. The data show that a greater number of debtors applied for motorcycle loans.

Table 6. **Loan Application Type Variable**

Variable	Category	Number of debtors
X_{10}	Motorcycle	140
	Car	17
Total debtors		157

3.2 Support Vektor Machine in Classfying Non-peforming Loans

This reasearch employed three types of kernel functions for the Support Vector Machine (SVM) classification: Linear [11], Polynomial [12], and Radial Basis Function (RBF) [13]. Each kernel was optimized by selecting the most suitable parameters. The optimal values obtained were C = 1 for the linear kernel, p = 3 and C = 1 for the polynomial kernel, and σ = 0.1 with C = 1 for the RBF kernel. After parameter tuning, the classification performance of each kernel was evaluated and compared to determine which kernel delivered the best results [14]. A summary of the evaluation metrics is presented in **Table 8**.

Table 8. **Performance Accuracy of SVM Using Linear, Polynomial, and Radial Kernels**

Kernel type	Accuracy	*Recall*	*Precision*
Kernel Radial	95.65%	80%	100%
Kernel Pollynomial	91.30%	60%	100%
Kernel Linear	**97.83%**	**90%**	**100%**

The accuracy results presented in **Table 8** indicate that the classification using the linear kernel outperformed the others in identifying non-performing loans at PT. Adira Finance Ambon Branch, achieving a total accuracy of 97.83%. In comparison, the polynomial kernel yielded an accuracy of 91.30%, while the radial kernel achieved 95.65%.

3.3 k-Nearest Neighbor in Classifying Non-Performing Loans

k-Nearest Neighbor (k-NN) method was applied to classify non-performing loans at PT. Adira Finance Ambon Branch by testing different values of k (2, 3, 4, and 5) [15], [16]. These values were used to identify the nearest neighbors for building the classification model based on the training data, with performance evaluated using the testing data. A comparison of the evaluation metrics for each classifier configuration is presented in **Table 9** below.

Table 9. **Performance Accuracy of k-NN with k Values of 2, 3, 4, and 5**

Nilai *k*	Accuracy model	*Recall*	*Precision*
2	69.56%	60%	37.5%
3	67.39%	40%	30.77%
4	73.91%	50%	41.66%
5	**78.26%**	**40%**	**50%**

The accuracy results shown in **Table 9** indicate that the classification using k = 5 performed best in identifying non-performing loans at PT. Adira Finance Ambon Branch, achieving a total accuracy of 78.26%. In comparison, the accuracy for k = 2 was 69.56%, for k = 3 was 67.39%, and for k = 4 was 73.91%.

3.4 Evaluating Best Performance Accuracy Between k-NN and SVM Models

This section presents a comparison between the SVM and k-NN methods in classifying non-performing loans at PT. Adira Finance Ambon Branch. The detailed performance metrics for both models are summarized in **Table 10**.

Table 10. Comparison of the Best Performance Accuracy Between k-NN and SVM Models

Model	Accuracy model	*Recall*	*Precision*	AUC
Kernel Linear	**97.83%**	**90%**	**100%**	**95%**
k=5	78.26%	40%	50%	64.44%

The performance comparison between the two classification models k-Nearest Neighbors (k-NN) with k=5 and Support Vector Machine (SVM) with a linear kernel shows that the SVM model achieved a higher accuracy of 97.83%. Its classification performance, as indicated by an Area Under the Curve (AUC) value of 95%, demonstrates excellent capability in identifying non-performing loans at PT. Adira Finance Ambon Branch [17]. In contrast, the k-NN model with k = 5 reached an accuracy of 78.26%. These results suggest that the SVM model provides more accurate overall predictions than k-NN in classifying non-performing loans in this case study.

3.5 Evaluating Model Accuracy Performance Using 80:20 and 90:10 Data Splits

This study also evaluates the accuracy performance of the SVM and k-NN models using 80:20 and 90:10 data splits as comparisons to the 70:30 ratio previously used for classifying non-performing loans at PT. Adira Finance Ambon Branch. The results of this comparison are presented in **Table 11** below:

Table 11. Accuracy Evaluation of SVM and k-NN Models Using 70:30, 80:20, and 90:10 Data Splits

		SVM			**k-NN**			
		Kernel Radial	**Kernel *Polynomial***	**Kernel Linear**	**k = 2**	**k = 3**	**k = 4**	**k = 5**
70: 30	Accuracy	95.65%	91.30%	**97.83%**	69.56%	67.39%	73.91%	78.26%
	Recall	80%	60%	90%	60%	40%	50%	40%
	Precision	100%	100%	100%	37.5%	30.77%	41.66%	50%
80:20	Accuracy	87.10%	90.32%	96.77%	77.42%	77.42%	77.42%	77.42%
	Recall	71.43%	57.14%	85.71%	42.86%	42.86%	28.57%	28.57%
	Precision	91.67%	100%	100%	50%	50%	50%	50%
90:10	Accuracy	86.67%	80%	93.33%	86.67%	73.33%	93.33%	86.67%
	Recall	66.67%	33.33%	100%	66.66%	66.66%	66.66%	33.33%
	Precision	91.67%	91.67%	91.67%	66.66%	40%	100%	100%

Table 11 shows that among the three data split proportions, the best accuracy performance was achieved by the SVM model with a linear kernel using the 70:30 split, reaching an accuracy of 97.83%. In comparison, under the same data split, the best-performing k-NN model with k = 5 achieved an accuracy of 78.26%.

4. CONCLUSION

Based on the results discussion that have been carried out, the conclusions of this study are as follows:

1. Suport Vector Machine SVM was tested using three kernel types: linear, polynomial, and radial basis function (RBF). Among them, the linear kernel yielded the highest accuracy (97.83%), with a recall of 90%, precision of 100%, and an AUC of 95%. These metrics indicate that the linear kernel is highly reliable and consistent in identifying risky loans.
2. k- Nearest Neighbor was evaluated using different values of k (2 to 5). The best result was achieved with k = 5, which gave an accuracy of 78.26%, recall of 40%, and precision of 50%. While it showed decent performance, it was significantly outperformed by the SVM model in all aspects.
3. Additional analysis using different training/testing splits (70:30, 80:20, and 90:10) also the value *Area Under The Curve* (AUC) confirmed that the SVM linear model consistently outperformed k-NN across various data proportions.

REFERENCES

[1] Hariyani, Restrukrisasi dan penghapusan kredit macet, vol. Vol.1, Jakarta, 2010.

[2] F. A. Hermawati, Data Mining, vol. 1, Yogyakarta: Andi, 2013.

[3] S. N. Asiyah and K. Fithriasari, "Klasifikasi Berita Online Menggunakan Metode Support Vector Machine dan K- Nearest Neighbor," *Jurnal Sains Dan Seni ITS,* vol. 5, no. 2, pp. 317-322, 2016.

[4] I. I. Aditia, R. Latuconsina dan A. Dinimaharawati, “Sistem Pendukung Keputusan Pemilihan Tempat Kerja,” dalam *e-Proceeding of Engineering*, 2021.

[5] J. Suntoro, Data Mining : Algoritma dan Implementasi dengan Pemrograman PHP, Jakarta: PT Elex Media Komputindo, 2019.

[6] S. Ulum, R. F. Alifa, P. Rizkika dan C. Rozikin, “Perbandingan Performa Algoritma KNN dan SVM dalam Klasifikasi Kelayakan Air Minum,” *Generation Journal,* vol. 7, no. 2, pp. 141-146, 2023.

[7] Z. Saputra, H. A. Supahri, R. Ismanizan dan Rahmaddeni, “Perbandingan Algoritma KNN (K-Nearest Neighbors), Naïve Bayes, Dan,” *Jurnal Ilmiah Sistem Informasi dan Teknik Informatika (JISTI),* vol. 7, no. 1, pp. 67-75, 2024.

[8] A. Rohman, “Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Penyakit Jantung,” *Neo Teknika: Jurnal Ilmiah Teknologi,* vol. 2, no. 2, 2016.

[9] F. Latuconsina, M. S. Noya van Delsen dan Yudistira, “Klasifikasi Menggunakan Metode Support Vector Machine (SVM) Multiclass pada Data Indeks Desa Membangun (IDM) di Provinsi Maluku,” *Journal of Mathematics, Computation, and Statistics (JMATHCOS),* vol. 7, no. 2, pp. 380-395, 2024.

[10] A. H. Talakua, G. Haumahu dan M. S. Noya Van Delsen, “Klasifikasi Penggunaan Alat Kontrasepsi di Kecamatan Salahutu Kabupaten Maluku Tengah Menggunakan Metode K-Nearest Neighbor (KNN),” *Proximal: Jurnal Penelitian Matematika dan Pendidikan Matematika,* vol. 7, no. 2, pp. 752-759, 2024.

[11] A. Kowalczyk, Support Vector Machines Succinctly, USA: Syncfusion, 2017.

[12] C.-W. Hsu, C.-C. Chang dan C.-J. Lin, “A Practical Guide to Support Vector Classi cation,” 2003. [Online]. Available: https://github.com/tpn/pdfs/blob/master/A%20Practical%20Guide%20to%20Support%20Vector%20Classification.pdf. [Diakses 2025].

[13] A. Ben-Hur dan J. Weston, “A User’s Guide to Support Vector Machines,” *Springer Protocols,* pp. 223-229, 2009.

[14] E. A. Sari, M. T. Br. Saragih, I. A. Shariati, S. Sofyan, R. A. Baihaqi dan R. Nooraeni, "Klasifikasi Kabupaten Tertinggal di Kawasan Timur Indonesia dengan Support Vector Machine," *JIKO: Jurnal Informatika dan Komputer,* vol. 3, no. 3, pp. 188-195, 2020.

[15] E. Prasetyo, Data Mining: Konsep dan Aplikasi menggunakan MATLAB, Yogyakarta: ANDI, 2012.

[16] Y. Yang dan X. Liu, "A re examination of text categorization methods," dalam *SIGIR '99: Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, USA, 1999.

[17] F. Gorunescu, Data Mining: Concepts, Models and Techniques, Berlin: Springer Science & Business Media, 2011.