

CLASSIFICATION OF STUNTING IN CHILDREN UNDER FIVE YEARS IN PADANG CITY USING SUPPORT VECTOR MACHINE

Izzati Rahmi^{1*}, Mega Susanti², Hazmira Yozza³, FriLianda Wulandari⁴

^{1,2,3,4} Department of Mathematics, Andalas University
Limau Manis, Padang, 25163, Indonesia

Corresponding author's e-mail: ^{1*} izzatirahmihg@sci.unand.ac.id

Abstract. Stunting is a nutritional problem in children characterized by a child's height that is less than twice the standard deviation of the median standard of children growth that has been determined by the WHO. Stunting is influenced by many factors. If the conditional of these factors were known, it could be expected to know whether a child is stunted or not. In this study, the prediction of stunting was carried out using the Support Vector Machine (SVM) classification method. SVM is a method to find the best hyperplane that can be used to separate two or more classes. In this study, the parameters of the SVM model that must be determined are the cost value and gamma. Based on the research results with parameters cost = 10 and gamma = 5, a classification estimation result with 100% accuracy can be obtained.

Keywords: Stunting, Support Vector Machine, Hyperplane.

Article info:

Submitted: 28th February 2022

Accepted: 3rd July 2022

How to cite this article:

I. Rahmi, M. Susanti, H. Yozza and F. Wulandari, "CLASSIFICATION OF STUNTING IN CHILDREN UNDER FIVE YEARS IN PADANG CITY USING SUPPORT VECTOR MACHINE", *BAREKENG: J. Math. & App.*, vol. 16, iss. 3, pp. 771-778, September, 2022.



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).
Copyright © 2022 Author(s)

1. INTRODUCTION

Children under five years of age are the most vulnerable group for having malnutrition problems [1]. One form of malnutrition is stunting. Stunting is a condition where toddlers have a length or height that is less than twice the standard deviation of the median growth standard defined by WHO. According to the result of monitoring nutritional status by the Ministry of Health office of the Republic of Indonesia, there has been a stunting problem in Padang, West Sumatera, Indonesia. In 2015, 21.1% of children under the age of five were stunted, and this figure increased to 22.6% in 2016 [2], [3]. However, this figure is still greater than WHO's standard, where a region is categorized to have no nutritional problems if the rate of stunting is less than 20% [4].

Stunting is a nutritional problem that is influenced by several factors. These factors include mothers and mothers-to-be, the condition of infants and toddlers, and the socioeconomic conditions of the family and environment [5], [6]. If the conditions of the factors that influence the incidence of stunting are known, it can be predicted whether a child is stunted or not.

Using a data mining classification approach can be used to predict whether a child is stunted or not. One of the methods that can be used in data mining classification is the Support Vector Machine (SVM). The SVM method is a classification approach developed in computer science since the 1990s and is getting more popular today [7], [8]. The principle of this method is to determine the best separator function that can separate two or more classes. In SVM, the function separator is known as a hyperplane. Classifiers on SVM can be carried out on data that can be separated linearly or on data that cannot be separated linearly. Classification is assisted by using kernel functions on data that cannot be separated linearly. Several kernel functions are commonly used, such as linear, polynomial, Radial Basis Function (RBF), and sigmoid [9]. It is complicated to illustrate how to determine the hyperplane using kernel functions.

The SVM method is a classification method with a complex process because the hyperplane to be determined depends on many predictor variables. The more predictors are used, the more the hyperplane will be in a higher dimensional space, so that the function is complicated to present explicitly. In addition to predictor variables, the amount of data used is also very influential in the classification process by SVM. The more observations that are used, the more parameters can be determined.

In this study, to clarify how classification works by the SVM method, the researchers will illustrate the classification process by taking simple examples of linear kernel functions. Afterwards, the researchers determined the best parameters to classify the incidence of stunting in toddlers in the city of Padang.

2. RESEARCH METHODS

2.1 Data Source

This study used primary data obtained directly from data sources, i.e., 440 toddlers. The population of interest in the study is children under five years of age in Padang, West Sumatra. The survey was conducted at several Posyandu (maternal and children's health services), daycares and housing, which were chosen using a purposive sampling technique. This primary data is related to ten factors that are predicted to affect nutritional status. Primary data collection was done by interview technique, using the questionnaire as a guide. The survey was conducted from August to September 2017.

2.2 Research Variables

The variables used in this study consisted of response variables and predictor variables. The response variable used is the incidence of stunting in toddlers (Y) that is categorized into 2: the incidence of stunting with ($y = 1$) and the incidence of not stunting with ($y = 2$). The predictor variables used are the variables that affect the stunting incidence. In this study, there are several continuous variables and several categorical variables. The continuous variables used are the child's age (X_2), the length of the baby at birth (X_3), and the number of dependents (X_5). The categorical variables used are gender (X_1), average income (X_4), employment status (X_6), breastfeeding status (X_7), mother's behavior status (X_8), mother's last education (X_9), and mother's knowledge status (X_{10}).

2.3 Definition of Stunting

Stunting was measured by the number of standard deviations that a child's length or height is below or above the median of the 2006 WHO growth reference population data on every child's length or height, age, and sex [10]. A number of standard deviations (z-score) less than -2 was considered stunting [11].

2.4 Classification

Classification is a way of categorizing objects based on the characteristics possessed by the classification object [10]. In the classification, there are dependent variables and independent variables (predictors). The dependent variable is divided into several classes or specified categories. The independent variables are the predicted variables that affect the result of classification.

2.5 Hyperplane

In a d -dimensional space, a hyperplane is a subspace of a certain dimension $d - 1$. For example, in a two-dimensional space, a hyperplane is a subspace of one dimension. In other words, the hyperplane is a line. In three-dimensional space, the hyperplane is a subspace of two dimensions. In other words, the hyperplane is a field. It won't be easy to visualize the shape of the hyperplane in more than three-dimensional space [12]. In two-dimensional space, the hyperplane is a line given by the Eq. (1):

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

where X_1 dan X_2 is the variable, $\beta_0, \beta_1, \beta_2$ are the parameter [13].

In three-dimensional space, a hyperplane is a plane that can be expressed by (2).

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 = 0 \quad (2)$$

Furthermore, Eq. (2) can be extended to find the hyperplane that is in d -dimensional space that can be expressed by using Eq. (3).

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_d X_d = 0 \quad (3)$$

2.6 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a machine learning classification method that finds the best hyperplane that serves as a data class separator. In the SVM method of classification, the hyperplane maximizes the distance between the hyperplanes with each of the closest data points for each class.

2.6.1 Hard Margin

Hard-margin is a term in SVM used when a hyperplane can completely separate both data classes. Suppose there is data available on d -dimensional space denoted by $x_i \in R^d$ and label each class, which is denoted by $y_i = -1, +1$ for $i = 1, 2, 3, \dots, n$ where n is the amount of the data. Assume the two classes can be completely separated so that the hyperplane can be defined in the form of Eq. (4) [12].

$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_d x_{id} = 0 \quad (4)$$

The sample x_i , which belongs to the class $y_i = -1$ can be formulated as a function that satisfies the inequality (5).

$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_d x_{id} \leq -1 \quad (5)$$

while the sample x_i , which belongs to the class $y_i = +1$ can be formulated as a function that satisfies the inequality (6)

$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_d x_{id} \geq +1 \quad (6)$$

Inequality (5) and (6) can be written in general by inequality (7).

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_d x_{id}) \geq 1 \quad (7)$$

Assume that $\mathbf{w} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_d \end{bmatrix}$ and $\mathbf{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \dots \\ x_{id} \end{bmatrix}$ as a result, it can be obtained $\mathbf{w}^T \mathbf{x}_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_d x_{id}$.

Then the inequality (7) can be written as

$$y_i(\mathbf{w}^T \mathbf{x}_i + \beta_0) \geq 1 \quad (8)$$

Getting the largest margin can be done by optimizing the distance between the hyperplane and its nearest data point. This can be formulated with the Quadratic Programming (QP) problem by finding the vector \mathbf{w} that minimizes the equation:

$$\tau(\mathbf{w}) = \frac{1}{2}(\mathbf{w}^T \mathbf{w}) \quad (9)$$

taking into account the inequality constraint

$$y_i(\mathbf{w}^T \mathbf{x}_i + \beta_0) \geq 1 \quad (10)$$

Problems (9) and (10) can be restated using Lagrange Multiplier is by adding a constraint function which is multiplied by the Lagrange multiplier ($\alpha_i \geq 0$) to the objective function (11), so that obtained

$$L(\mathbf{w}, \beta_0, \alpha_i) = \frac{1}{2}(\mathbf{w}^T \mathbf{w}) - \sum_{i=1}^n \alpha_i (y_i(\mathbf{w}^T \mathbf{x}_i + \beta_0) - 1) \quad (11)$$

The double formula usually called the Wolfe dual obtained in the problem is written in Eq. (12).

$$\max_{\alpha_i} L = \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \quad (12)$$

with

$$\alpha_i \geq 0 \text{ and } \sum_{i=1}^n \alpha_i y_i = 0 \quad (13)$$

2.6.2 Soft Margin

Using SVM soft margin, the result of classifying most of the data is true, but it allows misclassification to occur for some points around the dividing limit [13]. On the soft margin, the classification problem can be interpreted as a way to find the best hyperplane that can separate the two classes by allowing classification errors around the hyperplane. The double formula for this case is:

$$\max_{\alpha_i} L = \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \quad (14)$$

with

$$0 \leq \alpha_i \leq C \text{ and } \sum_{i=1}^n \alpha_i y_i = 0 \quad (15)$$

2.7 SVM on Nonlinearly Separable Data

If data cannot be separated linearly, the SVM soft margin cannot find the best hyperplane to minimize the number of misclassified data points. By using the SVM method, the problem was modified by incorporating kernel functions into the nonlinear SVM. The kernel is used to convert data into more dimensional space to separate the data linearly [13].

A kernel is a function that is used to measure the similarity of two observations [12]. The basis of the kernel is mapping the data into the new higher dimensional space using a function. The kernel mapping function can be formulated as follows [13],

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)\phi(\mathbf{x}_j) \quad (16)$$

where $\phi(\mathbf{x}_i)$ dan $\phi(\mathbf{x}_j)$ is the Hilbert Space. Eq. (14) can be transformed into (17).

$$\max_{\alpha_i} L = \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (17)$$

with

$$0 \leq \alpha_i \leq C \text{ and } \sum_{i=1}^n \alpha_i y_i = 0 \quad (18)$$

To predict the class of an object z , a discriminant function will be used whose basic operation is the sum of the weights of the similarity of object z and a collection of objects that have been previously selected (support vector). Assume S is the support vector, the function is formulated as follows [14]:

$$f(\mathbf{z}) = \sum_{x_i \in S} \alpha_i y_i K(\mathbf{z}, \mathbf{x}) + \beta_0 \quad (19)$$

If $f(\mathbf{z})$ is positive, then object \mathbf{z} belongs to class +1, and if $f(\mathbf{z})$ is negative, object \mathbf{z} belongs to class -1. In general, to predict the class of an object \mathbf{z} can use the function:

$$\text{sign}(f(\mathbf{z})) \begin{cases} +1, & f(\mathbf{z}) > 0 \\ -1, & f(\mathbf{z}) < 0 \end{cases} \quad (20)$$

3. RESULTS AND DISCUSSION

3.1. Application of SVM to Predict the Classification of Stunting Events

In classifying the data using the SVM method, it is necessary to determine the parameter value cost (C) and the kernel to be used. According to Vapnik in [9], a popular kernel used in the classifier is the Radial Basis Function (RBF) kernel, so that in this research, we will use the kernel function formula as follows:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (22)$$

Thus, to determine the best hyperplane for data classification by using the SVM method, it is necessary to determine the values of the two parameters, namely cost (C) and gamma (γ), which result in good accuracy. Several values will be used to choose the best parameter. The trial and error method can be used to find the value of the parameters C and γ [15]. Based on [16], the range of C values that can be used in SVM is between 10^{-2} to 10^2 . Some of the C values that will be used in this study are 0.01, 0.1, 1, 10, 100, and the γ are 1, 2, 3, 4 and 5.

In determining the best parameter values from the 440 data points used, the data is divided into two parts, namely training data and testing data. The distribution of the amount of training data and testing data can be seen in Table 2.

Table 2. Distribution of training data and testing data

	Percentage	Amount
Training Data	75%	330
Testing Data	25%	110
Total	100%	440

The distribution of training data and testing data is carried out randomly to obtain 330 training data and 110 testing data. Furthermore, several repetitions will be done to determine the best parameters to be used in the classification of stunting. The following will display the average accuracy of the ten possible parameter values that can be used in the classification process. Ten possible parameters were obtained after 100 repetitions.

Table 3. Average accuracy of the ten possibilities parameter values

Cost Value	Gamma Value	Training Accuracy	Testing Accuracy
0.01	1	68.24	67.992
1	1	94.534	97.27
1	2	97.212	98.544
1	3	98.286	98.983
1	4	98.64	98.483
1	5	99.09	99.09
10	2	99.7	100
10	3	99.9	100
10	4	99.85	100
10	5	100	100

Based on Table 3, it can be assumed that the best parameters that can be used to form a hyperplane in this study are the value of cost = 10 and gamma = 5. Furthermore, these values are used to predict the classification of the incidence of stunting in toddlers in the city of Padang. The following table presents the RBF kernel's prediction results with cost = 10 and gamma = 5.

Table 4. Confusion Matrix Results

Predicted	Actual	
	Stunting	Not stunting
Stunting	140	0
Not Stunting	0	300

Based on Table 4, it can be seen that 140 data can be truly predicted the incidence of stunting, and 300 classified as not stunting can be predicted correctly. Overall, the accuracy obtained in this study are:

$$Accuracy = \frac{TP + FP}{TP + TN + FP + FN} = \frac{140 + 300}{440} = 1$$

The accuracy value obtained is 100%. This explains that the prediction of classification of the incidence of stunting produced by using the SVM method in this study has a very high level of accuracy equal to 100%.

4. CONCLUSIONS

From the results of the research, it can be concluded that:

1. Based on the classification illustration described using the linear kernel function, the parameters that must be specified are α_i and β_0 . The number of parameters α that must be determined depends on the amount of data used. The value of α that is obtained by maximizing the margin that separates the two classification classes by using the double formula:

$$\max_{\alpha_i} L = \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

Furthermore, the value of β_0 is obtained using the constraint function, namely by choosing a certain α_i , which is a support vector.

2. Based on the research results that have been obtained, it is known that the best parameters used in predicting the classification of stunting in children under five in the city of Padang are cost = 10 and gamma = 5, with the kernel function used is the Radial Basis Kernel Functions (RBF). The accuracy obtained in this classification is about 100%. This case explains that all observations can be correctly predicted.
3. In this study, researchers used the Support Vector Machine method to predict the classification of two classes. In future research, it is suggested that researchers can use the Support Vector Machine (SVM) method to predict the classification of many classes by taking different parameter values.

REFERENCES

- [1] J. Hoddinott, J. A. Maluccio, J. R. Behrman, R. Flores, and R. Martorell, "Effect of a nutrition intervention during early childhood on economic productivity in Guatemalan adults," *Lancet*, vol. 371, no. 9610, pp. 411–416, Feb. 2008, doi: 10.1016/S0140-6736(08)60205-6.
- [2] The Ministry of Health Republic Indonesia, *Handbook of Nutritional Status Monitoring Results 2015*. Jakarta: The Ministry of Health Republic Indonesia, 2015.
- [3] The Ministry of Health Republic Indonesia, *Handbook of Nutritional Status Monitoring Results 2016*. Jakarta: The Ministry of Health Republic Indonesia, 2016.
- [4] M. de Onis, C. Garza, A. W. Onyango, and M.-F. Rolland-Cachera, "WHO growth standards for infants and young children," *Arch. Pédiatrie*, vol. 16, no. 1, pp. 47–53, 2009, doi: <https://doi.org/10.1016/j.arcped.2008.10.010>.
- [5] C. P. Stewart, L. Iannotti, K. G. Dewey, K. F. Michaelsen, and A. W. Onyango, "Contextualising complementary feeding in a broader framework for stunting prevention," *Matern. Child Nutr.*, vol. 9, no. S2, pp. 27–45, Sep. 2013, doi: <https://doi.org/10.1111/mcn.12088>.
- [6] J. P. Wirth et al., "Assessment of the WHO Stunting Framework using Ethiopia as a case study," *Matern. Child Nutr.*, vol. 13, no. 2, p. e12310, Apr. 2017, doi: <https://doi.org/10.1111/mcn.12310>.
- [7] M. A. Chandra and S. S. Bedi, "Survey on SVM and their application in imageclassification," *Int. J. Inf. Technol.*, vol. 13, no. 5, pp. 1–11, 2021, doi: 10.1007/s41870-017-0080-1.
- [8] D. Şen, C. Ç. Dönmez, and U. M. Yıldırım, "A Hybrid Bi-level Metaheuristic for Credit Scoring," *Inf. Syst. Front.*, vol. 22, no. 5, pp. 1009–1019, 2020, doi: 10.1007/s10796-020-10037-0.
- [9] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer, 1995.
- [10] D. K. Srivastava and L. Bhambhu, "Data Classification Using Support Vector Machine," *J. Theor. Appl. Inf. Technol.*, vol.

- 12, no. 1, pp. 1–7, 2010.
- [11] World Health Organization (WHO), *Child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: methods and development*. Geneva: WHO, 2006.
 - [12] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning with Applications in R*. New York: Springer, 2013.
 - [13] M. Awad and R. Khanna, *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*. Apress, 2015.
 - [14] L. Wang, *Support Vector Machines: Theory and Applications*. Berlin: Springer, 2005.
 - [15] J. Nayak, B. Naik, and D. H. Behera, “A comprehensive survey on support vector machine in data mining tasks: Applications & challenges,” vol. 8, pp. 169–186, Jan. 2015, doi: 10.14257/ijda.2015.8.1.18.
 - [16] C.-M. Huang, Y.-J. Lee, D. K. J. Lin, and S.-Y. Huang, “Model selection for support vector machines via uniform design,” *Comput. Stat. & Data Anal.*, vol. 52, no. 1, pp. 335–346, 2007.

