Journal of Theoretical and Applied Information Technology
15th December 2021. Vol.99. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

OPTIMIZATION OF PARAMETERS ON SUPPORT VECTOR MACHINES AT VARIOUS FACTOR LEVEL VALUES

ALIFYA AL ROHIMI[1], ADJI ACHMAD RINALDO FERNANDES[2], ENI SUMARMININGSIH[3]
[1]Student, University of Brawijaya, Department of Statistics, East Java, Indonesia
[2,3]Lecturer, University of Brawijaya, Department of Statistics, East Java, Indonesia
E-mail: [1]alifyaalrohimi@student.ub.ac.id, [2]fernandes@staff.ub.ac.id, [3]eni_stat@ub.ac.id

ABSTRACT

The purpose of this study is to optimize the parameters on Support Vector Machines (SVM) at various factor level values and compare the classification results. The data used in this study are secondary data from the 5C (Collateral, Character, Capacity, Condition of economy, and Capital) assessment at Bank X. Optimization of the parameters is carried out by determining each of the 5-factor level values for parameters C and γ. The determination of the 5 factor level values was carried out based on previous research, namely for the value of C = {0.5,0.75,1,10,100} and γ = {0.005,0.05,0.1,0.5,0.75}. The data is divided into two parts, namely training data and testing data with a ratio of 80:20. This comparison is based on the Pareto principle. This division is also done using 5-fold cross-validation. The kernel used in this study is the Kernel Radial Basis Function (RBF) because RBF can transform data into very high dimensions so that it can perform classification well. The conclusion that can be drawn based on this research is by using a kernel trick, especially using the RBF kernel, the results of parameter optimization are better. This is proven by the average level of accuracy using the RBF kernel with cross-validation using 5-fold reaching 90.38% while without the kernel trick it only reaches an average accuracy of 65%. Novelty in this study is the use of the 5C variable in the credit assessment at Bank X and the use of the 5 level factor value for parameter optimization of the SVM.

Keywords: *SVM, Radial Basis Function, Optimization, Factor Level Value*

1. INTRODUCTION

Support Vector Machine (SVM) is one of the superior methods of machine learning because it has good performance in solving classification and prediction cases. The principle of SVM is to find a classification model or a set of the optimal separator from classification data that is trained with an algorithm so that it can separate the data set into two or more different classes that can help predict categories from new data. The advantage of using SVM is that it can be analyzed theoretically using the concept of computational learning theory. The SVM method is proven to be a method that can improve the accuracy of classification results as found in previous studies which can achieve an accuracy of 98%. In machine learning, SVM are included in the supervised learning category with certain algorithms that analyze data for classification [1].

The development of the times demands that almost all lines of life work with data. Along with the development of technology, the data generated from every side of life is also increasing and incalculable. Data can be used to provide accurate information to the general public. With proper processing, data can be used as a source of information, risk assessment, forecasting, and can even be used for decision making. Statistics is the study of data processing. Starting from collecting data, presenting, processing, analyzing, drawing conclusions, to making decisions. One of the important roles of statistics is in the banking sector.

Banks offer several services to the public. One of them is credit. According to Law No. 10 of 1998, credit is the provision of money or equivalent claims, based on an agreement or loan agreement between a bank and another party that requires the borrower to repay his debt after a certain period with interest. The types of credit offered also vary. One type of credit is a Home Ownership Loan. According to [2], homeownership credit is one type of credit that aims to purchase a house, build a house, or renovate a house. Before providing homeownership loans to debtors, banks will consider the debtor's ability to fulfill their obligations to pay home mortgages. The problem that often arises in granting homeownership loans is the existence of debtors who have non-current homeownership loans. This can cause losses for the bank. Based on these problems, banks need to consider and supervise the implementation of mortgage loans.

Before giving credit to prospective debtors, the bank will carry out stages to determine the decision to grant credit to prospective debtors. This stage is known as credit scoring. Credit scoring is one of the main methods for developing credit risk assessment tools. Credit scoring is a method to evaluate the credit risk of prospective debtors with their credit scores obtained from the credit scoring model [3]. The aim is to classify prospective debtors based on predetermined variables. Statistical analysis that can be used in this problem is Support Vector Machines (SVM).

Although this method has advantages in terms of accuracy, these advantages also depend heavily on selecting the optimal parameters from the SVM parameters, namely C (cost) and γ (gamma). Therefore, the selection of parameter values is the focus of the problem in this study. The technique of selecting parameter values using a trial and error approach is not possible because there are so many combinations of values that can be used and even have infinite values so that an optimization technique is needed in selecting the parameter values which does not require too many experiments and requires a relatively short time.

This research will compare the parameter optimization with and without kernel trick. The kernel used in this study is the Radial Basis Function (RBF) kernel. SVM have high classification skills because they can learn data well from training data [4]. The selection of the RBF kernel is based on the consideration that the RBF kernel will transform data into a higher dimension so that it is expected to be able to perform better classification on homeownership credit Bank X customers.

2. LITERATURE REVIEW

Machine learning is a study that studies computational algorithms for several purposes, such as filtering, classifying, or detecting images or videos. Broadly speaking, machine learning is divided into three categories, namely unsupervised learning, supervised learning, and reinforcement learning. The main purpose of unsupervised learning is to group objects that are considered similar in a certain space or area, for example using K-Nearest Neighbor. While supervised learning has a goal to classify objects based on the characteristics inherent in these objects. Reinforcement learning itself is one of the techniques in machine learning where the algorithm learns something by taking a certain action and seeing the results of that action (learning based on previous experience). In machine learning, SVM fall into the category of supervised learning with certain algorithms that analyze data for classification [5].

2.1 Support Vector Machine

Support Vector Machines (SVM) are linear and non-linear data classification methods. SVM were first introduced by Vladimir Vapnik and two of his friends, Bernhard Boser and Isabelle Guyon. SVM are included in supervised learning. SVM work with algorithms that function to transform data into higher dimensions. In the new dimension, the SVM algorithm looks for the optimal linear separator hyperplane. For example, a decision boundary that separates one observation from another. With a suitable non-linear mapping to a sufficiently high dimension, data from the two classes can always be separated by a hyperplane. SVM define this hyperplane using the support vector classifier and margin (which is defined from the support vector classifier) [6]. SVM can work to search for linear and non-linear hyperplane.

SVM strive to find the optimal separator function (hyperplane) to separate the data set into 2 parts. Let D be a data set consisting of $(\mathbf{X}_1, y_1), (\mathbf{X}_2, y_2), ..., (\mathbf{X}_{|D|}, y_{|D|})$, where $\mathbf{X}_i$ is a set of independent variables with the response variable y_i hich consists of two categories, for example, +1 and -1. SVM aim to find the maximum marginal hyperplane (MMH). Margin is defined as the shortest distance between the data and the hyperplane. When the margin is bigger, the better the hyperplane is to separate a data set.

2.2 Hyperplane

The hyperplane is defined as follows:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0 \qquad (1)$$

where:

$\beta_0, \beta_1, \beta_2$ is the parameter to be predicted.

When expanded into p dimensions, the hyperplane is defined as follows:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p = 0 \qquad (2)$$

Suppose there is a data matrix **X** where n is the number of samples and *p* is the number of variables, then

$$\mathbf{x_1} = \begin{pmatrix} x_{11} \\ \vdots \\ x_{1p} \end{pmatrix}, ..., \mathbf{x_n} = \begin{pmatrix} x_{n1} \\ \vdots \\ x_{np} \end{pmatrix}, \qquad (3)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

where observation is divided into two classes, $y_i,...,y_n \in \{-1,1\}$ where -1 represents the first class (in this study the credit is not current) and 1 represents the other classes (current credit).

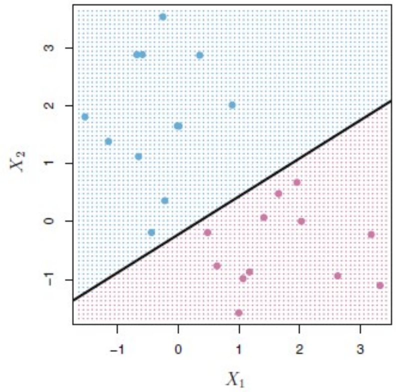


Figure 1. Separation of 2 Types of Class [7]

Figure 1 is an illustration of the observation which consists of two classes. Suppose the class in blue is the class with observations $y_i = 1$ while the purple one $y_i = -1$ so that the hyperplane is defined as follows.

$$\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_p X_{ip} > 0, y_i = 1 \quad (4)$$

and

$$\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_p X_{ip} < 0, y_i = -1 \quad (5)$$

The two inequalities can be written as

$$y_i(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_p X_{ip}) > 0, i = 1,...,n \quad (6)$$

2.3 Support Vector Machine

In general, if the data can be separated perfectly by a hyperplane, then there are many possible hyperplane options for separating the data. To build the best hyperplane, there must be a reason to choose 1 among the many hyperplanes available. The best hyperplane is the hyperplane that has the largest margin so that it can separate the two classes very well, this hyperplane is called the Maximal Margin Hyperplane (MMH).

MMH can be formed through optimization with the following constraints.

$$\max_{\beta_0,\beta_1,...,\beta_p,M} M$$

depend on $\sum_{j=1}^{p} \beta_j^2 = 1$,

$$y_i(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_p X_{ip}) \geq M, \forall i = 1,...,n \quad (7)$$

where M is a positive number.

The optimization with these constraints is simpler than it seems. The constraint in equation (7) guarantees that each object will be on the right side of the hyperplane in classifying objects, where the value of M is positive. For each object to be on the right side of the hyperplane we need constraints $y_i(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_p X_{ip}) > 0$. Parameter estimation is done by forming the inequality for each object. Then each of the inequalities that are formed is carried out by a substitution or elimination process to obtain the respective estimator values β.

2.4 Support Vector Classifier

Support Vector Classifier will classify objects depending on which side of the object is from the hyperplane where the hyperplane is selected to properly separate each object into two classes, but some objects may be classified as inappropriate. Support Vector Classifier is an object that has the closest distance to the hyperplane. The Support Vector Classifier can be determined through optimization with the following constraints.

$$\max_{\beta_0,\beta_1,...,\beta_p,\epsilon_1,...,\epsilon_n,M} M$$

depend on $\sum_{j-1}^{p} \beta_j^2 = 1$,

$$y_i(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_p X_{ip}) \geq M(1 - \epsilon_i), \quad (8)$$

$$\epsilon_i \geq 0, \sum_{i=1}^{n} \epsilon_i \leq C$$

where

C : non negative parameter.

M : margin distance

$\epsilon_1, ..., \epsilon_n$: slack variable

2.5 Kernel Function

SVM are an extension of the support vector classifier that results from enlarging the feature space in some way, using the kernel. The solution to the support vector classifier problem involves only the inner product of the observation (as opposed to the observation itself). The inner product of the two r-vectors a and b of the two observations is defined as $\langle a, b \rangle = \sum_{i=1}^{r} a_i b_i$, so that the inner product of the two observations x_i, x_i' defined as follows.

$$\langle x_i, x_i' \rangle = \sum_{j=1}^{p} x_{ij} x_{ij}' \quad (9)$$

so that the linear support vector classifier can be written as follows

$$f(x) = \beta_0 + \sum_{i=1}^{n} \alpha_i \langle x, x_i \rangle \quad (10)$$

where there are n parameters α_i, $i = 1, ..., n$.

To estimate parameters $\alpha_1, ..., \alpha_n$ and β_0, needed $\binom{n}{2}$ inner product $\langle x_i, x_i' \rangle$ among all observations so that a generalization of the inner product is needed in the following form.

$$K(x_i, x_i') \tag{11}$$

where K is a function of the Kernel. The purpose of the Kernel function is to calculate the similarity of two observations. There are several Kernel functions, the first is linear kernel. Linear kernel basically measures the similarity of a pair of observations using the Pearson correlation. The linear kernel is defined as follows.

$$K(x_i, x_i') = \sum_{j=1}^{p} x_{ij} x_{i'j} \tag{12}$$

so $\mathbf{f(x)} = \boldsymbol{\beta}_0 + \sum_{i=1}^{n} \alpha_i \langle \mathbf{x}, \mathbf{x_i} \rangle$, becomes

$$f(x) = \beta_0 + \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{p} x_{ij} x_{i'j}$$

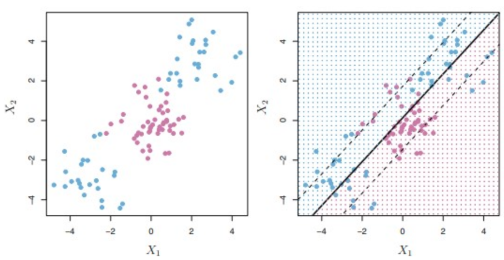


Figure 2. Draw Objects with Hyperplane[7]

The second kernel function is the polynomial kernel which is defined as follows.

$$K(x_i, x_i') = \left(1 + \sum_{j=1}^{p} x_{ij} x_{i'j}\right)^d \tag{14}$$

where d is the degree of the polynomial with conditions d > 1. This kernel function will transform the data to a higher dimension so that the optimal hyperplane can be determined. These dimensions correspond to the degrees used in kernel functions.

When a support vector classifier is combined with a non-linear kernel such as a polynomial kernel, the result is referred to as Support Vector Machines. So that $\mathbf{f(x)} = \boldsymbol{\beta}_0 + \sum_{i=1}^{n} \alpha_i \langle \mathbf{x}, \mathbf{x_i} \rangle$, becomes

$$f(x) = \beta_0 + \sum_{i=1}^{n} \alpha_i \left(1 + \sum_{j=1}^{p} x_{ij} x_{i'j}\right)^d \tag{15}$$

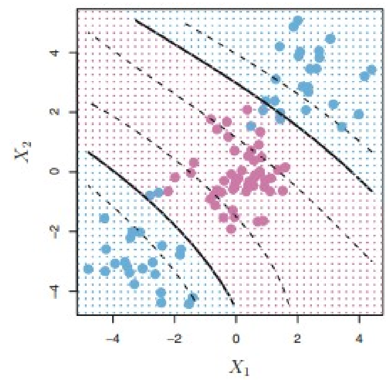


Figure 3. SVM with kernel polynomials [7]

The third kernel function is the kernel radial basis function (RBF) which is defined as follows.

$$K(x_i, x_i') = \exp\left(-\gamma \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2\right) \tag{16}$$

where γ is a positive constant. The RBF kernel works by transforming into an infinite dimensional space such that $\mathbf{f(x)} = \boldsymbol{\beta}_0 + \sum_{i=1}^{n} \alpha_i \langle \mathbf{x}, \mathbf{x_i} \rangle$, becomes

$$(x) = \beta_0 + \sum_{i=1}^{n} \alpha_i \exp\left(-\gamma \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2\right) \tag{17}$$

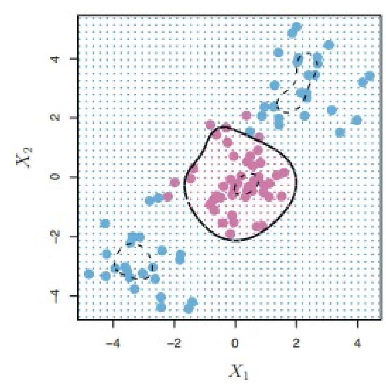


Figure 4. SVM with RBF kernel [7]

2.6 Support Vector Machine

K-folds Cross Validation is a technique for validation that is very popular in use. The k-folds validation method is very suitable for data cases where the number of samples is limited. To carry out the classification process, of course, the data is divided into training and testing, and when the data used for training is very little, the data used may be less representative. In k-folds cross-validation, data (D) is divided into k subsets $D_1, D_2, \ldots, D_k$ the same amount. The data used for training is data subsets k-1 which are combined and then applied to the remaining one data subsets as a result of testing. This process is repeated as many as k subsets and the results of classification accuracy are the average

Journal of Theoretical and Applied Information Technology
15th December 2021. Vol.99. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

results of each training and testing data. The k-folds that are commonly used are 3, 5, 10 and 20 [8].

2.7 Classification Method Performance Evaluation

Performance can be seen based on three criteria, namely accuracy, sensitivity, and specificity. Accuracy measures how correctly a diagnostic test identifies and excludes a certain condition, in other words, accuracy is used to measure the goodness of the model. In diagnostic tests, the terms sensitivity and specificity are also known. Sensitivity and specificity in diagnostic tests is a measure of the ability to correctly identify objects under reality [9]. The difference is that sensitivity measures the positive group while specificity measures the negative group. For example, in the case of credit collectibility, there are current and non-current loans. The current credit is referred to as the positive group and non-current credit is referred to as the negative group. The sensitivity in this example is the ability to correctly diagnose debtors with current and indeed current credit collectibility. Meanwhile, specificity is the ability to diagnose debtors with non-current and non-current credit collectibility. To get the value of accuracy, sensitivity, and specificity can use the Confusion Matrix in Table 2.1.

Table 1. Confusion Matrix

Actual	Prediction	
	Positive	Negative
Positive	*TP*	*FN*
Negative	*FP*	*TN*

TP : True Positive (number of correct predictions in positive class)
FP : False Positive (number of false predictions in positive class)
FN : False Negative (number of false predictions in negative class)
TN : True Negative (number of correct predictions in negative class)
Based on Table 2.1 the calculation of accuracy can be done with the following formula.

$$accuracy = \frac{TN + TP}{TN + TP + FN + FP}$$

To get an optimal and more specific classification, sensitivity and specificity can be tested. Sensitivity is a true positive level or a performance measure to measure a positive class while specificity is a true negative level or a performance measure to measure a negative class. The formula for sensitivity and specificity is as follows.

$$Sensitivity = \frac{TP}{(TP + FN)} \times 100\%$$

$$Specificity = \frac{TN}{(TN + FP)} \times 100\%$$

2.8 Credit collectibility and 5C Assesment

According to Law No. 10 of 1998, credit is the provision of money or equivalent claims, based on an agreement or loan agreement between a bank and another party that requires the borrower to repay his debt after a certain period with interest. The risk of lending the least favored by banks is when there are non-performing loans that occur due to the debtor's negligence in fulfilling their obligations to pay credit installments, both principal and interest [10].

Credit collectability is the condition of payment of principal or installments of principal and interest on credit by customers and the level of possibility of receiving back funds invested in securities or other investments. The accuracy of payment of principal and interest is classified into two, namely:

1) Performing Loans, namely on-time payment conditions, good account development, and no arrears and according to credit terms.
2) Non-Performing Loan, namely the condition of arrears of principal/interest of more than 90 days.

The predictor variable used in this study is the 5C assessment (Collateral, Character, Capacity, Condition of economy, and Capital) in a bank with the following explanation.

a. Collateral

Based on Law No. 10 of 1998 concerning Banking, banks are prohibited from providing credit without sufficient guarantees or collateral. Therefore, collateral is an important part to ensure that there will be no bank losses if the debtor fails in running his business or does not make credit payments as agreed. In this study, the indicators used are security documents and length of stay.

b. Character

In analyzing the character, it is closely related to the integrity of the prospective debtor because this integrity greatly determines the willingness to pay or return the credit along with the interest on the credit facilities obtained. Assessment of good faith to repay credit is indeed a bit difficult to predict. However, this can be done by seeking input through interbank information, either through Bank Indonesia as the central bank or through local commercial banks. This study uses several indicators to measure character, namely the size of the city, education, age, collectability status, and marital status.

Journal of Theoretical and Applied Information Technology
15th December 2021. Vol.99. No 23
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

c. Capacity

Capacity is an assessment of the ability of the prospective debtor regarding his ability to fulfill the obligations that will be stipulated in the credit agreement. Some things that must be considered in measuring the creditability of prospective debtors are as follows.

- Ability to provide funds for objects financed with bank credit.
- The ability to carry out the project according to the schedule that has been planned and determined.
- The ability to produce, market, and profit from their products.

In this study, several indicators were used to measure capacity, namely the form of business entity, credit period, Installment Income Ratio, occupation, work experience, and the number of dependents in the family.

d. Condition of Economy

Economic condition is an assessment of economic indicators that affect the business of prospective debtors and projects that will be financed with bank loans. The assessment includes:

- The current condition of the prospective debtor's business sector and its prospects.
- Provision of raw materials and the extent of dependence on imported raw materials.
- Government regulations governing the business of prospective debtors (if any).
- National and global economic conditions that support businesses or projects financed by banks.

In this study, the indicator used to assess the condition of the economy is the ownership of savings.

e. Capital

Prospective debtors must have a certain amount of money as business capital in carrying out their business and lack capital based on certain ratios in accordance with the bank's calculation policy. This lack of capital is financed by credit provided by the bank. Data regarding the capital of the prospective debtor is known through the financial statements submitted to the bank. In this study, the capital indicator used is Loan to Value.

3. RESEARCH METHODS

The data used in this study are secondary data from the 5C (Collateral, Character, Capacity, Condition of economy, and Capital) assessment at Bank X. The total data is 100 customers. Parameter optimization is done by determining each of the 5 factor level values for parameters C and γ. The determination of the 5 factor level values was carried out based on previous research, namely for the value C={0.5,0.75,1,10,100} and γ={0.005,0.05,0.1,0.5,0.75}. This determination is based on research conducted by Huang, Hung, Lee, Li, and Jiang [11] using C={10,50,100} and γ={2.4,5,10} then in research conducted by Erfanifard, Behnia and Moosavi [12] using C={100,200,300} and γ={0.2,0.3,0.4}. In machine learning, data is divided into two parts, namely training data and testing data. Training data is used to build or train a model / classifier. While the testing data is used to test the classifier that has been built to see how accurate the classification results are. Based on the Pareto principle, data sharing is done in a ratio of 80:20. As much as 80% of the data is training data and the remaining 20% is used as testing data [12]. This comparison is based on the Pareto principle. This division is also done using 5-fold cross validation. The kernel used in this study is the Kernel Radial Basis Function (RBF) because RBF can transform data into very high dimensions so that it can perform classification well.

4. RESULTS AND DISCUSSION

Begin with a simple lemma that bounds the square difference beetween the plus function $(x)_+$ and its smooth approximation $p(x,\alpha)$.

Lemma 1.

$$for\ x\in R\ and\ |x|<\rho: p(x,\alpha)^2-(x_+)^2 \le \left(\frac{\log 2}{\alpha}\right)+\frac{2\rho}{\alpha}\log 2$$

where $p(x,\alpha)$ is the p function of $p(x,\alpha)=x+\frac{1}{\alpha}\log(1+\varepsilon^{-\alpha})$ with smoothing parameter $\alpha>0$.

Proof: we consider two cases. For $0<x<\rho$,

$$p(x,\alpha)^2-(x_+)^2=\frac{1}{\alpha^2}\log^2(1+\varepsilon^{-\alpha x})+\frac{2x}{\alpha}\log(1+\varepsilon^{-\alpha x})$$

$$\le\left(\frac{\log 2}{\alpha}\right)^2+\frac{2\rho}{\alpha}\log 2$$

For $-\rho<x\le 0, p(x,\alpha)^2$ is a mnotonically increasing function, so we have

$$p(x,\alpha)^2-(x_+)^2=p(x,\alpha)^2\le p(0,\alpha)^2=\left(\frac{\log 2}{\alpha}\right)^2$$

Hence, $p(x,\alpha)^2-(x_+)^2\le\left(\frac{\log 2}{\alpha}\right)+\frac{2\rho}{\alpha}\log 2$.

According to Lemma 1. It can be proved that RBF Kernel Based SVM is able to approximate any continuous function on compact set with arbitrary accuracy, and this conclusion can be deduced to discrete function.

Theorem: For any continuous real functions g defined on a compact set

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$U \in R^n$ and any $\in > 0$, there exists a RBF kernel based SVM f formed by $y(x)=\sum_{i=1}^{n}\alpha_i k(x,x_i)+b$ verifies: $\sup_{x\in U}|f(x)-\mathrm{g}(x)|<\varepsilon$.

Lemma 2. Suppose Z is a set of continuous real functions on compact set U, if it satisfies the conditions below, then the universal closure of Z includes all of continuous funcction on U, namely, $(\mathrm{Z},\mathrm{d}_\infty)$ on $(C[\mathrm{U}],\mathrm{d}_\infty)$ is compact.

(1) Z is an algebra, that is, set Z is closed to addition, multiplication and scalar multiplication;

(2) Z can isolate each point on U, that is, to every $x,y \in > U$, if $x \neq y$, there $f \in Z$ which make $f(x)\neq f(y)$;

(3) Z is not zero at any point on U, to every $x,y \in > U$, there must exist $f \in Z$ which make $f(x)\neq 0$.

Proof: to prove the theorem, we should first prove that Y satisfies the three conditions of the lemma, $y(x)=\sum_{i=1}^{n}\alpha_i k(x,x_i)+b$ can be written as

$$y(x)=f(x)=\sum_{i=1}^{n}\alpha_i k(x,x_i)+b=\sum_{i=1}^{n}\alpha_i \exp\left(-\|x-x_i\|^2/2\sigma^2\right)$$

Where $\alpha_{i+1}=b, x_{i+1}=x$

(Y,d_∞) is an algebra.

Suppose $f_1,f_2\in Y$, then f_1,f_2 can be written as:

$$f_1(x)=\sum_{i_1=1}^{n_1=1}\alpha_{i1}k(x,x_{i1}), f_2(x)=\sum_{i_2=1}^{n_2=1}\alpha_{i2}k(x,x_{i2})$$

$$f_1(x)+f_2(x)=\sum_{i_1=1}^{n_1=1}\alpha_{i1}k(x,x_{i1})+\sum_{i_2=1}^{n_2=1}\alpha_{i2}k(x,x_{i2})=\sum_{i=1}^{n_1+n_2+2}\alpha_i k(x,x_i)\in Y$$

$$f_1(x)f_2(x)=\left(\sum_{i_1=1}^{n_1=1}\alpha_{i1}k(x,x_{i1})\right)\left(\sum_{i_2=1}^{n_2=1}\alpha_{i2}k(x,x_{i2})\right)=\sum_{i_1=1}^{n_1=1}\sum_{i_2=1}^{n_2=1}\alpha_{i1}\alpha_{i2}k(x,x_{i1})k(x,x_{i2})$$

$$=\sum_{i_1=1}^{n_1=1}\sum_{i_2=1}^{n_2=1}(\alpha_{i1}\alpha_{i2})\exp\left(-\left(\|x-x_{i1}\|^2+\|x-x_{i2}\|^2\right)/2\sigma^2\right)\in Y$$

$$cf_1(x)=\sum_{i_1=1}^{n_1=1}(c\alpha_{i1})k(x,x_{i1})\in Y$$

So (Y,d_∞) is an algebra and $y(x)=\sum_{i=1}^{n}\alpha_i k(x,x_i)+b$.

Before entering the main analysis, descriptive statistics were calculated used to find out the general picture of the data. The data used in this study is secondary data of 5C assessment with ordinal and nominal data scales. Descriptive statistics for ordinal data are presented in Table 4.1.

Table 2. Descriptive statistics

Variable	Minimum	Mean	Maximum
X_{12}: Length of Residence (Years)	0.50	8.86	20.00
X_{21}: City Size (Scale 1-15)	2.00	6.98	15.00
X_{22}: Education (Years)	12.00	15.53	16.00
X_{23}: Age (Years)	26.00	40.24	54.00
X_{33}: Credit Term (Years)	2.00	11.49	25.00
X_{34}: Installment Income Ratio	0.13	2.19	4.92
X_{36}: Work Experience (Months)	14.00	74.61	329.00
X_{37}: Number of Family Dependents (person)	0.00	1.40	6.00
X_5: Loan To Value	53.28	82.41	98.40

Based on Table 2, the shortest customer in their current residence is 6 years while the longest is 20 years with most of the customers occupying their current place of residence for 8.86 years. In addition, the customer's residence is also taken into account in this study. Where there are customers who live in areas with the narrowest city size, which is a scale of 2 and the widest is a scale of 15 with most of the customers residing in cities with an area of 6.98.

Most of the customers studied for 15.53 years with the lowest education being 12 years while the longest was 16 years. Bank X's customers consist of various ages. The youngest customer is 26 years old while the oldest is 54 years old with most of the customers being 40.24 years old. In addition, the credit period for each customer also varies, the shortest term for customer credit is 2 years while the longest is 25 years with most customers taking credit terms of 11.49 years.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

On the other hand, the installment income ratio for each customer is also different. The smallest is 0.13 while the largest is 4.92 with the most installment income ratio of 2.19. Customers also come from different work backgrounds. The minimum customer work experience is 14 months while the longest is 329 months with most customers having work experience of 74.61 months. The number of dependents for each customer also varies, ranging from those with no dependents to a maximum of 6 dependents with an average number of dependents of 1 to 2 people.

The financial term used to describe the size of the loan compared to the value of the property used as collateral is Loan to Value (LTV). In Bank X customer data, the lowest LTV is 53.28 and the highest is 98.40. Of the total customers, most customers have an LTV of 82.41.

The following is the calculation result in determining the optimal parameters for homeownership credit Bank X prospective customer data either without using a kernel trick or using the RBF kernel as shown in Table 1 and Table 2.

Table 3. Parameter optimization without kernel trick

No.	C	γ	SVM Accuracy Value (%)					Average Accuracy (%)
			Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	
1.	0.5	0.005	75	66.67	61.54	50	69.23	64.49
2.	0.5	0.05	58.33	83.33	84.62	50	46.15	64.49
3.	0.5	0.1	75	66.67	46.15	66.67	69.23	64.74
4.	0.5	0,5	75	66.67	53.85	66.67	61.54	64.74
5.	0.5	0.75	91.67	66.67	69.23	33.33	61.54	64.49
6.	0.75	0.005	83.33	58.33	69.23	66.67	46.15	64.74
7.	0.75	0.05	75	66.67	69.23	66.67	46.15	64.74
8.	0.75	0.1	58.33	100	53.85	50	61.54	64.74
9.	0.75	0,5	75	75	38.46	58.33	76.92	64.74
10.	0.75	0.75	66.67	58.33	69.23	58.33	69.23	64.36
11.	1	0.005	66.67	91.67	38.46	58.33	69.23	64.87
12.	1	0.05	83.33	66.67	69.23	50	53.85	64.62
13.	1	0.1	66.67	58.33	84.62	50	61.54	64.23
14.	1	0,5	50	66.67	69.23	58.33	76.92	64.23
15.	1	0.75	58.33	66.67	61.54	66.67	69.23	64.49
16.	10	0.005	83.33	58.33	76.92	58.33	46.15	64.62
17.	10	0.05	75	50	61.54	75	61.54	64.62
18.	10	0.1	50	75	61.54	66.67	69.23	64.49
19.	10	0,5	66.67	50	69.23	58.33	76.92	64.23
20.	10	0.75	91.67	75	53.85	41.67	61.54	64.74
21.	100	0.005	66.67	66.67	69.23	66.67	53.85	64.62
22.	100	0.05	58.33	66.67	84.62	58.33	53.85	64.36
23.	100	0.1	25	66.67	84.62	66.67	76.92	63.97
24.	**100**	**0,5**	**66.67**	**75**	**53.85**	**83.33**	**46.15**	**65**
25.	100	0.75	75	58.33	84.62	58.33	46.15	64.49

Table 3 shows the quality of the parameter optimization results. It can be seen that according to the results of the optimization of various factor levels in the data for homeownership credit Bank X customers, especially without using a kernel trick, the optimal parameter is shown by the combination of the value of C = 100 and γ = 0.5 with an average value of accuracy of 65%. When compared with using the RBF kernel, the results are as shown in Table 2 below.

Table 4. Optimization of parameters using the RBF kernel

No.	C	γ	SVM Accuracy Value (%)					Average Accuracy (%)
			Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	
1.	0.5	0.005	75.00	66.67	84.62	66.67	46.15	67.82
2.	0.5	0.05	100.00	83.33	92.31	83.33	76.92	87.18
3.	0.5	0.1	66.67	50.00	69.23	58.33	76.92	64.23
4.	0.5	0,5	83.33	75.00	61.54	50.00	76.92	69.36
5.	0.5	0.75	83.33	91.67	69.23	83.33	76.92	80.90
6.	0.75	0.005	50.00	66.67	69.23	58.33	76.92	64.23
7.	0.75	0.05	83.33	83.33	76.92	100.00	92.31	87.18
8.	0.75	0.1	83.33	58.33	84.62	58.33	46.15	66.15
9.	0.75	0,5	83.33	58.33	92.31	50.00	69.23	70.64
10.	0.75	0.75	91.67	83.33	84.62	91.67	84.62	87.18

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

No.	C	γ	SVM Accuracy Value (%)					Average Accuracy (%)
			Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	
11.	1	0.005	100.00	66.67	69.23	33.33	69.23	67.69
12.	1	0.05	58.33	66.67	84.62	58.33	53.85	64.36
13.	1	0.1	100.00	83.33	84.62	83.33	84.62	87.18
14.	1	0,5	75.00	83.33	84.62	91.67	92.31	85.38
15.	1	0.75	91.67	58.33	92.31	83.33	92.31	83.59
16.	10	0.005	83.33	75.00	84.62	83.33	100.00	85.26
17.	10	0.05	75.00	66.67	53.85	66.67	61.54	64.74
18.	10	0.1	83.33	75.00	69.23	50.00	69.23	69.36
19.	**10**	**0,5**	**91.67**	**100.00**	**92.31**	**83.33**	**84.62**	**90.38**
20.	10	0.75	100.00	83.33	84.62	100.00	76.92	88.97
21.	100	0.005	58.33	83.33	84.62	50.00	46.15	64.49
22.	100	0.05	75.00	58.33	61.54	66.67	69.23	66.15
23.	100	0.1	91.67	83.33	92.31	66.67	84.62	83.72
24.	100	0,5	91.67	50.00	76.92	75.00	84.62	75.64
25.	100	0.75	83.33	66.67	69.23	75.00	46.15	68.08

Table 4 shows that the optimal value using the criteria for various level values is indicated by the combination of the value of C = 10 and γ = 0.5 with an average accuracy of 90.38%. It can be seen that there is a significant increase in the resulting accuracy value after using the RBF Kernel. The following is a visualization to describe the accuracy of each fold from the combination of each level for each factor.

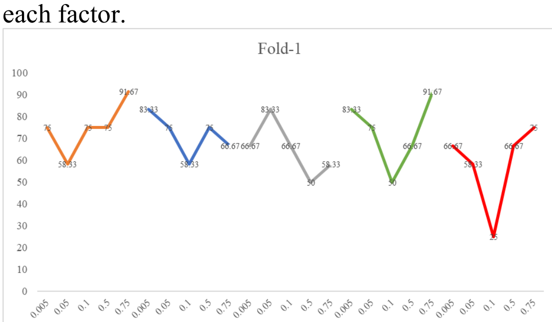
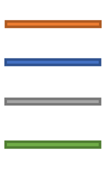


Figure 5. Accuracy Results on Fold-1

where:



: C = 0.5

: C = 0.75

: C = 1

: C = 10

: C = 100

Figure 5 shows the results of the accuracy in fold-1 that the value of C = 0.5 and C = 10 gives a high contribution in improving the relationship because it produces the maximum value of each in the experiment γ = 0.75 while the lowest is the experiment at C = γ = 1.

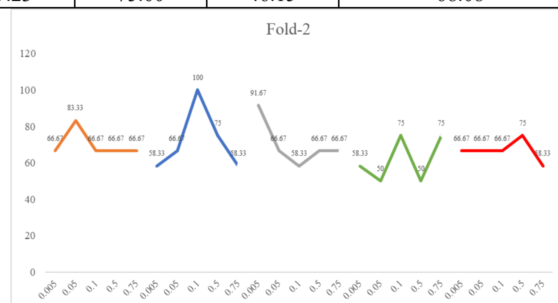


Figure 6. Accuracy Results on Fold-2

Figure 6 shows the results of the accuracy in fold-2 that the maximum accuracy is 100% when C = 0.75 with γ = 0.1 while the lowest accuracy value is 50% when C = 10 with γ = {0.05,0.5}.

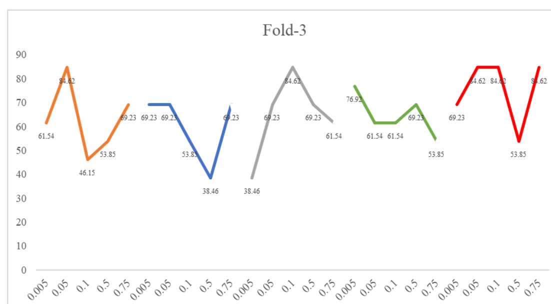


Figure 7. Accuracy Results on Fold-3

Figure 7 shows the results of the accuracy on fold-3 that the maximum accuracy value is 84.62% which is found when C = 0.5 with γ = 0.05 and C = 100 with γ = {0.05,0.1}. While the lowest accuracy value is at C = 0.75 and C = 1 with γ values respectively 0.5 and 0.005.

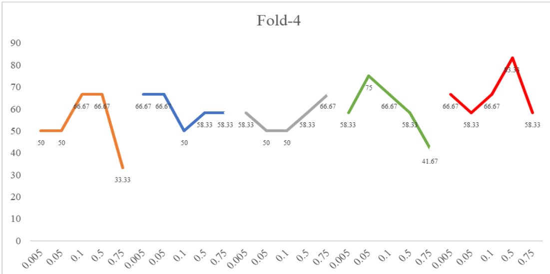


Figure 8. Accuracy Results on Fold-4

Figure 8 shows the results of the accuracy on the fold-4 that the highest accuracy value is 83.33% when C = 100 and γ = 0.5 while the lowest accuracy value is when C = 0.5 with γ = 0.75 at 33.33%.

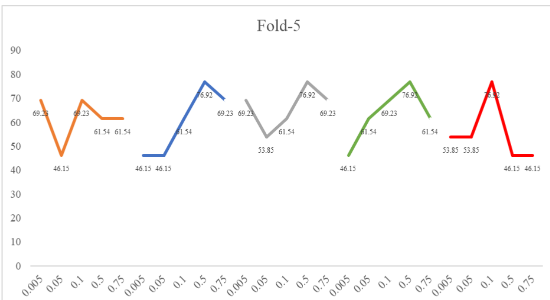


Figure 9. Accuracy Results on Fold-5

Figure 9 shows the results of the accuracy of the fold-5 that the highest accuracy value is 76.92% at C = {0.75,1,10,100} with different dengan values. While the lowest accuracy value is 64.15%.

5. CONCLUSION AND SUGGESTION

The conclusion that can be drawn based on this research is by using a kernel trick, especially using the RBF kernel, the results of parameter optimization are better. This is proven by the average level of accuracy using the RBF kernel with cross-validation using 5-fold reaching 90.38% while without the kernel trick it only reaches an average accuracy of 65%. Suggestions for further research are expected to use other factor level values or at different intervals in determining the optimal parameter location in the SVM.

REFERENCES:

[1] Pantouw, E. 2008. *230+ Sumber Pinjaman untuk Usaha Anda*. Yogyakarta: Gradien Mediatama.

[2] Yu, L., Wang, S., Lai, K.K., dan Zhou, L. 2008. *Bio-Inspired Credit Risk Analysis.* Heidelberg: Springer.

[3] Budianto, A., Ariyuana, R., & Maryono, D. “The closest C-neighboring comparison (KNN) and supports vector machines (SVM) in the introduction of motorized vehicle plate characters”. Scientific Journal of Engineering and Vocational Education, (11/1), 2018. 27-35.

[4] Mitchell, T. “Machine Learning”. New York: McGraw Hill. 1997.

[5] Han, J., Kamber, M., & Pei, J. “DATA MINING Concepts and Techniques”. Waltham: Elsevier. 2012.

[6] James, G., Witten, D., Hastie, T., & Tibshirani, R. “An Introduction to Statistical Learning with Application in R”. New York: Springer. 2017.

[7] Bolon, V., Sanchez, N., & Alonso, A. “Feature Selection for High-Dimensional Data. Artificial Intelligence: Foundation, Theory, and Algorithms”. A Coruna, Spain: Springer International Publishing Switzerland. 2015.

[8] Wong, H.B. dan Lim, G.H., 2011. Measures of diagnostic accuracy: sensitivity, specificity, PPV and NPV. *Proceedings of Singapore healthcare*, *20*(4), pp.316-318.

[9] Leon, B. dan Ericson, S., 2007. *Manj Aktiva Pasiva Bank Nondevisa*. Grasindo.

[10] Huang, M.-L., Hung, Y.-H., Lee, W. M., Li, R. K., & Jiang, B.-R. “SVMRFE Based Feature Selection and Taguchi Parameters Optimization for Multiclass SVM Classifier”. The Scientific World Journal, 2014. 1-10.

[11] Erfanifard, Y., Behnia, N., & Moosavi, V. Tree crown delineation on UltraCam-D aerial imagery with SVM classification technique optimised by Taguchi method in Zagros woodlands. International Journal of Image and Data Fusion, (5/4), 2014. 300-314.

[12] Loelianto, I., Thayf, M. S. S., & Angriani, H. 2020. Implementasi Teori Naive Bayes Dalam Klasifikasi Calon Mahasiswa Baru STMIK Kharisma Makassar. SINTECH (Science and Information Technology) Journal, 3(2), 110-117.