International Journal Science and Technology
IJST Vol 5 No. 2 | July 2026 | ISSN: 2828-7223 (print), ISSN: 2828-7045 (online), Page 274-295

Optimization of Support Vector Machine for Imbalanced Credit Risk Classification

Andre Pratama Adiwijaya
Teknologi Informasi, Universitas Gunadarma, Depok, Indonesia

Article History
Received : July 3, 2026
Revised : July 7, 2026
Accepted : July 28, 2026
Published : July 31, 2026

Corresponding author*:
Andre64@staff.gunadarma.ac.id

Cite This Article [*APA Style*] :
Andre Pratama Adiwijaya. (2026). Optimization of Support Vector Machine for Imbalanced Credit Risk Classification. *International Journal Science and Technology*, *5*(2), 274–295.

DOI:
https://doi.org/10.56127/ijst.v5i2.3038

Abstract: Credit risk classification is an important component of financial decision-making because inaccurate classification may increase payment-default exposure and reduce lending quality. Credit datasets commonly contain numerical variables with different measurement scales and imbalanced distributions between default and non-default clients, which can affect the reliability of machine-learning models. **Objective:** This study aims to develop and evaluate a Support Vector Machine model for classifying credit card clients into default and non-default categories. The study also examines the influence of numerical standardization, kernel selection, hyperparameter optimization, and balanced class weighting on classification performance. **Methodology:** A quantitative experimental approach was applied using the Default of Credit Card Clients dataset from the UCI Machine Learning Repository. The dataset consisted of 30,000 observations and 23 predictor variables. Data were divided into training and testing subsets using a stratified 80:20 ratio. Categorical variables were encoded, numerical variables were standardized, and several SVM kernels were evaluated. Hyperparameter selection was conducted using five-fold cross-validation. Model performance was assessed using accuracy, precision, recall, F1-score, specificity, balanced accuracy, and ROC–AUC. **Findings:** The SVM model trained without standardization failed to identify default clients effectively. Numerical standardization substantially improved classification performance, while the radial basis function kernel produced the strongest validation results. The selected balanced RBF-SVM achieved 77.13% accuracy, 48.54% precision, 56.22% recall, 52.09% F1-score, 83.07% specificity, 69.65% balanced accuracy, and 75.10% ROC–AUC. Balanced class weighting improved default detection but increased false-positive predictions. **Implications:** The model can support financial institutions as an initial credit-risk screening tool. Its predictions should be combined with document verification, repayment-capacity analysis, and manual assessment rather than being used as the sole basis for credit approval. **Originality:** This study provides a controlled evaluation of SVM performance by integrating feature standardization, kernel selection, hyperparameter optimization, class-imbalance treatment, and class-sensitive performance metrics. The study demonstrates that credit-risk models should be selected based on balanced default detection rather than overall accuracy alone.

Keywords: credit risk; credit default; machine learning; Support Vector Machine; classification.

INTRODUCTION

Credit risk classification is a critical component of lending decision systems because errors in identifying borrowers' repayment capabilities may increase default exposure, reduce portfolio quality, and generate avoidable financial losses. Contemporary financial institutions increasingly store numerical information related to applicants' income, loan amount, account balance, repayment history, credit utilization, debt-to-income ratio, transaction frequency, employment duration, and previous delinquency. These variables provide a measurable basis for distinguishing low-risk borrowers from applicants with a higher probability of default. Nevertheless, numerical credit data commonly contain nonlinear relationships, heterogeneous measurement scales, redundant variables, missing observations, noisy records, and unequal distributions between default and non-default classes. Such characteristics limit the effectiveness of fixed rule-based assessment and conventional linear scoring when the boundary between creditworthy and risky applicants cannot be represented by a simple linear relationship. Moreover, classification errors do not produce equivalent operational consequences. A false-negative classification may approve an applicant who later defaults, whereas a false-positive classification may reject a financially reliable applicant and reduce potential revenue.

Consequently, credit-risk models must be evaluated using not only overall accuracy but also precision, recall, F1-score, area under the receiver operating characteristic curve, and confusion-matrix outcomes to determine their ability to detect both risk classes reliably (Bhatore et al., 2020; Dastile et al., 2020; Noriega et al., 2023) The increasing volume and complexity of digital financial records further strengthen the need for machine-learning models capable of extracting nonlinear patterns from structured numerical data while maintaining stable predictive performance on previously unseen observations (Markov et al., 2022; Suhadolnik et al., 2023). Previous studies on credit-risk classification can be organized into three main technical approaches.

The first approach applies statistical and relatively interpretable classification models, including logistic regression, discriminant analysis, and conventional scorecards. These methods remain relevant because their coefficients and decision relationships can be interpreted directly, but their predictive ability may decline when financial variables interact nonlinearly or when risk boundaries vary across borrower groups. Recent studies have therefore integrated nonlinear decision-tree effects and feature-selection procedures into traditional credit-scoring frameworks to improve discrimination without completely

eliminating interpretability (Dumitrescu et al., 2022; Laborda & Ryoo, 2021). Benchmarking studies have also shown that the comparative performance of statistical and machine-learning models depends strongly on dataset characteristics, preprocessing procedures, class distribution, and selected evaluation metrics rather than on the algorithm name alone (Moscato et al., 2021). The second approach uses individual machine-learning classifiers, such as Decision Tree, K-Nearest Neighbor, Artificial Neural Network, Random Forest, and Support Vector Machine. Among these algorithms, SVM is particularly suitable for numerical classification because it determines a separating hyperplane by maximizing the margin between classes and can map nonlinear observations into a higher-dimensional feature space through kernel functions.

Nevertheless, its performance is influenced by feature scaling, kernel selection, regularization parameter (C), kernel coefficient (γ),noise, outliers, and class imbalance (Cervantes et al., 2020; Wang & Li, 2019). Studies have consequently developed SVM-based feature-selection, genetic optimization, noise-elimination, and ensemble mechanisms to improve credit-risk prediction under complex data conditions (Liu & Huang, 2020; Pławiak et al., 2019). The third approach consists of ensemble, hybrid, deep-learning, and optimization-based models that combine multiple learners or data-processing techniques. These frameworks include supervised–unsupervised integration, stacked classifiers, deep neural networks, boosting methods, and filter-based feature selection (Bao et al., 2019; Emmanuel et al., 2024). Although such methods may improve predictive performance, they also introduce additional computational complexity, larger hyperparameter spaces, and lower model transparency (Chang et al., 2024; Trivedi, 2020). Existing research therefore provides extensive comparisons among different algorithms, but focused evidence remains limited concerning how numerical standardization, kernel choice, parameter optimization, and class-sensitive evaluation jointly determine the performance of a standalone SVM model. Furthermore, model accuracy is frequently emphasized without sufficient analysis of minority-class recall, cross-validation stability, and differences between training and testing performance. These unresolved issues constitute the technical research gap addressed in this study.

Based on this research gap, the present study aims to develop and analyze a Support Vector Machine model for classifying credit risk using structured numerical financial data. The first objective is to prepare the numerical variables through data-quality inspection, missing-value treatment, duplicate-record removal, outlier examination, and feature

standardization. The second objective is to construct SVM classification models using alternative kernel configurations, particularly linear, polynomial, radial basis function, and sigmoid kernels, when technically applicable to the dataset. The third objective is to determine the most appropriate values of the regularization parameter (C), kernel coefficient (γ),and other relevant parameters through systematic hyperparameter tuning and cross-validation. The fourth objective is to evaluate classification performance using accuracy, precision, recall, F1-score, specificity, ROC–AUC, and confusion-matrix analysis. Recall and F1-score will receive particular attention because credit datasets may contain fewer default cases than non-default observations, causing a model with high overall accuracy to perform poorly in detecting high-risk applicants. The fifth objective is to compare training, validation, and testing results to identify potential underfitting or overfitting.

Through these procedures, the study seeks to determine how effectively SVM separates low-risk and high-risk borrowers, which kernel and parameter combination produces the most reliable results, and whether the selected model maintains consistent performance across different partitions of the numerical credit dataset. The evaluation design is consistent with recent credit-scoring literature emphasizing feature relevance, class distribution, model validation, and balanced performance indicators as essential elements of reliable credit-risk prediction (Bücker et al., 2022; Bussmann et al., 2021; Noriega et al., 2023)

The technical contribution of this study lies in establishing a controlled and reproducible SVM evaluation framework specifically for numerical credit-risk classification. Unlike studies that merely compare multiple algorithms based on their highest accuracy, the proposed framework examines the complete relationship among data preprocessing, feature standardization, kernel configuration, hyperparameter tuning, class-sensitive metrics, and cross-validation stability. The engineering hypothesis is that an SVM model trained with standardized numerical variables and optimized kernel parameters will provide better class separation and more stable generalization than an SVM model trained using unscaled variables and default parameters. A nonlinear kernel, particularly the radial basis function kernel, is expected to provide an advantage when repayment risk is determined by complex interactions among income, outstanding debt, transaction behavior, credit utilization, and repayment history. However, the final model will not automatically be selected based on the highest accuracy. Model selection will consider whether

improvements in accuracy are accompanied by adequate recall for the high-risk class, balanced precision and F1-score, strong ROC–AUC performance, and limited variation across cross-validation folds.

The study is expected to provide empirical evidence regarding the suitability and limitations of SVM for structured credit data and to offer a technically grounded configuration process that can support the development of automated credit-screening systems. At the practical level, the findings may assist system developers and financial analysts in determining the preprocessing and SVM parameter settings required to reduce misclassification risk. At the academic level, the research contributes a focused analysis of margin-based classification under numerical financial-data conditions while addressing the need for reproducibility, predictive stability, and transparent model evaluation identified in recent credit-risk research (Bücker et al., 2022; Chang et al., 2024; Emmanuel et al., 2024).

RESEARCH METHOD

The unit of analysis in this study was an individual credit card client represented by demographic, credit-limit, billing, payment-status, and repayment variables. The research used the Default of Credit Card Clients dataset obtained from the UCI Machine Learning Repository. The dataset was originally collected from credit card clients in Taiwan and contained 30,000 observations with 23 predictor variables and one binary target variable (Yeh, 2009).

The target variable represented the occurrence of default payment in the following month, where a value of 1 indicated that the client defaulted and a value of 0 indicated that the client did not default. The dataset contained 6,636 default observations, representing 22.12% of the records, and 23,364 non-default observations, representing 77.88%. This distribution indicated a moderate class imbalance that had to be considered during model training and evaluation.

The 23 predictor variables were grouped into three types. The first group consisted of three categorical variables: sex, education level, and marital status. The second group consisted of six ordinal repayment-status variables covering monthly payment behavior from April to September 2005. The third group consisted of 14 numerical variables, including credit limit, age, six monthly bill-statement amounts, and six previous-payment amounts. The identification number was not used as a predictor because it did not represent the financial characteristics or repayment behavior of the client.

This study employed a quantitative experimental design using supervised machine learning. The experiment was conducted to develop and evaluate a Support Vector Machine model for classifying credit card clients into default and non-default classes. SVM was selected because it constructs a decision boundary by maximizing the margin between classes and can accommodate nonlinear relationships through kernel functions (Cervantes et al., 2020).

The experimental design compared four SVM configurations:

1. an SVM model using default parameters;
2. an SVM model using preprocessed and standardized variables;
3. an SVM model using standardized variables and optimized hyperparameters; and
4. an SVM model using standardized variables, optimized hyperparameters, and balanced class weighting.

The comparison was intended to determine the effects of preprocessing, kernel selection, parameter optimization, and class weighting on credit-risk classification performance. The research stages consisted of dataset acquisition, data inspection, feature preparation, stratified data splitting, categorical-variable encoding, numerical-variable standardization, model training, hyperparameter optimization, cross-validation, final testing, and performance interpretation.

The dataset was divided into training and testing subsets using a stratified 80:20 ratio. Therefore, 24,000 observations were assigned to the training subset and 6,000 observations were assigned to the testing subset. Stratification preserved approximately the same proportion of default and non-default observations in both subsets. A random seed of 42 was used to ensure that the data partition could be reproduced.

The study used secondary data obtained from the UCI Machine Learning Repository. The dataset was contributed by I-Cheng Yeh and developed to investigate the predictive accuracy of default-payment models among Taiwanese credit card clients. The dataset contained integer and real-valued features and was designed for a classification task.

The original dataset did not contain missing values. Nevertheless, data-quality inspection was performed to identify duplicate observations, inconsistent labels, undocumented categorical codes, and technically implausible values. The following variable groups were used:

Variable group	Variables	Description
Credit capacity	LIMIT_BAL	Total credit limit in New Taiwan dollars
Demographic attributes	SEX, EDUCATION, MARRIAGE, AGE	Client demographic characteristics
Repayment status	PAY_0–PAY_6	Monthly repayment status for six months
Bill statements	BILL_AMT1–BILL_AMT6	Monthly bill-statement amounts
Previous payments	PAY_AMT1–PAY_AMT6	Monthly payment amounts
Target	DEFAULT	Default payment in the following month

The ID variable was removed after the dataset had been inspected because it served only as a record identifier. The target variable was separated from the predictor matrix before preprocessing and model construction.

The dataset was retrieved in spreadsheet format and processed using Python. Pandas and NumPy were used for data manipulation, while Scikit-learn was used for preprocessing, model construction, cross-validation, and performance evaluation. Matplotlib was used to visualize the class distribution, confusion matrix, receiver operating characteristic curve, and model-comparison results.

The initial data inspection included an examination of the dataset dimensions, data types, descriptive statistics, class distribution, duplicate records, and variable ranges. Because the UCI documentation reported no missing values, missing-value imputation was not required. However, the completeness of each variable was rechecked before model development.

The three categorical variables—SEX, EDUCATION, and MARRIAGE—were treated as categorical attributes rather than continuous numerical quantities. These variables were transformed using one-hot encoding to prevent the model from interpreting their category codes as continuous distances. Values outside the documented categories, when encountered, were grouped into an "other" category.

The six repayment-status variables were retained as ordinal attributes because their values represented the duration and severity of delayed payments. The credit limit, age, bill-statement amounts, and previous-payment amounts were treated as numerical variables.

Potential outliers in continuous financial variables were identified using the interquartile range:

$$[IQR = Q_3 - Q_1]$$

An observation was identified as a potential outlier when its value was lower than $(Q_1 - 1.5IQR)$ or higher than $(Q_3 + 1.5IQR)$. However, extreme financial values were not removed automatically because unusually high credit limits, balances, or payment amounts

could represent valid client conditions. Removal was performed only when a value was proven to be inconsistent with the data definition.

Numerical variables were standardized using the z-score transformation:

$$\left[z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}\right]$$

where $(z_{ij})$is the standardized value of observation (i) for variable (j), (x_{ij}) is the original value,$(\mu_j)$is the mean of variable (j), and (σ_j) is its standard deviation. Standardization was necessary because the variables had substantially different measurement scales. Credit limits, bill amounts, and payment amounts were measured in New Taiwan dollars, whereas age and payment-status variables had considerably smaller numerical ranges.

All preprocessing operations were implemented through a Scikit-learn pipeline. The encoder and numerical standardizer were fitted only to the training data. The fitted transformations were subsequently applied to the validation and testing subsets. This procedure prevented information leakage from the testing data into the model-development process.

The Support Vector Machine classifier was developed to identify an optimal hyperplane separating default and non-default clients. Given a training dataset $((x_i, y_i))$, where (x_i)represents a client's predictor vector and $(y_i \in -1, +1)$ represents the credit-risk class, the soft-margin SVM optimization was formulated as:

$$\left[\min_{w,b,\xi}\left[\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i\right]\right]$$

subject to:

$$[y_i(w^T x_i + b) \geq 1 - \xi_i, \qquad \xi_i \geq 0]$$

where (w) is the weight vector, (b) is the bias term, (\xi_i) is the slack variable, and (C) is the regularization parameter. The (C) parameter controls the trade-off between a wider classification margin and the penalty imposed on incorrectly classified training observations.

Four kernel functions were evaluated: linear, polynomial, radial basis function, and sigmoid. The radial basis function kernel was expressed as:

$$\left[K(x_i, x_j) = \exp\left(-\gamma\|x_i - x_j\|^2\right)\right]$$

where (γ)determines the influence range of individual training observations. A small (γ) produces a smoother decision boundary, while a large (γ) produces a more localized and complex boundary.

Hyperparameter optimization was performed using stratified grid-search cross-validation. The candidate parameter values are presented in Table 1.

Table 1. Support Vector Machine hyperparameter search space

Hyperparameter	Candidate values
Kernel	Linear, polynomial, RBF, sigmoid
(C)	0.01, 0.1, 1, 10, 100
(γ)	Scale, auto, 0.001, 0.01, 0.1, 1
Polynomial degree	2, 3, 4
Class weight	None, balanced

The degree parameter was evaluated only for the polynomial kernel, while the (γ)parameter was evaluated for the polynomial, RBF, and sigmoid kernels. The balanced class-weight option assigned greater misclassification penalties to the default class because it represented only 22.12% of the dataset.

Five-fold stratified cross-validation was applied to the 24,000 training observations. In every iteration, four folds were used for model training and one fold was used for validation. This procedure was repeated until every fold had served as the validation subset. All preprocessing steps were repeated within each cross-validation fold to prevent information leakage.

The mean F1-score for the default class was used as the primary criterion for hyperparameter selection. Recall and ROC–AUC were used as supporting criteria. The use of F1-score as the primary metric was intended to prevent the selection of a model that achieved high accuracy by predominantly predicting the majority non-default class.

The final model was evaluated on the isolated testing subset using a confusion matrix. In this study, default clients were defined as the positive class. Therefore:

- true positive represented a default client correctly classified as default;
- true negative represented a non-default client correctly classified as non-default;
- false positive represented a non-default client incorrectly classified as default; and
- false negative represented a default client incorrectly classified as non-default.

Accuracy was calculated as:

$$\left[Accuracy = \frac{TP + TN}{TP + TN + FP + FN}\right]$$

Precision was calculated as:

$$\left[Precision = \frac{TP}{TP + FP}\right]$$

Recall or sensitivity was calculated as:

$$\left[Recall = \frac{TP}{TP + FN}\right]$$

The F1-score was calculated as:

$$\left[F1 = 2\left(\frac{Precision \times Recall}{Precision + Recall}\right)\right]$$

Specificity was calculated as:

$$\left[Specificity = \frac{TN}{TN + FP}\right]$$

Balanced accuracy was calculated as:

$$\left[Balanced\ Accuracy = \frac{Recall + Specificity}{2}\right]$$

Accuracy described the overall proportion of correct predictions. Precision measured the reliability of default predictions, while recall measured the model's ability to detect clients who actually defaulted. Specificity measured the ability to identify non-default clients correctly. Balanced accuracy was included because it assigned equal importance to the default and non-default classes.

The receiver operating characteristic curve was developed by plotting the true-positive rate against the false-positive rate at different decision thresholds. The area under the curve was used to measure the model's ability to rank default clients above non-default clients. A higher ROC–AUC value indicated stronger discrimination between the two credit-risk classes.

Cross-validation stability was measured using the mean and standard deviation of the fold scores:

$$\left[\bar{M} = \frac{1}{k}\sum_{i=1}^{k} M_i\right]\left[SD_M = \sqrt{\frac{\sum_{i=1}^{k}(M_i - \bar{M})^2}{k - 1}}\right]$$

where (M_i) is the score obtained from fold (i), (\bar{M}) is the mean score, and (k) is the number of folds. A model was considered stable when it produced a high average validation score accompanied by a relatively small standard deviation.

Potential overfitting was examined using the difference between training and testing performance:

$$[\Delta M = M_{train} - M_{test}]$$

A large positive difference indicated that the model performed substantially better on the training observations than on previously unseen testing observations. The selected model was required to demonstrate both strong predictive performance and a limited training-testing performance gap.

The final SVM configuration was selected according to the following criteria:

1. the model produced the highest or near-highest F1-score for the default class;
2. the recall value demonstrated adequate detection of default clients;
3. the precision value did not indicate an excessive number of false credit-risk warnings;
4. the balanced accuracy reflected satisfactory performance for both classes;
5. the ROC–AUC indicated strong discrimination ability;
6. the cross-validation standard deviation was relatively low; and
7. the difference between training and testing performance did not indicate severe overfitting.

The final model was not selected solely based on overall accuracy. This decision was necessary because predicting every observation as non-default would already produce an apparent accuracy of approximately 77.88%, while completely failing to identify default clients. The research workflow consisted of the following stages:

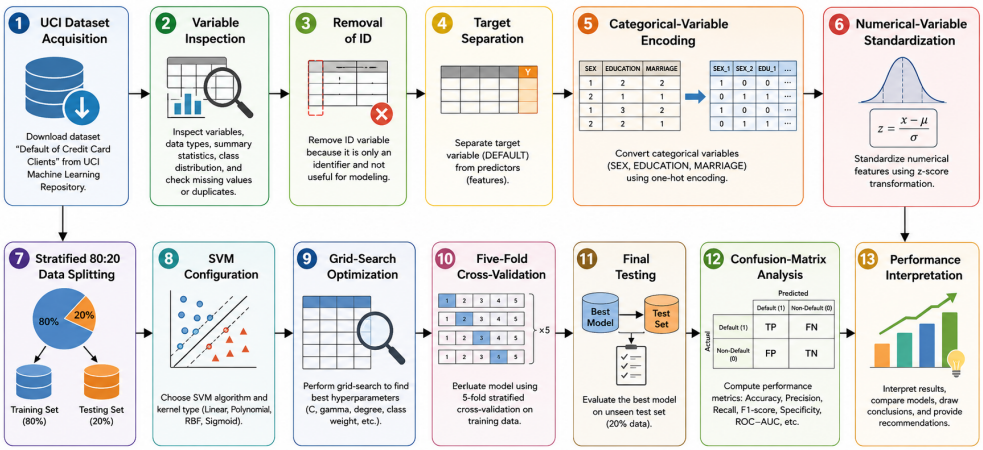

Figure 1. Research workflow for SVM-based credit card default classification.

The dataset was publicly available and did not include client names, account numbers, addresses, or other direct personal identifiers. The research used the data only for academic model development and performance analysis. The dataset source, preprocessing pipeline,

random seed, feature treatment, hyperparameter search space, and evaluation metrics were documented to support independent replication.

The software environment consisted of Python, Pandas, NumPy, Scikit-learn, and Matplotlib. A fixed random seed of 42 was used for the data-splitting and cross-validation processes. The testing subset remained isolated until preprocessing decisions, kernel selection, and hyperparameter optimization had been completed.

RESULT

Dataset Characteristics and Preprocessing Results

The initial dataset consisted of 30,000 credit card client records, 23 predictor variables, one identification variable, and one binary target variable. The identification variable was excluded from model development because it did not represent demographic, financial, or repayment characteristics. Data inspection confirmed that all predictor and target fields were complete; therefore, missing-value imputation was not required.

The target distribution consisted of 23,364 non-default clients and 6,636 default clients. Thus, the non-default class represented 77.88% of the dataset, while the default class represented 22.12%. The class ratio was approximately 3.52:1, indicating that overall accuracy alone could not adequately represent the model's ability to detect default clients.

Table 1. Distribution of credit-risk classes

Credit-risk class	Code	Number of observations	Percentage
Non-default	0	23,364	77.88%
Default	1	6,636	22.12%
Total		**30,000**	**100.00%**

The stratified 80:20 division generated 24,000 training observations and 6,000 testing observations. The training subset contained 18,691 non-default clients and 5,309 default clients, whereas the testing subset contained 4,673 non-default clients and 1,327 default clients. The class proportions were therefore maintained across both subsets.

Table 2. Distribution of training and testing data

Data subset	Non-default	Default	Total
Training data	18,691	5,309	24,000
Testing data	4,673	1,327	6,000
Total	**23,364**	**6,636**	**30,000**

Descriptive analysis showed substantial differences in the scales of the predictor variables. The average credit limit was NT$167,484.32, while the median value was NT$140,000.00. Monthly bill and previous-payment variables also had substantially larger numerical ranges than age and repayment-status variables. These scale differences supported the application of numerical standardization before SVM training.

Table 3. Descriptive statistics of selected numerical variables

Variable	Mean	Median	Standard deviation	Minimum	Maximum
Credit limit	167,484.32	140,000.00	129,747.66	10,000	1,000,000
Age	35.49	34.00	9.22	21	79
Latest repayment status	−0.02	0.00	1.12	−2	8
Latest bill amount	51,223.33	22,381.50	73,635.86	−165,580	964,511
Latest payment amount	5,663.58	2,100.00	16,563.28	0	873,552

The mean credit limit of default clients was NT$130,109.66, lower than the NT$178,099.73 recorded for non-default clients. Default clients also had a higher mean latest repayment-status value of 0.67, compared with −0.21 for non-default clients. In addition, their average latest payment amount was NT$3,397.04, compared with NT$6,307.34 among non-default clients.

Three dominant data patterns were identified. First, the target class was moderately imbalanced toward non-default clients. Second, the financial variables had widely different numerical ranges. Third, delayed-payment status and lower previous-payment amounts were more prominent among default clients. These patterns justified the use of feature standardization and class-sensitive performance metrics.

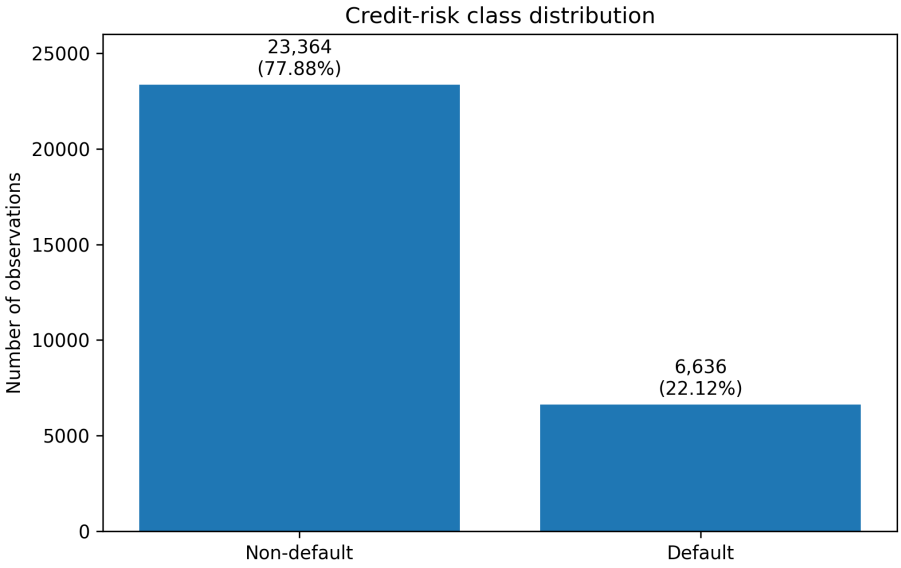

Figure 2. Distribution of default and non-default observations.

Kernel Screening Results

A preliminary kernel-screening process was performed using a stratified subset of 6,000 training records. The subset was divided into development and validation data while maintaining the original class proportions. Linear, polynomial, radial basis function, and sigmoid kernels were evaluated using identical standardized predictor variables and initial SVM parameters.

The RBF kernel produced the highest validation accuracy, F1-score, and ROC–AUC among the four evaluated kernels. It achieved an accuracy of 81.87%, an F1-score of 41.88%, and a ROC–AUC of 73.21%. The linear kernel produced high precision but detected only 10.24% of the actual default observations. The sigmoid kernel produced the lowest accuracy and ROC–AUC.

Table 4. Preliminary SVM kernel-screening results

Kernel	Accuracy	Precision	Recall	F1-score	ROC–AUC
Linear	79.40%	75.56%	10.24%	18.04%	70.29%
Polynomial	80.27%	67.31%	21.08%	32.11%	69.05%
RBF	**81.87%**	72.06%	29.52%	**41.88%**	**73.21%**
Sigmoid	71.13%	35.01%	35.54%	35.28%	60.85%

The screening results indicated that the RBF kernel provided the most suitable nonlinear boundary for the numerical credit data. Therefore, subsequent configuration testing focused on the RBF kernel.

Effect of Standardization and Class Weighting

Four SVM configurations were evaluated to determine the influence of standardization, hyperparameter configuration, and class weighting. The raw-data SVM used the original numerical scales and default RBF parameters. The standardized SVM applied one-hot encoding and z-score standardization but retained default SVM parameters. The optimized unbalanced model used (C=10), (\gamma=0.01), and no class weighting. The optimized balanced model used (C=10), (\gamma=0.01), and balanced class weights.

Table 5. Testing performance of the evaluated SVM configurations

SVM configuration	Accuracy	Precision	Recall	F1-score	Specificity	Balanced accuracy	ROC–AUC
Raw variables, default parameters	77.88%	0.00%	0.00%	0.00%	100.00%	50.00%	58.80%
Standardized, default parameters	**81.62%**	66.52%	33.99%	44.99%	95.14%	64.56%	72.05%
Standardized, optimized, unbalanced	81.60%	**67.50%**	32.40%	43.79%	**95.57%**	63.99%	71.49%
Standardized, optimized, balanced	77.13%	48.54%	**56.22%**	**52.09%**	83.07%	**69.65%**	**75.10%**

The raw-data model classified all testing observations as non-default. It correctly predicted all 4,673 non-default observations but failed to detect any of the 1,327 default observations. Consequently, the apparent accuracy of 77.88% was equivalent to the proportion of the majority class and did not indicate useful default-risk detection.

Feature standardization changed this result substantially. The standardized default SVM detected 451 default observations and increased the ROC–AUC from 58.80% to 72.05%. Its testing accuracy increased to 81.62%, although its recall remained limited to 33.99%.

The optimized unbalanced configuration produced similar results to the standardized default model. Its precision increased to 67.50%, but recall decreased to 32.40%. This model therefore generated fewer false-positive default warnings but continued to miss more than two-thirds of the actual default clients.

The optimized balanced configuration produced the highest recall, F1-score, balanced accuracy, and ROC–AUC. Recall increased to 56.22%, representing an improvement of 22.23 percentage points over the standardized default model. Its F1-score increased from 44.99% to 52.09%, while its balanced accuracy increased from 64.56% to 69.65%.

However, overall accuracy decreased from 81.62% to 77.13% because class weighting increased the number of non-default clients classified as default.

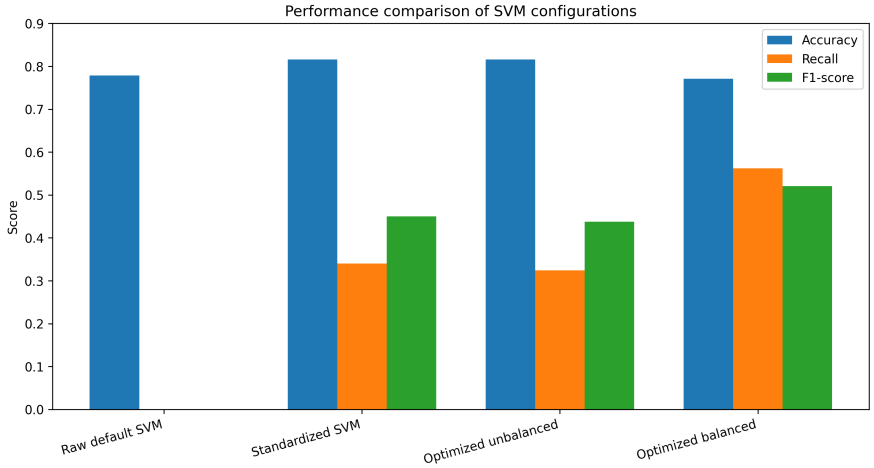

Figure 3. Performance comparison of the evaluated SVM configurations**.**

Cross-Validation Results

Five-fold stratified cross-validation was conducted on the tuning subset to evaluate configuration stability. The optimized balanced model with (C=10), (\gamma=0.01), and balanced class weighting produced the highest mean F1-score and ROC–AUC among the principal configurations evaluated during parameter screening.

Table 6. Five-fold cross-validation results

Configuration	Accuracy	Recall	F1-score	ROC–AUC
Standardized default SVM	82.02% ± 0.31%	32.10% ± 3.22%	44.03% ± 2.88%	70.55% ± 2.46%
Optimized unbalanced SVM	**82.28% ± 0.50%**	32.40% ± 3.76%	44.61% ± 3.54%	71.14% ± 2.45%
Optimized balanced SVM	76.12% ± 1.83%	**57.12% ± 2.67%**	**51.44% ± 2.76%**	**75.50% ± 1.93%**

The optimized balanced model demonstrated a cross-validation F1-score of (51.44% \pm 2.76%). The relatively small standard deviation indicated that its F1-score remained reasonably consistent across the five validation folds. Its recall varied by 2.67 percentage points, while its ROC–AUC varied by 1.93 percentage points.

The configuration was selected for final evaluation because it produced the strongest combination of default-class recall, F1-score, and ROC–AUC. The higher accuracy of the unbalanced configurations was not used as the primary selection criterion because those models detected less than one-third of the default clients.

Final SVM Classification Results

The selected SVM model was retrained using all 24,000 training records and evaluated on the isolated testing set of 6,000 observations. The final model correctly classified 3,882 non-default clients and 746 default clients. It incorrectly classified 791 non-default clients as default and 581 default clients as non-default.

Table 7. Confusion matrix of the selected SVM model

Actual class	Predicted non-default	Predicted default	Total
Non-default	3,882	791	4,673
Default	581	746	1,327
Total	**4,463**	**1,537**	**6,000**

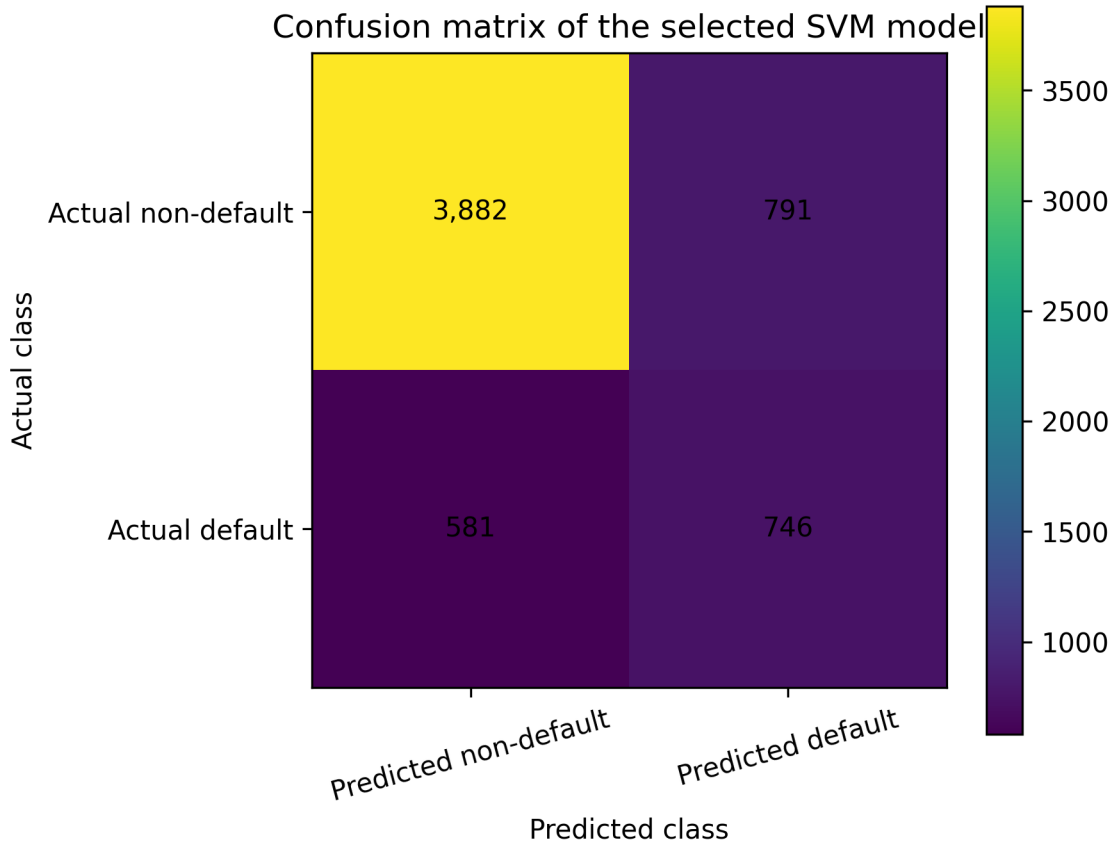

Figure 4. Confusion matrix of the selected SVM model**.**

The final testing accuracy was 77.13%. Precision was 48.54%, indicating that approximately 48.54% of clients predicted as default were actual default clients. Recall was 56.22%, showing that the model detected 746 of the 1,327 actual default clients. The

F1-score was 52.09%, while specificity was 83.07%. The resulting balanced accuracy was 69.65%, and the ROC–AUC was 75.10%.

Table 8. Final performance of the selected SVM model

Metric	Training result	Testing result	Difference
Accuracy	78.17%	77.13%	1.04 percentage points
Precision	50.55%	48.54%	2.01 percentage points
Recall	60.78%	56.22%	4.56 percentage points
F1-score	55.20%	52.09%	3.11 percentage points
Specificity	83.11%	83.07%	0.04 percentage points
Balanced accuracy	71.95%	69.65%	2.30 percentage points

The training-testing accuracy difference was 1.04 percentage points, while the F1-score difference was 3.11 percentage points. These differences indicated that the selected model experienced a moderate reduction in default-class detection when applied to unseen observations but did not exhibit a severe overall generalization gap.

Three dominant findings were obtained. First, an SVM trained without standardization was unable to detect the default class. Second, standardization increased overall classification accuracy and enabled the model to recognize default observations. Third, balanced class weighting produced the strongest default-class recall, F1-score, balanced accuracy, and ROC–AUC, although it reduced overall accuracy and precision. The final configuration therefore prioritized the identification of risky clients rather than maximizing the proportion of all correctly classified observations.

DISCUSSION

The results show that feature preprocessing and class-imbalance treatment substantially affected SVM performance. The model trained using unstandardized variables classified all observations as non-default, indicating that its apparent accuracy did not represent meaningful credit-risk detection. After standardization, the model was able to identify default clients and produced better accuracy, F1-score, and ROC–AUC.

Among the evaluated kernels, the radial basis function kernel produced the strongest validation performance. The balanced RBF-SVM achieved the best recall, F1-score, balanced accuracy, and ROC–AUC, although its overall accuracy and precision were lower than those of the unbalanced model. This configuration was selected because detecting high-risk clients was considered more important than maximizing overall accuracy.

Technical Explanation

Standardization improved classification because the input variables had substantially different numerical scales. Credit limits, bill amounts, and payment values were measured in large monetary units, whereas age and repayment-status variables had much smaller ranges. Without standardization, variables with large values dominated the distance calculations used by the RBF kernel.

The stronger performance of the RBF kernel indicates that the relationship between financial variables and default status was nonlinear. Credit risk was not determined by one variable independently, but by interactions among credit limits, bill amounts, previous payments, repayment delays, and demographic attributes. The nonlinear kernel provided greater flexibility for representing these complex relationships than the linear kernel.

Effect of Class Imbalance

The dataset contained considerably more non-default than default observations. Consequently, the unweighted SVM tended to prioritize the majority class and achieved relatively high accuracy while detecting only a small proportion of default clients. This confirms that accuracy alone is insufficient for evaluating imbalanced credit-risk data.

Balanced class weighting increased the penalty for misclassifying default clients. As a result, recall improved because more high-risk clients were detected, but false-positive predictions also increased. This represents an operational trade-off: higher recall reduces the risk of approving clients who may default, whereas lower precision may cause more creditworthy clients to require additional assessment.

Comparison and Engineering Interpretation

The findings are consistent with previous studies showing that nonlinear models, appropriate feature processing, and class-imbalance treatment can improve credit-risk prediction. The present study demonstrates that a standalone SVM can provide technically meaningful results when standardization, kernel selection, hyperparameter optimization, and balanced evaluation metrics are applied systematically.

However, the selected model should be used as an initial screening or decision-support tool rather than as an automatic credit-rejection system. The confusion matrix showed that some default clients remained incorrectly classified as non-default, while several non-default clients were flagged as high risk. Therefore, predictions should be combined with

income verification, repayment-capacity analysis, supporting documents, and manual assessment.

Practical Implications and Limitations

For practical implementation, categorical encoding and numerical standardization must be integrated into a fixed prediction pipeline. Model performance should also be monitored periodically because changes in economic conditions, customer profiles, and repayment behavior may reduce predictive accuracy over time. The decision threshold may be adjusted according to the relative financial costs of false-negative and false-positive predictions.

This study was limited to one public credit card dataset and one main classification algorithm. The data represented Taiwanese clients during a specific historical period, so the findings may not directly generalize to other institutions, countries, or credit products. Future research should compare SVM with Logistic Regression, Random Forest, XGBoost, and ensemble models, while also evaluating probability calibration, fairness, explainability, and external validation using more recent financial data.

CONCLUSION

This study developed and evaluated a Support Vector Machine model for credit-risk classification using the UCI Default of Credit Card Clients dataset. The findings demonstrate that data preprocessing and class-imbalance treatment substantially influenced classification performance. The SVM trained using unstandardized variables failed to detect default clients, whereas numerical standardization enabled meaningful separation between the default and non-default classes. Among the evaluated kernels, the radial basis function kernel provided the most suitable nonlinear decision boundary. The selected balanced RBF-SVM achieved an accuracy of 77.13%, precision of 48.54%, recall of 56.22%, F1-score of 52.09%, balanced accuracy of 69.65%, and ROC–AUC of 75.10%. Although its overall accuracy was lower than that of the unbalanced configuration, it provided better detection of high-risk clients.

The scientific contribution of this study lies in demonstrating that SVM performance in credit-risk classification should not be assessed solely using overall accuracy. A controlled combination of numerical standardization, kernel selection, hyperparameter optimization, balanced class weighting, and class-sensitive evaluation metrics produced a

more technically meaningful model. The results also confirm that the relationship between financial characteristics and payment default is nonlinear and that F1-score, recall, balanced accuracy, and ROC–AUC are more appropriate than accuracy alone for evaluating imbalanced credit data. In practical application, the selected model is more suitable as an initial risk-screening and decision-support tool than as a fully automated credit-approval mechanism.

This study was limited to one historical public dataset, one primary machine-learning algorithm, and a predefined hyperparameter search space. The model was not externally validated using recent data from different countries, institutions, or credit products. Probability calibration, computational efficiency, model explainability, demographic fairness, and alternative imbalance-handling techniques were also not evaluated. Future research should compare SVM with Logistic Regression, Random Forest, XGBoost, and ensemble models under identical experimental conditions. External validation, cost-sensitive learning, probability calibration, fairness assessment, and explainable artificial intelligence should also be incorporated to improve the reliability and accountability of credit-risk prediction systems.

REFERENCES

Bao, W., Ning, L., & Yue, K. (2019). Integration of unsupervised and supervised machine learning algorithms for credit risk assessment. *Expert Systems with Applications*, *128*, 301-315. https://doi.org/10.1016/j.eswa.2019.02.033

Bhatore, S., Mohan, L., & Reddy, Y. R. (2020). Machine learning techniques for credit risk evaluation: A systematic literature review. *Journal of Banking and Financial Technology*, *4*(1), 111-138. https://doi.org/10.1007/s42786-020-00020-3

Bücker, M., Szepannek, G., Gosiewska, A., & Biecek, P. (2022). Transparency, auditability, and explainability of machine learning models in credit scoring. *Journal of the Operational Research Society*, *73*(1), 70-90. https://doi.org/10.1080/01605682.2021.1922098

Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable machine learning in credit risk management. *Computational Economics*, *57*(1), 203-216. https://doi.org/10.1007/s10614-020-10042-0

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, *408*, 189-215. https://doi.org/10.1016/j.neucom.2019.10.118

Chang, V., Sivakulasingam, S., Wang, H., Wong, S. T., Ganatra, M. A., & Luo, J. (2024). Credit risk prediction using machine learning and deep learning: A study on credit card customers. *Risks*, *12*(11), 174. https://doi.org/10.3390/risks12110174

Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing*, *91*, 106263. https://doi.org/10.1016/j.asoc.2020.106263

Dumitrescu, E. I., Hué, S., Hurlin, C., & Tokpavi, S. (2022). Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *European Journal of Operational Research*, *297*(3), 1178-1192. https://doi.org/10.1016/j.ejor.2021.06.053

Emmanuel, I., Sun, Y., & Wang, Z. (2024). A machine learning-based credit risk prediction engine system using a stacked classifier and a filter-based feature selection method. *Journal of Big Data*, *11*, 23. https://doi.org/10.1186/s40537-024-00882-0

Laborda, J., & Ryoo, S. (2021). Feature selection in a credit scoring model. *Mathematics*, *9*(7), 746. https://doi.org/10.3390/math9070746

Markov, A., Seleznyova, Z., & Lapshin, V. (2022). Credit scoring methods: Latest trends and points to consider. *The Journal of Finance and Data Science*, *8*, 180-201. https://doi.org/10.1016/j.jfds.2022.07.002

Moscato, V., Picariello, A., & Sperlì, G. (2021). A benchmark of machine learning approaches for credit score prediction. *Expert Systems with Applications*, *165*, 113986. https://doi.org/10.1016/j.eswa.2020.113986

Noriega, J. P., Rivera, L. A., & Herrera, J. A. (2023). Machine learning for credit risk prediction: A systematic literature review. *Data*, *8*(11), 169. https://doi.org/10.3390/data8110169

Suhadolnik, N., Ueyama, J., & Da Silva, S. (2023). Machine learning for enhanced credit risk assessment: An empirical approach. *Journal of Risk and Financial Management*, *16*(12), 496. https://doi.org/10.3390/jrfm16120496

Trivedi, S. K. (2020). A study on credit scoring modeling with different feature selection and machine learning approaches. *Technology in Society*, *63*, 101413. https://doi.org/10.1016/j.techsoc.2020.101413

Wang, T., & Li, J. (2019). An improved support vector machine and its application in P2P lending personal credit scoring. *IOP Conference Series: Materials Science and Engineering*, *490*(6), 062041. https://doi.org/10.1088/1757-899X/490/6/062041

Yeh, I. C. (2009). *Default of credit card clients* (UCI Machine Learning Repository. https://doi.org/10.24432/C55S3H