

Classification of Creditworthy Customer Using Support Vector Machine Algorithm

Ahmad ^{a,1,*}, Sartika Lina Mulani Sitio ^{a,2}, Nur Rofiq ^{a,3}

^a University of Pamulang, Jl. Raya Puspitek, Kec. Pamulang, Kota Tangerang Selatan, 15310, Indonesia

¹ dosen02594@unpam.ac.id ^{*}; ² dosen00847@unpam.ac.id; ³ dosen00376@unpam.ac.id

^{*}corresponding author

ARTICLE INFO

Article history:
Published
September 6, 2025

Keywords:
Confusion Matrix
Credit Eligibility
Customer Classification
Data Preprocessing
Support Vector Machine

ABSTRACT

The increase in the number of credit applications in the banking and financial institutions sector requires an efficient and accurate creditworthiness assessment system. Manually conducted assessments tend to be time-consuming and prone to subjectivity, so they can have an impact on errors in credit decision-making. The main problem faced is how to classify customers appropriately into creditworthy or non-creditworthy categories based on historical data. To overcome this, this study proposes the use of the Support Vector Machine (SVM) algorithm as an artificial intelligence-based solution that is able to handle classification problems with a high level of accuracy. The purpose of this study is to develop a model for classifying customer creditworthiness using the SVM algorithm and optimizing parameters to improve model performance. The methods used include the data preprocessing stage (handling missing values, categorical data encoding, and normalization), data division into training and test data, SVM model training, performance evaluation with accuracy, precision, recall, F1-score metrics, and parameter tuning using Grid Search and visualization through heatmaps. Kernel comparisons are also done to obtain the best configuration of the model. The results of the study show that the SVM model with the RBF kernel provides the best test performance reaching 87%, which means that the model is very good at recognizing potential customers who are not creditworthy. These results show that the SVM algorithm is effective in classifying customers' creditworthiness, so that it can be used as a decision-making tool in a more objective and efficient credit selection process.

Copyright © 2025 by the Authors.

I. Introduction

Lending is one of the main services in the banking industry and financial institutions. The credit provided to customers must go through a strict selection process to ensure that the risk of default can be minimized. One of the main challenges in this process is how to predict whether a potential customer is eligible for credit based on available data and history [1]. Errors in this decision-making can lead to significant financial losses for the institution. So far, the creditworthiness assessment process relies heavily on manual or simple rule-based methods determined by credit analysts. However, this approach is prone to subjective bias, limited historical data, and inefficiency in handling large amounts of data [2]. Therefore, a technology- and data-based approach is needed to improve accuracy and efficiency in the credit selection process. One of the relevant solutions is the application of machine learning techniques, especially classification algorithms, to model the patterns and characteristics of customers who are eligible and unworthy of receiving credit. Among the various algorithms available, the Support Vector Machine (SVM) is known to have a high ability to handle binary classification problems and perform well on high-dimensional data [3].

Research by [4] shows that SVM has a high level of accuracy in predicting creditworthiness compared to other algorithms such as K-Nearest Neighbors and Decision Tree. Another study by [5] states that the RBF kernel in SVM is very effective in modeling the complex relationship between input variables and outputs. Then the research conducted by [6] proves that the use of SVM can



improve the accuracy of credit decisions on real banking datasets. Based on these problems, this study aims to develop a creditworthy customer classification system using the SVM algorithm. This system will be built on historical credit data that has been processed with preprocessing techniques to eliminate blank values, encode categorical data, and normalize numerical features to match the characteristics of SVM inputs. The methods used in this study include several important stages: data cleaning, categorical value encoding, data normalization with StandardScaler, model training with SVM using the Radial Base Function (RBF) kernel, and performance evaluation using confusion matrix and classification report. In addition, a parameter tuning process is also carried out using GridSearchCV to obtain optimal performance from the model built.

The main advantages of this approach are its ability to handle high-dimensional data as well as its resistance to overfitting [7]. Using the RBF kernel, SVMs can map non-linear data to higher-dimensional feature spaces so that they can be separated linearly together. This provides an advantage in complex case classifications, such as creditworthiness involving many interrelated factors [8]. With this research, it is hoped that an accurate and reliable classification model can be obtained in helping financial institutions make objective and data-based lending decisions. In addition to increasing operational efficiency, this system also has the potential to reduce the risk of losses due to errors in the assessment of prospective customers.

II. Method

This study uses a quantitative approach by applying machine learning methods, especially the Support Vector Machine (SVM) algorithm, to build a customer classification model based on credit history data. This research method consists of several main stages as follows [9][10]:

A. Data Collection Methods

The data used is a Credit Approval dataset from the UCI Machine Learning Repository, which contains information on customer attributes such as age, occupation, marital status, income, and credit application decision results. This dataset was chosen because it is representative and commonly used in creditworthiness studies.

B. Preprocessing Data

The preprocessing stage is essential for the data to be optimally used by the SVM model. This process includes [11]:

1. Data Cleanup: Deletes rows with empty values (NaN) derived from the symbol ?.
2. Label Encoding: Converts categorical data into numerical forms so that the model can process.
3. Normalization: Using StandardScaler to standardize numerical features to have a uniform scale (mean = 0 and std = 1), which is important for SVM performance.

C. Modeling

The SVM model is built using the Radial Base Function (RBF) kernel because of its ability to handle non-linear data. The main parameters tested and adjusted are:

1. C (regularization): Controls the trade-off between misclassification on training data and decision margins.
2. Gamma: Determines the extent of the influence of a single training data.

D. Training and Testing

The dataset is divided into two parts:

1. Training data: To train the model.
2. Test data: To evaluate the performance of the model that has been trained.
The distribution was done randomly with a general proportion of 80% for training and 20% for testing.

E. Model Evaluation

Evaluation is conducted using the following metrics:

1. Confusion Matrix: To see the number of true and false predictions of each class.
2. Classification Report: Provides precision, recall, f1-score, and accuracy values.
3. Cross Validation (optional): To test the stability of the model.

F. Tuning Parameter

To optimize model performance, parameter tuning was performed using GridSearchCV, by trying various combinations of C and gamma values. The best results are measured based on the highest accuracy score in the test data.

III. Results and Discussion

A. Goods Distribution Table

The following is a table of goods distribution data obtained from PT Acel Furniture:

Table 1. Performace Metrics

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
0	0.86	0.92	0.89	77
1	0.89	0.80	0.84	61
accuracy			0.87	138
macro avg	0.87	0.86	0.87	138
Weighted avg	0.87	0.87	0.87	138

B. Performance Metrics

The performance of the Support Vector Machine (SVM) classification model was evaluated using several performance metrics: accuracy, precision, recall, and F1 score. This metric provides a comprehensive understanding of the effectiveness of the model in classifying credit customer eligibility. Results for each training and test data partition. The results of the accuracy, precision, recall and F1 score scores for the various training-to-test separations can be seen in the figure below:

Table 2. Performace Metrics

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
0	0.86	0.92	0.89	77
1	0.89	0.80	0.84	61
accuracy			0.87	138
macro avg	0.87	0.86	0.87	138
Weighted avg	0.87	0.87	0.87	138

Table 3. Train Set Metrics and Test Set Metrics

	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Confusion Matrix</i>
Train set metrics	0.86	0.86	0.82	0.84	[[273 33] [42 204]]
Test set metrics	0.87	0.89	0.80	0.84	[[71 6] [12 49]]

The test is performed on data that the model never saw during training. It reflects actual performance in the real world. The confusion metrics for the data set are:

$$\begin{bmatrix} 71 & 6 \\ 12 & 49 \end{bmatrix}$$

Table 4. Confusion Metrics

	<i>Prediksi 0</i>	<i>Prediksi 1</i>
Kelas 0 (nyata)	71 (TN)	6 (FP)
Kelas 1 (nyata)	12 (FN)	49 (TP)

True Negative (TN) = 71 → Class 0 (not creditworthy) data was correctly predicted.

False Positive (FP) = 6 → Class 0 data is incorrectly classified as creditworthy.

False Negative (FN) = 12 → Class 1 data is incorrectly classified as not creditworthy.

True Positive (TP) = 49 → Class 1 (creditworthy) data was correctly predicted.

The following is a picture of the confusion metrics for data test using SVM

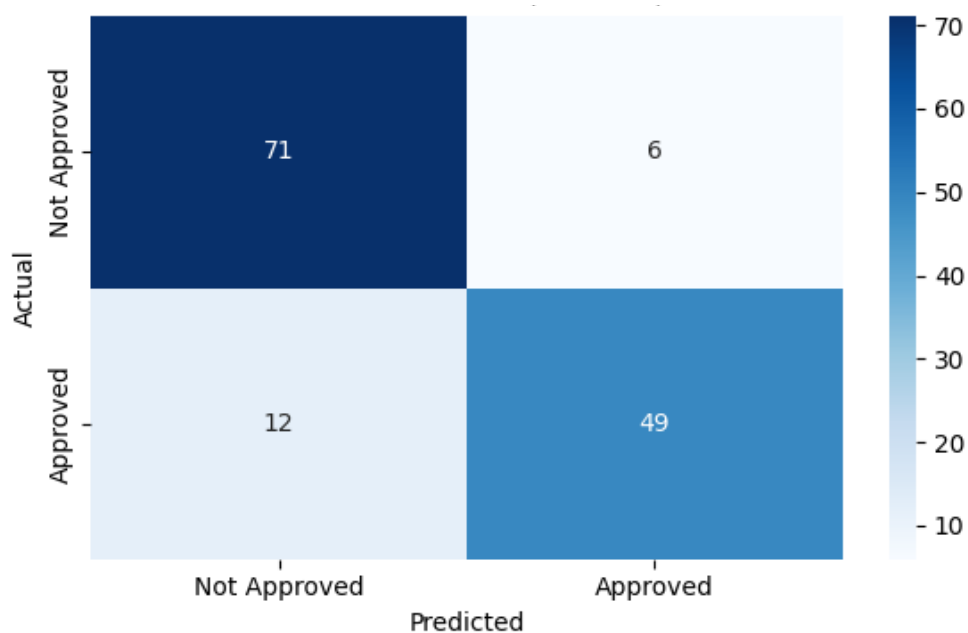


Fig. 1. Confusion Matrix Data Test

Based on the confusion matrix displayed, the SVM model managed to classify most of the customers correctly. Out of a total of 138 test data, as many as 71 customers who were indeed not creditworthy were correctly predicted as unqualified, and 49 credit-eligible customers were also correctly predicted as eligible. This shows that the model has high accuracy and is able to distinguish well between two main classes: creditworthy and non-creditworthy. With only 18 cases of prediction errors (6 false positives and 12 false negatives), the model can be said to be quite reliable in the creditworthiness selection process. However, the existence of 12 false negatives shows that there are a number of customers who actually deserve credit, but are predicted as unworthy by the model. This mistake is more crucial in the context of financial institutions, because it risks ignoring potential potential customers. On the other hand, the number of false positives (customers who are not eligible but predicted to be deserving) is 6, which also needs to be considered because it can increase the risk of bad loans. Therefore, while the model's performance is generally good, the balance between recall

and precision needs to be considered further, especially if it is used in a sensitive decision-making process such as credit approval.

C. Kernel Comparison and Parameter Turning

The figure below shows the learning curve of the Support Vector Machine (SVM) model with the RBF kernel used for the classification of customer creditworthiness. The graph shows trends in training accuracy and validation accuracy in the amount of training data used. It was seen that the training accuracy was consistently high, ranging from 81% to 88%, indicating that the model was able to learn patterns from the training data well. Validation accuracy also increases progressively as the data increases, from about 68% to more than 84%, reflecting the model's increased generalization of new data. The relatively small gap between the training and validation curves at the end of the graph indicates that the model is not significantly overfitting, and the model performance is fairly stable and optimal with larger data sizes.

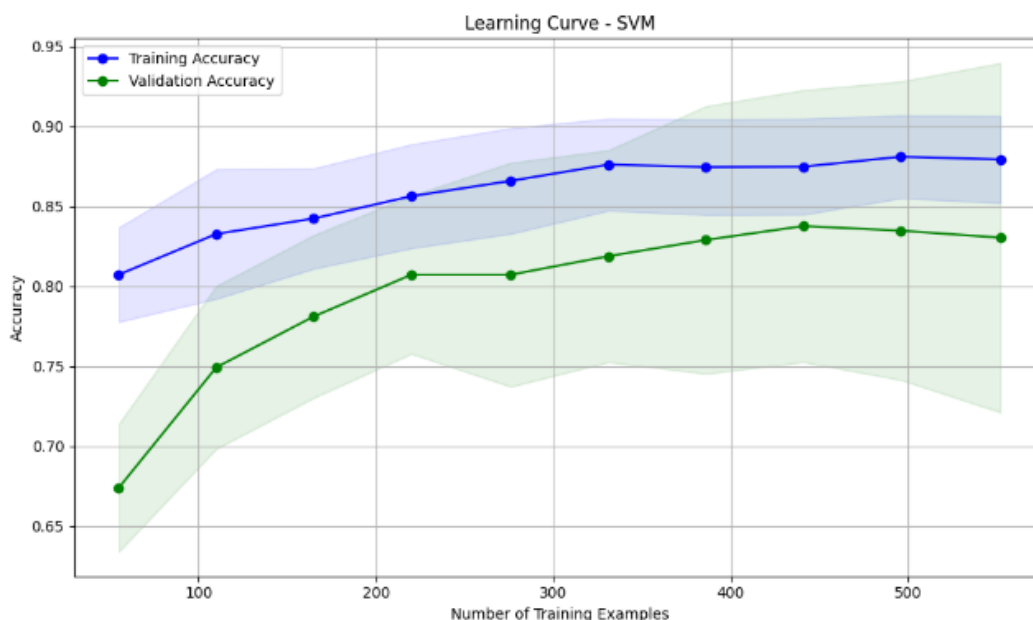


Fig. 2. Learning Curve - SVM

D. Tuning Parameters

The image below is a heatmap of the results of the Grid Search process for the selection of the best parameters in the SVM (Support Vector Machine) model with the RBF kernel. The two main parameters tested were C (regularization parameter) and gamma (kernel coefficient) and the numbers in each box represented the cross-validation accuracy score obtained for a given combination of C and gamma pairs. The darker the blue color, the higher the accuracy. The highest accuracy value was achieved at the combination of $C=10$ and $\gamma=0.001$, which was 0.855 or 85.5%. From this heatmap visualization, it can be concluded that the optimal parameter selection greatly affects the performance of the SVM model. Too large a gamma value (e.g. 1.0) and a C value that is too small (e.g. 0.1) result in low accuracy. In contrast, the combination of $C=10$ and $\gamma=0.001$ yielded the best results, suggesting that a balanced model between regularization and sensitivity to data can achieve optimal classification performance. It also shows the effectiveness of grid search in helping the process of tuning parameters systematically.

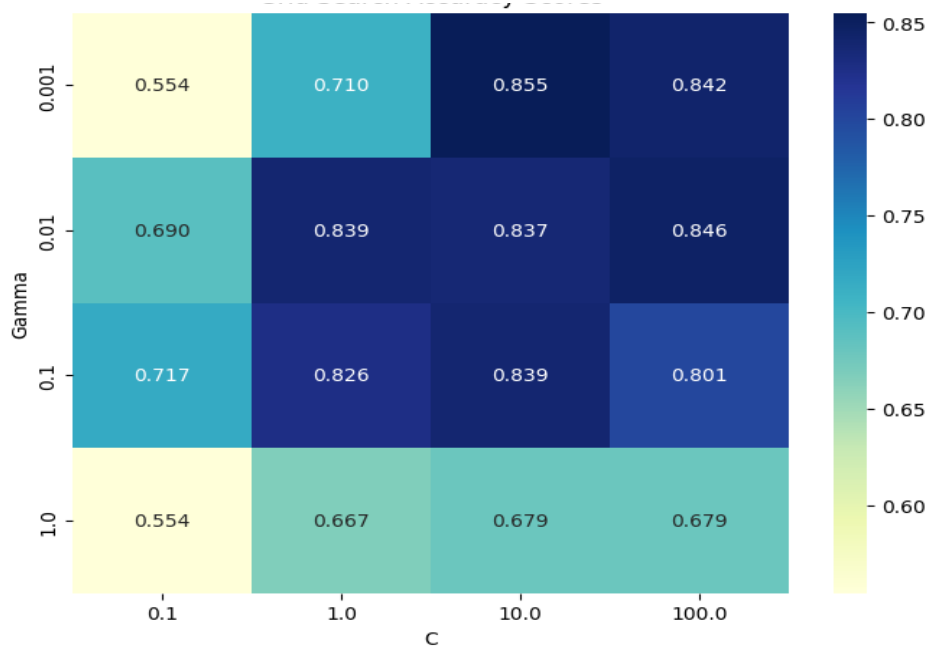


Fig. 3. Tuning Parameters

IV. Conclusion

Based on the results of the research that has been conducted using the Support Vector Machine (SVM) algorithm with the Radial Base Function (RBF) kernel, it can be concluded that the selection of model parameters through the Grid Search method has an important role in improving classification performance. The combination of $C=10$ and $\text{gamma}=0.001$ parameters was proven to produce the highest accuracy of 85.5% at the validation stage, which was then able to produce 87% accuracy in the test data based on the results of the classification report. This suggests that a well-tuned model is able to classify data with a low error rate. In addition, the confusion matrix analysis showed that the model was able to classify the data fairly evenly between the "Approved" and "Not Approved" classes, with high precision, recall, and f1-score values in both classes. The learning curve shown also shows that the model has performance stability between training and validation data, and does not show significant overfitting symptoms. Therefore, the SVM approach with the RBF kernel and parameter optimization via Grid Search can be considered effective and can be recommended for use in future cases of similar classifications.

References

- [1] M. R. Givari, M. R. Sulaeman, and Y. Umidah, "Perbandingan Algoritma SVM, Random Forest Dan XGBoost Untuk Penentuan Persetujuan Pengajuan Kredit," *NUANSA Inform.*, vol. 16, no. 1, 2022, doi: 10.25134/nuansa.v16i1.5406.
- [2] D. Pakuan Putra and B. Agus Wardijono, "Analisis Akurasi Penerapan Algoritma Support Vector Machine Menggunakan Kernel Radial Basis Function pada Penentuan Kelayakan Kredit (Studi Kasus German Kredit Data)," *J. Ilmiah KOMPUTASI*, vol. 19, no. 2, 2020.
- [3] S. Samsuri, "Penentuan Kelayakan Dan Besaran Pinjaman Pada Koperasi Di Banjarmasin Memanfaatkan Support Vector Machine (SVM) Dan Regresi Linier Berganda," *J. Sains Komput. dan Teknol. Inf.*, vol. 4, no. 2, 2022, doi: 10.33084/jsakti.v4i2.2838.
- [4] P. Golbayani, I. Florescu, and R. Chatterjee, "A comparative study of forecasting corporate credit ratings using neural networks, support vector machines, and decision trees," *North Am. J. Econ. Financ.*, vol. 54, 2020, doi: 10.1016/j.najef.2020.101251.
- [5] N. Rtayli and N. Enneya, "Selection features and support vector machine for credit card risk

- identification,” in *Procedia Manufacturing*, 2020. doi: 10.1016/j.promfg.2020.05.012.
- [6] A.- Amrin and O.- Pahlevi, “Implementation of Logistic Regression Classification Algorithm and Support Vector Machine for Credit Eligibility Prediction,” *J. INFORMATICS Telecommun. Eng.*, vol. 5, no. 2, 2022, doi: 10.31289/jite.v5i2.6220.
- [7] I. P. Casuarina, M. N. Hayati, and D. S. Prangga, “Klasifikasi Status Pembayaran Kredit Barang Elektronik dan Furniture Menggunakan Support Vector Machine Classification of Credit Payment Status for Electronic and Furniture Using Support Vector Machine,” *J. Eksponensial*, vol. 13, no. 1, 2022.
- [8] K. Amzile and M. Habachi, “Assessment of Support Vector Machine performance for default prediction and credit rating,” *Banks Bank Syst.*, vol. 17, no. 1, 2022, doi: 10.21511/BBS.17(1).2022.14.
- [9] D. Zhang, B. Bhandari, and D. Black, “Credit Card Fraud Detection Using Weighted Support Vector Machine,” *Appl. Math.*, vol. 11, no. 12, 2020, doi: 10.4236/am.2020.1112087.
- [10] F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan, and M. Ahmed, “Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms,” *IEEE Access*, vol. 10, 2022, doi: 10.1109/ACCESS.2022.3166891.
- [11] Salya Rater, Balkhaya, The Influence of Social Media Digital Marketing Analysis on New Student Acceptance in a University, *J. Inotera*, Vol. 8, No. 1, pp. 150-154, January-June 2023
- [12] M. G. Kibria and M. Sevkli, “Application of Deep Learning for Credit Card Approval: A Comparison with Two Machine Learning Techniques,” *Int. J. Mach. Learn. Comput.*, vol. 11, no. 4, 2021, doi: 10.18178/ijmlc.2021.11.4.1049.