Nonlinear Dynamics and Systems Theory, **25** (6) (2025) 621–629

Analysis and Prediction of Stunting Rate in East Java Province Using Support Vector Regression and Decision Tree Method

T. Herlambang [1], F. Fauzi [2], B. Suharto [3*], I. Kurniastuti [1], Z. Othman [4], R.W. Abdullah [5], and K. Oktafianto [6]

[1] *Department of Information System, Universitas Nahdlatul Ulama Surabaya, Indonesia.*
[2] *Undergradute Program of Information System, Faculty of Economy Business and Digital Technology, Universitas Nahdlatul Ulama Surabaya.*
[3] *Department of Tourism and Hospitality, Faculty of Vocational, Airlangga University, Surabaya.*
[4] *Department of Diploma Studies, Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Melaka Malaysia, Malaysia.*
[5] *Center of Advanced Computing Technology(C-ACT), Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Melaka Malaysia, Malaysia.*
[6] *Department of Mathematics, University of PGRI Ronggolawe, Indonesia.*

Received: October 22, 2024; Revised: November 16, 2025

Abstract: Currently, in the health sector, Indonesia is facing several problems that the government is focusing on, one of which is stunting. The problem of stunting is a focus for the government since it ranks second after the problem of maternal mortality during childbirth. Stunting is a term used in the health sector describing a condition of growth failure in children under five due to chronic malnutrition in the first 1000 days of life. The causes of stunting can be identified from low nutritional intake and health status of pregnant women at risk of giving birth to babies with low body weight and below-standard baby length. With the current advances in the field of information technology, the stunting rate can be estimated using a machine learning method. There are numerous machine learning methods for prediction, such as Support Vector Regression (SVR), Decision Tree, K-Nearest Neighbour (K-NN) and many more. In this study, we aim to compare two prediction methods, namely Support Vector Regression (SVR) and Decision Tree, and determine how both methods succeed in predicting well. The Support Vector Regression (SVR) method achieved the best error value of 0.137 and the Decision Tree method had the best error value of 0.164.

Keywords: *stunting; prediction; machine learning; Support Vector Regression; decision tree.*

Mathematics Subject Classification (2020): 62J05, 70-10, 90Bxx.

* Corresponding author: `mailto:bambang.suharto@vokasi.unair.ac.id`

1 Introduction

Stunting is the focus of health concerns for the Indonesian government and the East Java provincial government. Stunting is a condition of malnutrition associated with insufficient nutrients in the past so it is considered a chronic nutritional problem [1]. Children experiencing stunting are more likely to grow into unhealthy and poor adults. Stunting of children is also associated with increased vulnerability to disease, both infectious and non-communicable diseases (NCDs), as well as increased risk of overweight and obesity [2].

The stunting problem in East Java initially occurred in areas with a low economic level, a low level of education, and a culture of early marriage, typically in Madura and the eastern part of East Java [3]. East Java provincial Nutritional Status Monitoring (DPSG) data shows that the prevalence of stunting among children under five in East Java is 27.1%, consisting of 17.6% short and 9.5% very short. The prevalence of stunting among children under five in one of the Madura Island districts, namely Bangkalan, is the highest in East Java, at 53.2%, with a prevalence of very short toddlers of 27.4% [4].

To address this issue, a method is required to predict the stunting rate in East Java province with the help of relevant advances in information technology. Prediction, in its definition, is the activity of estimating what will happen in the future using past data [5]. Through accurate and computerized prediction methods, decision making on policies to be taken to deal with this stunting problem becomes more measurable. In this research, the authors use the Support Vector Regression (SVR) and Decision Tree methods to predict the rate of stunting in East Java.

The Support Vector Regression (SVR) method has been widely used in various prediction studies. The Support Vector Regression (SVR) method is a derivative of the Support Vector Machine (SVM) method which has gained recognition for its ability to handle high-dimensional data and work well on relatively small datasets but with a large number of features [6]. Meanwhile, Decision Tree is a learning method that constructs a prediction model in the form of a tree structure. The tree consists of nodes representing decisions based on attribute values, and branches representing the results of those decisions [7]. In the previous research, both methods were used to predict the risk of stunting in families [8].

In this study, the Support Vector Regression (SVR) and Decision Tree methods were used to compare the performance of each method in predicting the rate of stunting in East Java. This approach is expected to provide a clear view of the stunting rate, making it easier for the parties involved to make decisions.

2 Research Methods

The dataset used in this study comes from the official website of the Ministry of Home Affairs (`https://aksi.bangda.kemendagri.go.id/emonev/`) which consists of 156 rows and 6 columns with a time span ranging from January 02, 2021 to December 18, 2024. The data in this study were analyzed using the Python programming language with a dataset as in Table 1 below.

After the data are obtained and explored, the next step is to carry out the stages of the research methodology one by one as shown in Figure 1 below.

1. **Problem Identification**: This research raises a case study on the prediction of stunting prevalence rates in East Java province.

Year	Regency	Number of Toddlers	Short Stunting	Very Short Stunting	Preva-lence(%)
02/01/2021	Pacitan	3068	2337	653	9.7
15/01/2021	Ponorogo	44307	482	1784	14.9
18/01/2021	Trenggalek	36144	2932	574	9.7
21/01/2021	Tulungagung	63329	167	431	3.3
31/01/2021	Blitar	76118	4258	1395	7.4
02/02/2021	Kediri	79244	8445	2327	13.6
15/02/2021	Malang	88532	651	1351	8.9
...	...	...	...	...	...
18/12/2024	Kota Batu	10945	106	269	12.1

Table 1: Dataset.

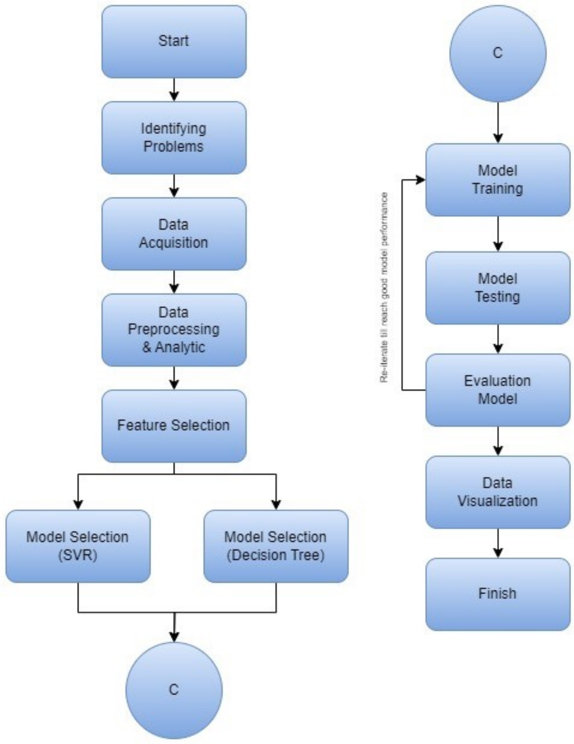

Figure 1: Research Methods.

2. **Data Acquisition**: The data used in this study is secondary data sourced from the official website of the Ministry of Home Affairs (`https://aksi.bangda.kemendagri.go.id/emonev/`) which consists of 156 rows and 6 columns with a time span ranging from January 02, 2021 to December 18, 2024.

3. **Data Preprocessing & Analytics**: The data preprocessing and analytics are basically used to see the initial condition of the dataset obtained from the source. At this stage, the dataset is identified from the data type to missing values that may occur in the dataset.

4. **Feature Selection**: This study uses the Pearson Product correlation analysis technique to find a linear relationship between two variables having a normal distribution [9]. Below is the function of the Pearson Product

$$r_{xy} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}}. \tag{1}$$

Notes:

- r_{xy}: Relationship coefficient
- N: Number of samples used
- ΣX: The total score of the question
- ΣY: Sum of total scores

5. **Model Selection**: Support Vector Regression (SVR) is a development algorithm of the Support Vector Machine (SVM) algorithm introduced by Cortes and Vapnik [10]. Like SVM, SVR also uses the best hyperplane in the form of a regression function by making the error as small as possible. The function of SVR can generally be written as follows:

$$f(x) = w\varphi(x) + b \tag{2}$$

with

- $f(x)$: Regression function
- w: Vector
- b: bias

and the decision boundary equation

$$W_x + b = +\varepsilon,$$

$$W_x + b = -\varepsilon$$

so that the hyperplane fulfills the equation

$$-\varepsilon < y - (W_x + b) < +\varepsilon$$

with the minimization function

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} |\xi_i|$$

and the constraint function

$$|y_i - w_i x_i| \leq \varepsilon + |\xi_i|.$$

Then, the Decision Tree algorithm begins by calculating the entropy value to measure the level of uncertainty or impurity in the dataset [11]. The equation for finding the entropy value is as follows:

$$\text{Entropy}(S) = \sum_{i=1}^{m} -p(w_i|S) \cdot \log_2(p(w_i|S)). \tag{3}$$

Description:

- S: The case set being analyzed.
- m: Total number of different classes within the data set S.
- $p(w_i|S)$: Probability of occurrence of class w_i in the data set S.

The next step is to calculate the gain value, which is a measure of how much information is obtained from separating data based on certain attributes. The gain calculation function is as follows:

$$\text{Gain}(S, J) = \text{Entropy}(S) - \sum_{i=j}^{n} p(v_i|S) \cdot \text{Entropy}(S_i). \tag{4}$$

Notes:

- S: The case set being analyzed
- J: Features/attributes considered in data splitting
- n: Number of classes in the node
- $p(v_i|S)$: Proportion of v values appearing in the class in the node
- S_i: The entropy of the composition of v values for the j-th class in the i-th data node.

6. **Model Training**: At this stage, the predicted values of the Support Vector Regression (SVR) and Decision Tree algorithms are trained based on the distribution of training data and testing data to obtain error values and accuracy values.
7. **Model Testing**: Model testing of learning outcomes against prepared testing data.
8. **Model Evaluation**: At the evaluation stage, the model trained and tested is calculated for accuracy based on the resulting error value. This research uses the Root Mean Square Error (RMSE) method to calculate the error value generated by the model. The function of the Root Mean Square Error (RMSE) is as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \widehat{y_i})^2}{n}} \tag{5}$$

with

- n: Quantity of data
- y_i: Actual value at the i-th data
- $\widehat{y_i}$: Predicted value at the i-th data

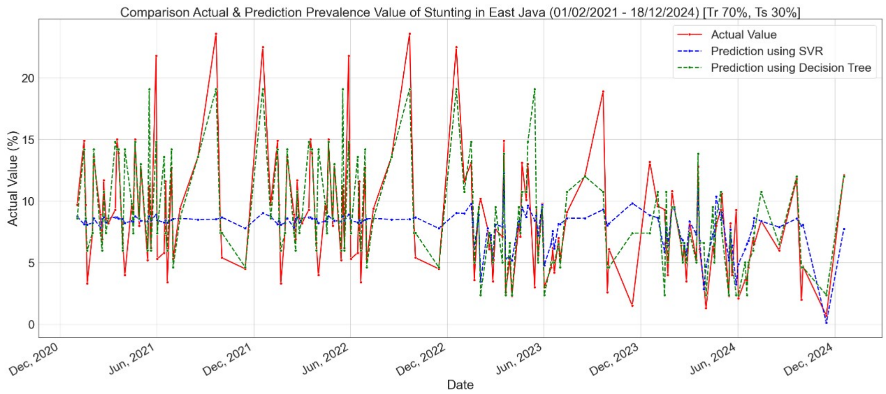

Figure 2: First Simulation Plot SVR Decision Tree (70% training data, 30% testing data).

3 Result and Discussion

The results of the testing simulation carried out with the proportion of training data and testing data are shown in Figure 2 to Figure 5.

The first simulation shown by Figure 2, produced SVR algorithm parameter values, that is, Cost = 1, Epsilon = 0.001, and Gamma = 0.1 on the Radial Basis Function kernel with a ratio of 70% training data and 30% testing data, with the SVR algorithm forecasting results not yet close to the actual value or less than optimal and the Decision Tree algorithm results close to the actual value. The results show that the first simulation produced an RMSE value of 0.1647. Meanwhile, the Decision Tree algorithm with the parameters of a max depth = 100 and min samples split = 10, produced an RMSE value of 0.2004.

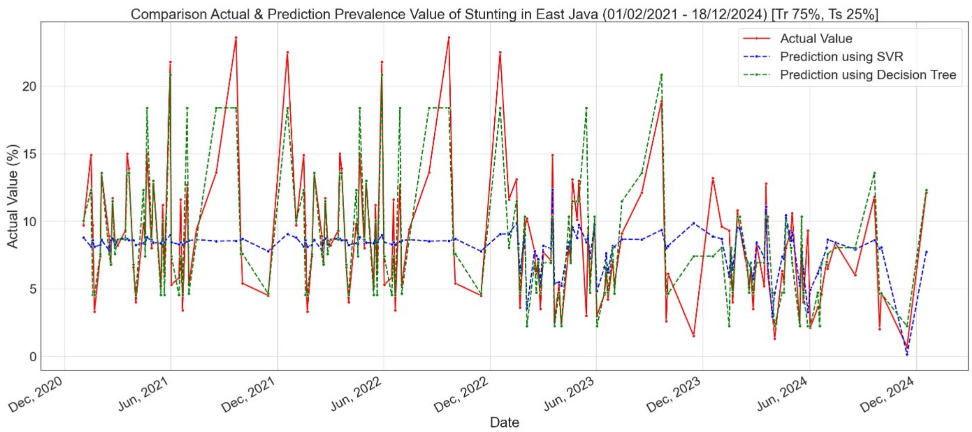

Figure 3: Second Simulation Plot SVR Decision Tree (75% training data, 25% testing data).

The results of the second simulation presented on Figure 3, show the parameter values of the SVR algorithm, namely Cost = 1 Epsilon = 0.001 Gamma = 0.1 on the

Radial Basis Function kernel with a ratio of 75% training data and 25% testing data with forecasting results close to the actual value or maximum enough. The first simulation resulted in an RMSE value of 0.1448. Meanwhile, the Decision Tree algorithm with the parameters such as max depth = 100 and min samples split = 10, produced an RMSE value of 0.1751.

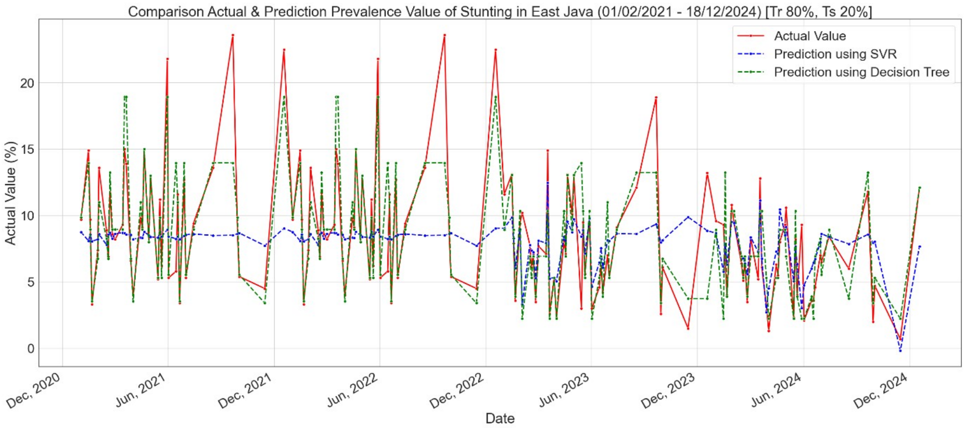

Figure 4: Third Simulation Plot SVR Decision Tree (80% training data, 20% testing data).

In the third simulation shown in Figure 4, the parameter values of the SVR algorithm are Cost = 1 Epsilon = 0.001 Gamma = 0.1 on the Radial Basis Function kernel with a ratio of 80% training data and 20% testing data with forecasting results close to the actual value or maximum enough. These first simulation results have an RMSE value of 0.1450. Meanwhile, using the Decision Tree algorithm with the parameters of max depth = 100 and min samples split = 10, the RMSE value is 0.1642.

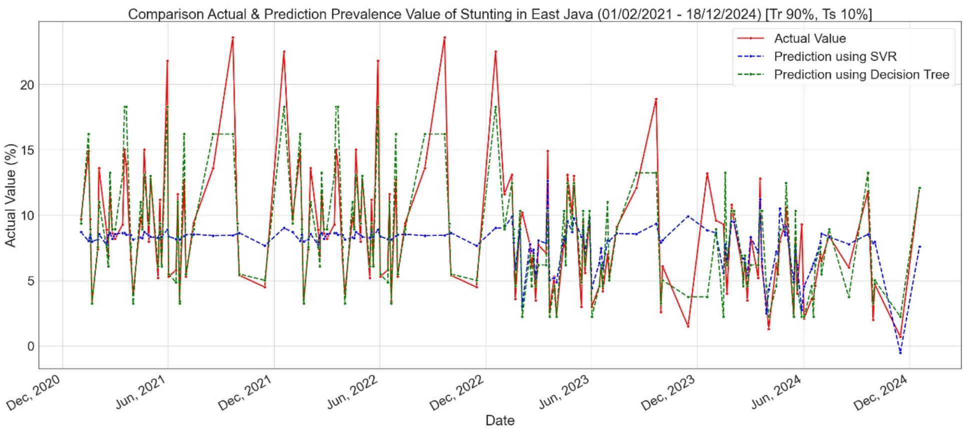

Figure 5: Fourth Simulation Plot SVR Decision Tree (90% training data, 10% testing data).

In the fourth simulation presented in Figure 5, the parameter values of the SVR algorithm are Cost = 1 Epsilon = 0.001 Gamma = 0.1 on the Radial Basis Function kernel with a ratio of 90% training data and 10% testing data with forecasting results close to the actual value or maximum enough. The results of this first simulation have an RMSE value of 0.1377. When using the Decision Tree algorithm with the parameters of max depth = 100 and min samples split = 10, the RMSE value is 0.2030.

The recapitulation of the simulation results of the Support Vector Regression (SVR) and Decision Tree algorithms can be seen in the table below.

Percentage Comparison of Training Data and Testing Data	**RMSE value by Support Vector Regression (SVR)**	**RMSE value by Decision Tree**
70% : 30%	0.1647	0.2004
75% : 25%	0.1448	0.1751
80% : 20%	0.1450	**0.1642**
90% : 10%	**0.1377**	0.2030

Table 2: Comparison of RMSE values.

In the table above, it can be seen that the maximum results produced by the Support Vector Regression (SVR) algorithm occurred in the fourth simulation with an RMSE value of 0.1377, while the best results of the Decision Tree algorithm occurred in the third simulation with an RMSE value of 0.1642.

4 Conclusion

Based on the simulation results, it can be concluded that the SVR algorithm produces the best simulation results in the fourth simulation with an RMSE value of 0.1377 and the Decision Tree algorithm produces the best simulation results in the third simulation with an RMSE value of 0.1642. The results prove that the Support Vector Regression and Decision Tree methods are able to provide good prediction results and fulfil the objective of this study and could be optimized in the next study.

Acknowledgment

The authors would like to thank LPPM-Universitas Nahdlatul Ulama Surabaya (UNUSA) for their support in providing the facility for this research. Additionally, gratitude is extended to the Centre of Research and Innovation Management of Universiti Teknikal Malaysia Melaka (UTeM) for sponsoring the publication fees under the Tabung Penerbitan CRIM UTeM.

References

[1] A. N. P. Wulandari, A. T. Devanti, A. F. Salsabila, R. A. Ardito, A. Amanullah, and E. Rahmawati. Preventing Stunting Through the GPS (Stunting Prevention Movement) Program in Murukan Village, Jombang Regency, East Java. *Gembira Journal: Community Service* **2** (2) (2024) 540–545.

[2] I. C. Dewi, and N. R. N. Auliyyah. Stunting Education as a Means to Minimize High Stunting Rates in Gambiran Village, Kalisat District. *JIWAKERTA: Scientific Journal of Practical Work Insights* **1** (2) (2020) 25–29.

[3] M. A. Syahputra, and A. C. Sitohang. The Effect of Poverty and Stunting on the Human Development Index (HDI) in East Java From 2019 to 2023. *GEMAH RIPAH: Business Journal* **4** (1) (2024) 74–79.

[4] T. Septiyanto, F. A. Mufarroha, D. R. Anamisa, and A. Jauhari. Predicting the Number of Stunting Cases in Madura Using Machine Learning. *JoMMiT: Multimedia and IT Journal* **7** (2) (2023) 70–76.

[5] D. A. Ningsih, R. Buaton, and A. Sihombing. Predicting Child Stunting Rates in Langkat District Using Multiple Linear Regression: (Case Study: PPKB-PPA Office of Langkat District). *Saturnus: Journal of Technology and Information Systems* **2** (4) (2024) 131–145.

[6] I. S. I. Putri, R. S. Pradini, and M. Anshori. Decision Tree Regression for Predicting the Prevalence of Stunting in East Nusa Tenggara Province. *Journal of Information Technology and Computers* **10** (2) (2024) 413–427.

[7] V. Asy'ari, M. Y. Anshori, T. Herlambang, I. W. Farid, D. F. Karya, and M. Adinugroho. Forecasting average room rate using k-nearest neighbor at Hotel S. In: *International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA)*), IEEE, Nov. 2023, 496–500.

[8] D. Novita, T. Herlambang, V. Asy'ari, A. Alimudin, and H. Arof. Comparison of K-Nearest Neighbor and Neural Network For Prediction International Visitor in East Java. *BAREKENG: Journal of Mathematics and Applied Sciences* **18** (3) (2024) 2057–2070.

[9] M. Y. Anshori, V. Asy'Ari, T. Herlambang, and I. W. Farid. Forecasting Occupancy Rate Using Neural Network at Hotel R. *International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA)*. (2023) 347–351.

[10] F. A. Susanto, M. Y. Anshori, D. Rahmalia, K. Oktafianto, D. Adzkiya, P. Katias and T. Herlambang. Estimation of Closed Hotels and Restaurants in Jakarta as Impact of Corona Virus Disease (Covid-19) Spread Using Backpropagation Neural Network. *Nonlinear Dynamics and Systems Theory* **22** (4) (2022) 457–467.

[11] M. Y. Anshori, T. Shawyun, D. V. Madrigal, D. Rahmalia, F. A. Susanto, T. Herlambang and D. Adzkiya. Estimation of Closed Hotels and Restaurants in Jakarta as Impact of Corona Virus Disease Spread Using Adaptive Neuro Fuzzy Inference System. *International Journal of Artificial Intelligence (IJAI)* **11** (2) (2022) 462–472.