Indonesian Journal of Data and Science

Volume 4 Issue 2
ISSN 2715-9930
https://doi.org/10.56705/ijodas.v4i2.68

Research Article

Spatial Prediction of Stunting Incidents Prevalence Using Support Vector Regression Method

Andi Widya Mufila Gaffar[1,*], Sugiarti[2], Dewi Widyawati[3], Andi Muhammad Kemal Arief Hidayat Paharuddin[4], Andi Vania Anastasia[5]
[1]Universitas Muslim Indonesia, Makassar, Indonesia, widya.mufila@umi.ac.id
[2]Universitas Muslim Indonesia, Makassar, Indonesia, sugiarti.sugiarti@umi.ac.id
[3]Universitas Muslim Indonesia, Makassar, Indonesia, dewiwidyawati@umi.ac.id
[4]Universitas Muslim Indonesia, Makassar, Indonesia, kemalarif92@gmail.com
[5]Universitas Muslim Indonesia, Makassar, Indonesia, andivaniaanastasia25@gmail.com

Correspondence should be addressed to Andi Widya Mufila Gaffar; widya.mufila@umi.ac.id

Received 10 June 2023; Accepted 18 June 2023; Published 31 July 2023

Abstract:
Stunting in toddlers is a major nutritional problem faced by Indonesia, with a high incidence rate occurring in several provinces across the country. This nutritional issue can occur at any age, starting from the prenatal stage, infancy, childhood, adolescence, adulthood, and even in the elderly. To reduce the prevalence of stunting in affected provinces, prevention efforts are essential, including predicting the spread of stunting incidents in each region. Therefore, this research conducted spatial prediction of the prevalence rate of stunting incidents using Machine Learning, specifically Support Vector Machine based Regression. The results of this study produced a prediction model with an RMSE (Root Mean Square Error) value of 0.008689303 and a multiple correlation coefficient of 0.65912721. Based on these findings, the predictive model utilized demonstrated satisfactory performance in predicting the prevalence rate of stunting incidents in each area.

Keywords: Stunting, Prediksi, Spasial, Machine learning, Support Vector Regression.
Dataset link:

1. Introduction

The development of information technology has had a significant impact on various fields, including the healthcare sector. Consequently, the need for information has become more crucial to be accurate, precise, and timely. Diverse information, regardless of its nature, whether positive or negative, can influence the emergence of specific issues, particularly in the context of health [1]. One of the current pressing health problems is stunting, which has a remarkably high incidence rate in several provinces in Indonesia, as seen in **Figure 1**. Stunting in toddlers is a major nutritional problem faced by Indonesia. Based on data from the Nutritional Status Monitoring (PSG) over the past three years, stunting has the highest prevalence compared to other nutritional issues such as undernutrition, wasting, and overweight. The prevalence of stunting in toddlers increased from 27.5% in 2016 to 29.6% in 2017 [2].

Figure 1. Proportion of Poor Nutrition and Undernutrition in Toddlers by Province [3]

Nutritional problems can occur at all ages, starting from the prenatal stage, infancy, childhood, adolescence, adulthood, and even in old age. The first two years of life are critical as they present the primary opportunity to prevent growth and developmental disorders. During this period, permanent nutritional disturbances may occur, which cannot be recovered even if nutritional needs are met in later stages [4].

Stunting can have both short-term and long-term effects. Short-term effects include delayed cognitive, motor, and language development, as well as the risk of disabilities, infectious diseases, and mortality. Long-term effects include the risk of degenerative diseases such as high blood pressure, diabetes, coronary heart disease, and stroke. Further impacts are observed in adulthood, such as reduced work efficiency [5].

To reduce the prevalence of stunting incidents, prevention is necessary by predicting the occurrence and distribution of stunting cases in each region. According to Labolo's study [6], which monitored and collected data in Gorontalo province, there was an issue of underestimation due to inaccurate monthly data collection, as it was based solely on estimated cases from local health centers. Hence, prediction of the number of stunting cases was done using K-Nearest Neighbor (K-NN) and Support Vector Machine (SVM) algorithms with backward elimination feature selection.

Byna's research [7] aimed to improve the accuracy of stunting incidence prediction using the Backward Elimination Method with Support Vector Machine (SVM) algorithm, resulting in higher accuracy in classification and more precise outcomes. This approach can offer valuable solutions for healthcare experts in determining stunting occurrences.

Addressing the causes of stunting requires supportive prerequisites, including: (a) political commitment and policy implementation, (b) government and cross-sector involvement, and (c) capacity for execution [8]. Therefore, this study aims to predict the spatial prevalence rate of stunting incidents using machine learning, specifically Support Vector Regression, as a foundation for early prevention of stunting.

2. Method

The stages of research in predicting the spatial prevalence rate of stunting incidents using the Support Vector Regression method are as follows:

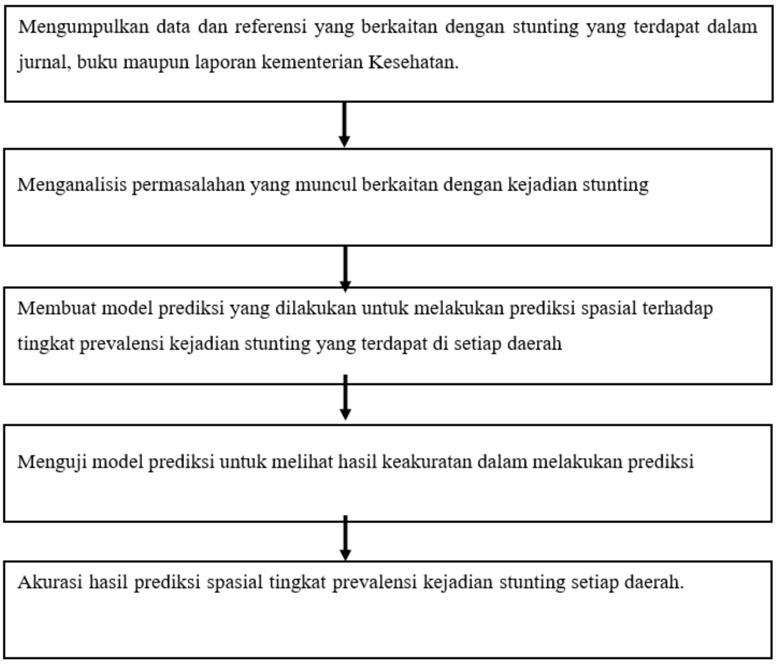

Figure 2. Flowchart of Research Stages

Based on **Figure 2**, the explanation of the research stages conducted is as follows:

1. Data Collection and References Collecting data sources and references related to stunting. Identifying the variables to be observed/measured for use in the prediction model. Data sources can be obtained from journals, books, and reports related to stunting.
2. Problem Analysis Analyzing the requirements of the existing problem. Based on the analysis results, it was found that the prediction model heavily utilizes machine learning methods, where the models used can learn autonomously. The machine learning method employed is Support Vector Regression.
3. Prediction Model Creation In this stage, the prediction model is developed by coding and utilizing both the training data and test data, which were obtained from the dataset collected. The stages of creating the prediction model can be seen in **Figure 3**.

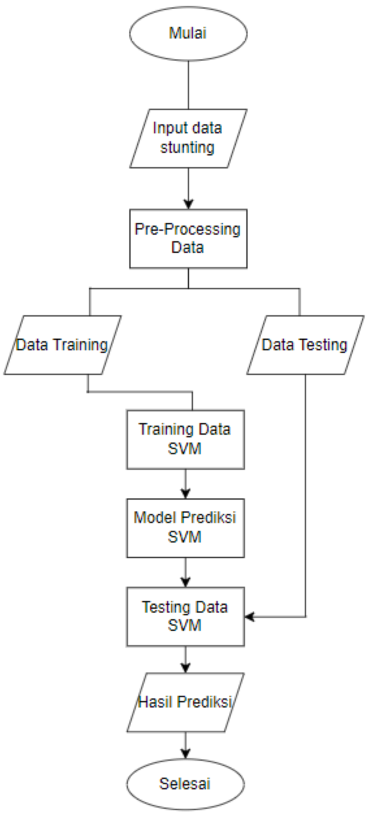

Figure 3. Flowchart of Prediction Model Stages

4. Prediction Model Testing

   After the coding of the prediction model is constructed, the model is then tested to assess its accuracy and precision in predicting the spread of stunting incidents.

Machine Learning

Machine Learning (ML) is one application of Artificial Intelligence (AI) that focuses on developing a system capable of learning on its own without being programmed repeatedly. ML requires data (training data) for the learning process before producing a result. In simple terms, Machine Learning can be described as computer programming to achieve specific criteria/performance using a set of training data or past experiences [9].

Support Vector Regression

Support Vector Machine is one of the methods developed in machine learning for classification and regression problems. Support Vector Regression (SVR) is a type of optimization model that can be used to model nonlinear processes, thus minimizing prediction errors and model complexity [10].

Table 1. Types of Kernel Functions

Jenis Kernel	Rumus
Linear	$x^T x$
Polynomial	$(\gamma x^T x_i + 1)^p$
Radial Basis Function (RBF)	$\exp(-\gamma \parallel (x^T x_i \parallel)^2$
Tangent hyperbolic (sigmoid)	$\tanh(\gamma x^T x + \beta^1)$, dimana $\beta, \beta^1 \epsilon \Re$

The SVM algorithm in regression cases aims to construct a hyperplane that is as close to the data points as possible. Therefore, the optimal hyperplane (separating line) is the one that best fits all input data with a minimum error ε by mapping the input vectors into a higher-dimensional space [11]. The learning process to find the support vector points requires a kernel function to efficiently solve the feature transformation to a new higher-dimensional space [12].

The prevalence rate of stunting incidents is predicted by building a regression model using the SVR algorithm, considering the search for the most optimal hyperplane [13], as illustrated in **Figure 4**.

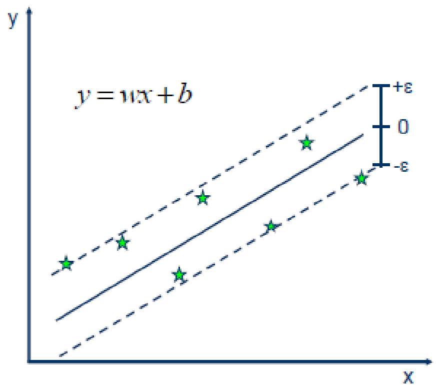

Figure 4. Optimal Regression Function (Hyperplane) [14]

The regression line function as the optimal hyperplane is denoted by Equation 1.

$$y_i = f(x) = wx + b \qquad w \in X, b \in \mathcal{R} \tag{1}$$

where f(x) represents the regression line function, b is the bias, and w is the weight.

The constructed prediction model is then analyzed and evaluated to determine its effectiveness and accuracy. The evaluation of the model can be performed using Root Mean Squared Error (RMSE) and multiple correlation coefficient (R2). RMSE is a procedure used to measure the average error in predicting the model using the SVR method, formulated in Equation 2 [15].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y'_i - y_i)^2} \tag{2}$$

where n is the number of data, "y'"$_i$ is the predicted value with output iteration i, and y_i is the actual value with output iteration i.

The multiple correlation coefficient (R2) is a relative measure of the relationship between the actual data denoted by Y_i and the predicted data denoted by "$\hat{Y}_i$." It indicates the strength of the relationship between the two variables. The formulation is given in Equation 3 [16].

$$R^2 = 1 - \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} \tag{3}$$

where Y_i is the observed response for iteration i, "$\hat{Y}_i$" is the predicted response for iteration i, and "$\bar{Y}$" is the mean observation.

3. Results and Discussion

The dataset used in developing the prediction model consists of data on the prevalence rate of stunting incidents and socio-economic conditions, specifically poverty line data. This research's prediction modeling applies the Support Vector Regression method, implemented in R programming using the e1017 package. The first step involves normalizing the dataset within the range of 0 to 1, followed by dividing the data into training and testing sets with a 70:30 ratio and conducting 10 trial runs, as shown in Appendix 7. The training is performed using several kernels, such as radial basis function, polynomial, linear, and sigmoid. The constructed prediction model is then analyzed and evaluated to determine its effectiveness and accuracy. The evaluation is carried out through the calculation of Root Mean Squared Error (RMSE) and multiple correlation coefficient (R2), as presented in **Table 2**.

Table 2. RMSE from the Use of Various Kernel Functions

Kernel	RMSE	
Radial Basis Function	**0.008689303**	
Polynomial	0.010244304	
Linear	0.009437419	
Sigmoid	0.039237599	

The prediction model is built using the default gamma parameter in R, which has a value of 0.125. The result of the prediction using SVR is an RMSE of 0.008689303 for the prevalence rate of stunting incidents. The comparison of the RMSE between actual and predicted values for the prevalence rate of stunting incidents in the year 2018 is shown in **Figure 7**. The comparison graph is not satisfactory since, at some points, the actual and predicted data are far apart.

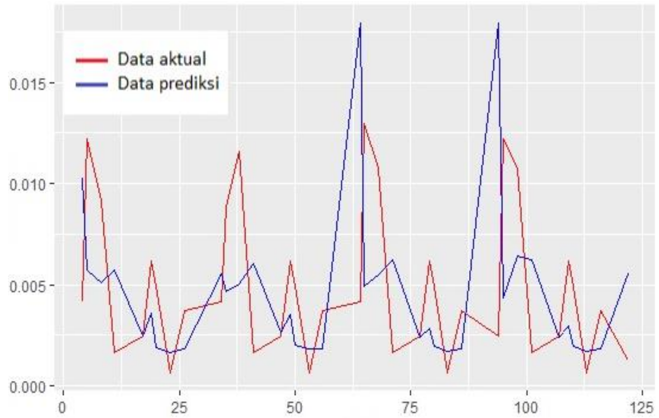

Figure 5. Comparison of Actual and Predicted Prevalence Rates of Stunting Incidents in 2018

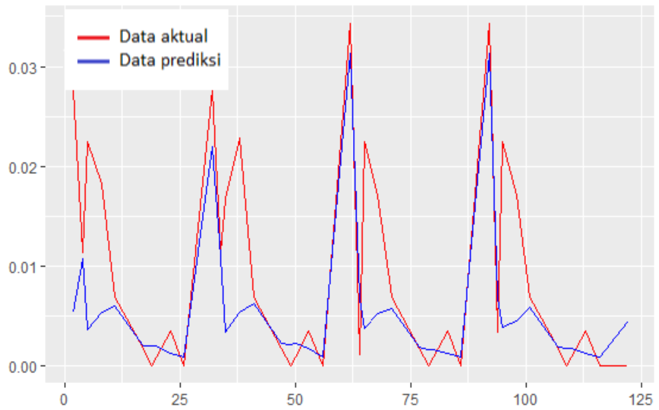

Figure 6. Comparison of Actual and Predicted Prevalence Rates of Stunting Incidents in 2019

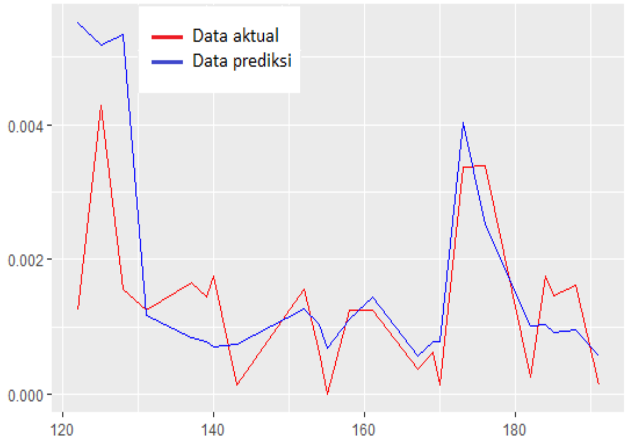

Figure 7. Comparison of Actual and Predicted Prevalence Rates of Stunting Incidents in 2020

Parameter tuning is conducted to obtain better results and performance in the prediction model. The tuning parameter is tested within a range of gamma values from 0.1 to 10, with 10 trial runs, as shown in the RMSE Table. The results of parameter tuning are presented in **Table 3**. The best-performing model utilizes a gamma value of 0.5, an epsilon value of 0.1, and a cost value of 1, using the Radial Basis Function kernel.

Table 3. Tuning Parameter Results

Gamma (kernel RBF)	RMSE	R^2
0.1	0.009633450	0.65327093
0.125	0.009181059	0.63403621
0.5	**0.008689303**	**0.65912721**
1	0.009129295	0.64935258
5 10	0.009386928 0.009442372	0.63986155 0.64151199

A gamma value of 0.5 is considered the optimal parameter because, during the trial runs, RMSE values decrease from 0.1 to 0.5 but increase when gamma is set to 1. On the other hand, the R2 values consistently increase with higher gamma values, indicating stronger multiple correlation coefficients. The RMSE value obtained using a gamma parameter of 0.5 is 0.008689303, and the multiple correlation coefficient is 0.65912721.

4. Conclusion

This research has successfully built a prediction modeling for the prevalence rate of stunting incidents by applying the Support Vector Machine based Regression method. The prediction model was constructed with the best gamma parameter value of 0.5, an epsilon value of 0.1, and a cost value of 1, using the Radial Basis Function kernel. The prediction model resulted in an RMSE value of 0.008689303 for the prevalence rate of stunting incidents and a multiple correlation coefficient of 0.65912721. Based on these findings, the utilized prediction model demonstrates satisfactory performance in predicting the prevalence rate of stunting incidents in each region.

This research can be further developed by conducting predictions on a smaller scale, such as predicting at the district or village level, and by adding factors like maternal nutrition and exclusive breastfeeding as additional variables in the prediction process.

Acknowledgments

Gratitude is extended to the Lembaga Penelitian dan Pengembangan Sumberdaya (LP2S) Universitas Muslim Indonesia (UMI) for the financial support provided to facilitate the implementation of this research.

References

[1] H. Purwanto and S. Dalis, “Perancangan Sistem informasi Pelayanan Kesehatan Peduli Remaja (PKPR) Berbasis Web Pada Puskesmas,” *SEMNASTEK*, 2017.

[2] Pusat Data dan Informasi, “Situasi Balita Pendek (Stunting) di Indonesia,” Jakarta, 2018.

[3] Kementerian Kesehatan RI, “Hasil Utama RISKESDAS 2018,” Jakarta, 2018.

[4] M. de Onis, M. Blossner, and E. Borghi, “Prevalence and Trends of Stunting,” *Public Health Nutr*, vol. 15, no. 1, 2020.

[5] P. Anindita, “Hubungan Tingkat Pendidikan Ibu, Pendapatan Keluarga, Kecukupan Protein dan Zinc dengan Stunting (Pendek) Pada Balita Usia 6-35 Bulan di Kecamatan Tembalang Kota Semarang,” *Jurnal Kesehatan Masyarakat*, vol. 1, no. 2, pp. 617–626, 2012.

[6] A. Y. Labolo, S. Mooduto, A. Bode, and D. I. C. R, “Penerapan Algoritma Spport Vector Machinedan K-Nearest Neighbor Menggunkan Feature Selection Backward Elimination Untuk Prediksi Status Penderita Stunting Pada Balita,” *TECNOSCIENZA*, vol. 6, no. 2, 2022.

[7] A. Byna and F. N. Anisa, “Backward Elimination Untuk Meningkatkan Akurasi Kejadian Stunting dengan Analisis Algoritma Support Vector Machine,” *Dinamika Kesehatan : Jurnal Kebidanan dan Keperawatan*, vol. 9, no. 2, 2018.

[8] Direktorat Jenderal Pembangunan dan Pemberdayaan Masyarakat Desa, “Panduan Fasilitasi Konvergensi Pencegahan Stunting di Desa,” Jakarta, 2018.

[9] R. Primartha, *Belajar Machine Learning*, 1st ed. Bandung: Informatika, 2018.

[10] A. Suárez Sánchez, P. J. García Nieto, P. Riesgo Fernández, J. J. del Coz Díaz, and F. J. Iglesias-Rodríguez, “Application of an SVM-based regression model to the air quality study at local scale in the Avilés urban area (Spain),” *Math Comput Model*, vol. 54, no. 5–6, pp. 1453–1466, Sep. 2011, doi: 10.1016/j.mcm.2011.04.017.

[11] T. Trafalis and H. Ince, “Support vector machine for regression and applications to financial forecasting,” in *IJCNN*, Jul. 2000.

[12] S. Maji, A. C. Berg, and J. Malik, “Efficient Classification for Additive Kernel SVMs,” *IEEE Trans Pattern Anal Mach Intell*, vol. 35, no. 1, pp. 66–77, Jan. 2013, doi: 10.1109/TPAMI.2012.62.

[13] T. Cheng, J. Wang, and X. Li, “The support vector machine for nonlinear spatio-temporal regression,” in *Proc Geocomputation*, 2007.

[14] N. Deng, Y. Tian, and C. Zhang, *Support Vector Machines Optimization Based Theory, Algorithms, and Extensions*, 1st ed. New York : CRC Press Taylor, 2013.

[15] M. Esmaeili, A. Salimi, C. Drebenstedt, M. Abbaszadeh, and A. A. Bazzazi, “Application of PCA, SVR, and ANFIS for modeling of rock fragmentation,” *Arabian Journal of Geosciences*, vol. 8, no. 9, pp. 6881–6893, 2015.

[16] N. R. Draper and H. Smith, *Applied regression analysis*, vol. 326. John Wiley & Sons, 2014.