Jurnal Ilmiah Edutic : Pendidikan dan Informatika /Vol.12, No.2 : 2025
ISSN 2407-4489
pp. 141-149

Population Growth Rate Prediction Using the Support Vector Regression Method

Rendy Yudha Pratama[a], Tahta Herdian Andika[b], Ulfa Isni Kurnia[c]

[a,c] *Faculty of Teacher Training and Education, Universitas Aisyah Pringsewu, Lampung, Indonesia*
[b] *Faculty of Informatics Engineering, Universitas Aisyah Pringsewu, Lampung, Indonesia*
email: : [a] nurullaili@itmnganjuk.ac.id, [b] nurul.laili.ft@um.ac.id, [c] triomaulana@itmnganjuk.ac.id, [d]muhammadin@kahuripan.ac.id, [e]umaranggana.ft@um.ac.id

Abstract

The high and continuously increasing population growth rate in Indonesia is an important issue as it impacts various sectors including the economy, society, politics, and national defense. Consequently, relevant entities such as the Social Services Department and the National Family Planning Agency (BKKBN) analyze factors associated with population growth rate to formulate policies aimed at achieving balanced population growth. Additionally, population growth prediction is utilized by the Population and Civil Registration Agency (Dispendukcapil) for budget planning and other needs. This research focuses on predicting the population growth rate using the Support Vector Regression (SVR) method and comparing the performance of linear and Gaussian RBF kernels. The study employs a time series dataset of population figures from March 2017 to December 2022. The prediction process involves data normalization, SVR training to obtain updated Lagrange multiplier values, and SVR testing to generate prediction results and error rates using Mean Absolute Percentage Error (MAPE). The testing results indicate an MAPE value of 0.0985% for the linear kernel and 0.38192% for the Gaussian RBF kernel. The findings of this study provide highly accurate prediction results that can support government agencies in population-based planning, budget allocation, and the formulation of long-term demographic policies. The superior performance of the linear kernel demonstrates its potential as a reliable tool for practical decision-making in urban planning and regional development.

Keywords: Population, prediction, support vector regression.

ISSN 2407-4489
pp. 141-149

INTRODUCTION

Population growth is a critical demographic indicator that influences national development, public service planning, resource allocation, and socio-economic stability. The continuous increase in Indonesia's population requires government institutions to anticipate future demographic changes in order to support effective planning in areas such as education, employment, health, and infrastructure. Accurate forecasting of population growth has therefore become essential for agencies such as BKKBN, Dispendukcapil, and regional planning offices in designing informed and sustainable population-related policies.

To address the need for accurate forecasting, machine learning techniques have been increasingly used to model population dynamics. Among these techniques, Support Vector Regression (SVR) is known for its ability to handle non-linear patterns and small-sample datasets effectively. This research applies SVR to predict the population growth rate in Lampung Province and compares the performance of linear and Gaussian RBF kernels to determine the most suitable model for demographic forecasting.

The productive age of a worker is a period in which a person has the ability and opportunity to work. This period generally starts from the age of 15 to 64 years. The productive age of workers is an important factor in determining the economic growth of a country. ((Bappenas), 2021), ((BPS), 2020a).

High economic growth requires a large and qualified workforce. The high productive age of workers indicates that there are many people who are ready to work and contribute to economic growth (Widiastuti & Asnawi, 2020).

The growth rate of the productive age of workers can differ between men and women. This is due to several factors, such as:

a. Gross Participation Rate (APK): Female APKs are generally lower than men. This means that more women are out of work than men.
b. Marriage and Pregnancy: Marriage and pregnancy can affect women's participation in the labor force.
c. Gender Discrimination: Gender discrimination in the workplace can hinder women from reaching their full potential (Widiastuti & Asnawi, 2020).

Lampung Province has a large productive age population. However, the unemployment rate in Lampung is still relatively high. This shows that there is great potential to improve the quality of the workforce in Lampung (BPS), 2020b), (Lampung, 2021).

Support Vector Regression (SVR) is a machine learning method used for regression tasks, which is to predict continuous values based on patterns in the data. SVR is a variant of Support Vector Machines (SVM) that was originally developed for classification problems.

METHOD

2.1 Support Vector Machine (SVM)

The basic principle of SVR is to turn regression problems into optimization problems. The main goal of SVR is to find functions that can map input data into higher feature spaces, where the relationship between input and output becomes more linear. In SVR, data points that are around the decision boundary or within a certain distance of that boundary are called "support vectors". These support vectors are used to build regression models and predict the output values for new data.

Basically, SVR involves several key concepts, including:

a. Kernel functions: Kernel functions are used to map data into higher feature spaces. Kernels can be linear kernels, Gaussian RBF kernels, polynomial kernels, or sigmoid kernels. Kernel functionality allows SVR to handle non-linear data efficiently (Zhou et al., 2021).

ISSN 2407-4489
pp. 141-149

b. Parameters C and ε: Parameter C controls the trade-off between the error penalty and the complexity penalty of the model. The ε parameter controls the width of the corridor (epsilon-tube) around the accepted regression function. These parameters can be adjusted to optimize the performance of the SVR model (Alshammari & Hussain, 2023)
c. Objective function: SVR tries to find regression functions that have as few errors as possible and at the same time have a wide margin. The objective function in the SVR involves penalties for prediction errors and penalties for model complexity (Pham & Tran, 2022)

The kernel function in Support Vector Regression (SVR) is used to map data into higher feature spaces. The kernel function converts the native data space into a feature space that has higher dimensions, where the relationship between input and output becomes more linear. The kernel functionality allows the SVR to handle non-linear data efficiently. Some common types of kernel functions used in SVR include:

a. Linear Kernel: Linear kernel functions are the simplest kernel functions. This function results in a linear mapping of the data into the same feature space. In SVR, a linear kernel is used when the input data has a linear relationship with the output that you want to predict (Alshammari & Hussain, 2023).
b. Gaussian RBF (Radial Base Function) kernel: The Gaussian RBF kernel function uses Gaussian functions to map data into an infinite feature space. This kernel is especially useful when the input data has a complex or non-linear pattern. The Gaussian function of RBF produces a mapping that illustrates the similarity between the input data and the specified central point (Zhou et al., 2021)
c. Polynomial Kernel: The polynomial kernel function maps data into feature spaces that have higher dimensions using polynomial functions. Polynomial kernels are useful when the input data has a non-linear pattern and can be approximated with polynomials (Pham & Tran, 2022)
d. Sigmoid Kernel: The sigmoid kernel function maps data into a higher feature space using sigmoid functions. This function is often used in cases where the input data has binary characteristics (Jain & Kumar, 2021).

Normalization and denormalization is a process used in data processing to change the scale or range of values of data to suit the needs of analysis or modeling. Here is a brief explanation of normalization and denormalization:

Normalization is the process of converting data into a specific or standard range of values. The goal of normalization is to eliminate the difference in scale between various attributes or features in the data, so that attributes with a larger scale do not dominate analysis or modeling. Normalization also helps maintain the stability and convergence of machine learning algorithms (Kaur & Singh, 2021)
Some commonly used normalization methods include:

a. Min-Max Scaling: Converts data into a range of [0, 1] by taking into account the minimum and maximum values of attributes.
b. Z-Score Normalization: Converts data into a normal distribution with a mean of zero and a standard deviation of one.
c. Decimal Scaling: Shifts the decimal point of an attribute so that its value is in the range [0, 1] or [-1, 1] (Rathore et al., 2022)

Denormalization is the process of returning data to the original scale or range of values after normalization. This is useful when predictions or analysis results have been obtained on data that has been normalized, and we want to return the results to their original scale for further interpretation or use.

Denormalization is done by using information about the scale or range of the original value before normalization, such as the minimum and maximum values of attributes. Using this information, the

ISSN 2407-4489
pp. 141-149

normalized data can be converted back to its original scale using the appropriate formula or transformation (Wang & Liu, 2021)(Zhang et al., 2023)

Mean Absolute Percentage Error (MAPE) is an evaluation metric used to measure the relative error rate in making predictions or forecasts. MAPE measures the average percentage of the absolute difference between the actual value and the predicted value, divided by the actual value, and then used as an average.

The formula for calculating MAPE is as follows:

$$MAPE = (1/n) * \Sigma(|(actual - predicted)| / actual) * 100 \quad (1)$$

Where:

- n is the number of samples or observations.
- Σ is a sigma symbol that indicates the total.

MAPE provides information about the average percentage of relative error from the prediction to the actual value. The lower the MAPE value, the better the prediction performance (Ramzan et al., 2023)(Ahmed et al., 2021).

2.2 Data

Population data was obtained from the Indonesian Central Statistics Agency of Lampung Province in March 2017 – December 2022 in the form of time series data.

2.3 Planning

This study aims to compare the performance of linear kernels and Gaussian kernels of RBF in the context of population growth rate prediction. The steps taken in this study are as follows:

a. Calculate the value of the difference between months on population data. This step is carried out to obtain information about the difference in the number of people between months.
b. Normalize data. Population data is normalized to have a specific or standard range of values. Normalization helps in eliminating the difference in scale between attributes in the data.
c. Conduct training on Support Vector Regression (SVR). SVR is used to obtain the update lagrange multiplier value in the last iteration of sequential learning. Sequential learning iterations are performed to update SVR parameters based on training data.
d. Using the update lagrange multiplier value in the SVR testing stage. At this stage, the calculation of the f(x) regression function is carried out to predict the population growth rate.
e. Evaluate predictions by producing predictive values of population growth rate, population differences every month, and MAPE values. The difference in the number of positive population indicates population growth is increasing, while the negative difference indicates population growth is decreasing. MAPE is used as an evaluation metric to measure the relative error rate of a prediction against the actual value.

The Support Vector Regression flowchart is used as a visual representation of the steps taken in the study to predict the population growth rate. Shown in figure 1.

ISSN 2407-4489
pp. 141-149

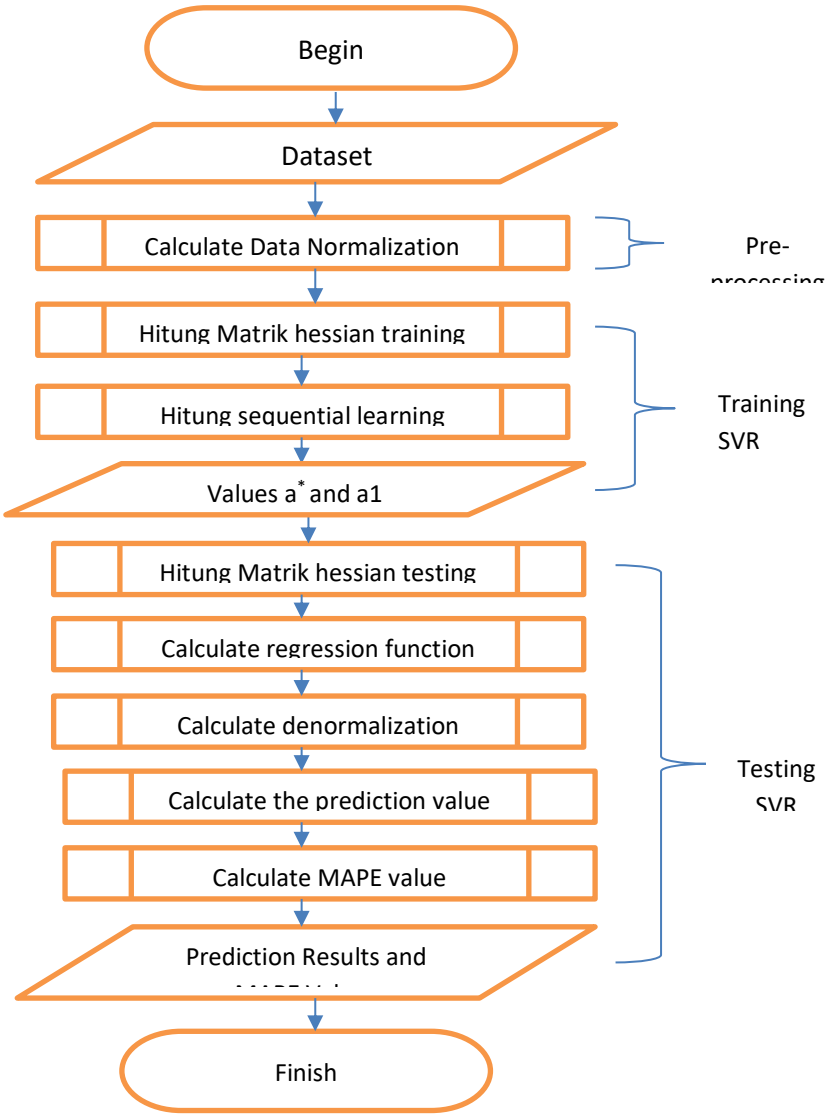


Fig 1. Research flowchart

RESULTS AND DISCUSSION

3.1 Complexity Value Testing

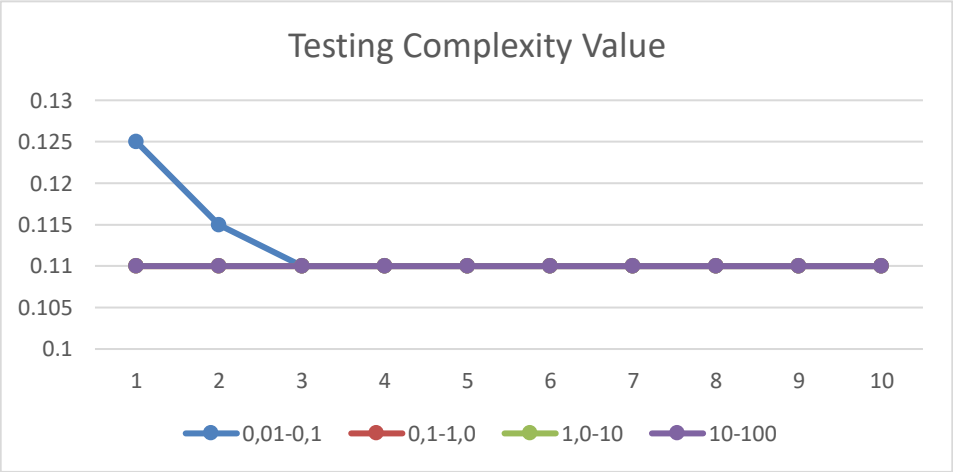


Fig 2. Complexity value testing

ISSN 2407-4489
pp. 141-149

Graphs in the 0.1-1.0 and 1.0-10 value range are invisible because they have the same MAPE value as the graph in the 10-100 value range. The complexity value is a penalty value for data that is outside the area of the ε-incentive loss function or has deviations (Rahman & Kim, 2021). The test results show that the value of the greater complexity does not tolerate errors so that the error rate generated is constant

3.2 Epsilon Value Testing

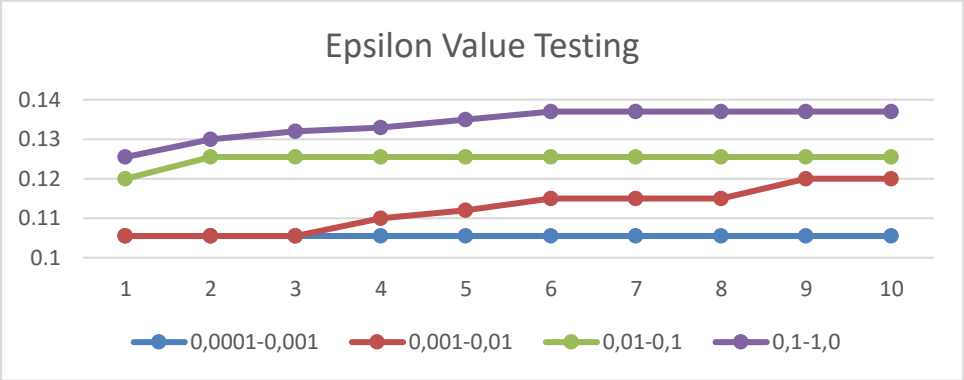


Fig 3. Epsilon value testing

The epsilon value is used to control the width of the ε-incentive loss function which is a representation of the distance between the data point and the hyperplane line. An epsilon value that is too large results in a large error rate because it does not match the data point.

3.3 Constant Value Testing of Learning Rate

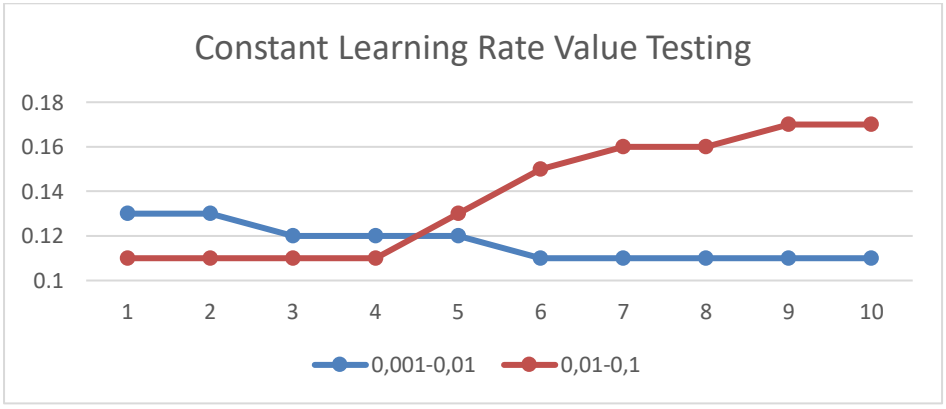


Fig 4. Constant value testing of learning rate

The learning rate constant is a value for calculating the learning rate (gamma) in the sequential learning process. The test results showed that the constant value of the learning rate had an effect on the delta alphastar and delta alpha in the stop condition sequential learning so that it had an impact on the error rate produced.

ISSN 2407-4489
pp. 141-149

3.4 Lambda Value Testing

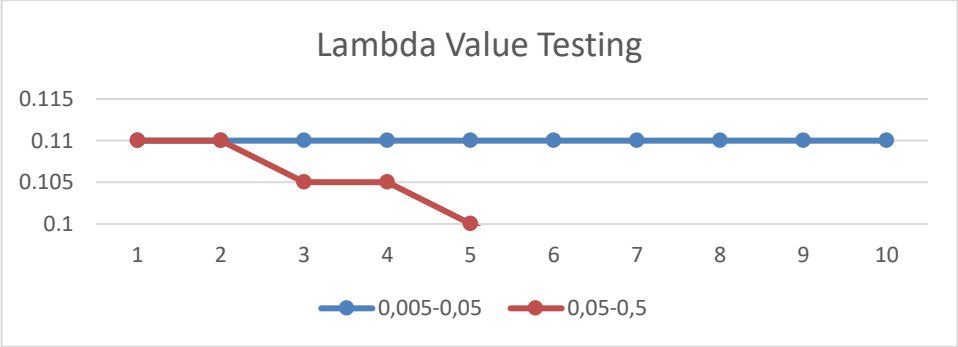


Fig 5. Lambda value testing

The lambda value is the scalar constant used on the kernel mapping space (Tang et al., 2021). Testing on both value ranges resulted in a continuously decreasing MAPE value. A steadily decreasing error rate occurs because a large lambda value causes the convergence to be slow and causes a small error rate.

3.5 Iteration Number Testing

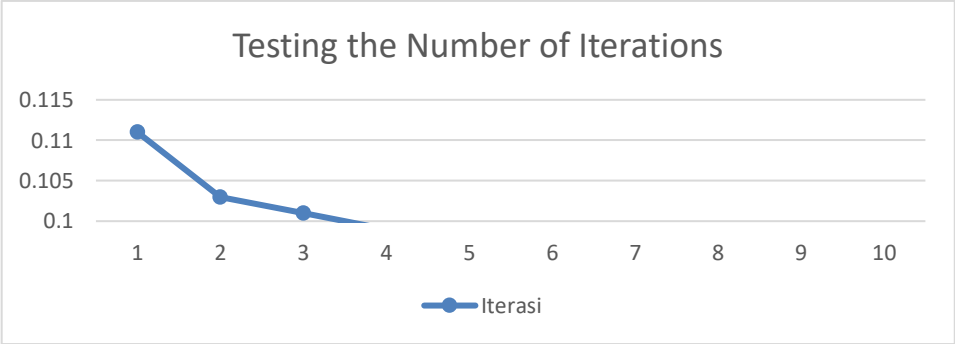


Fig 6. Iteration count testing

Based on the graph of the test results of the number of iterations, the more iterations used, the smaller the error rate. In addition, the value of the error rate also depends on other stop conditions, namely $\max(|\delta\alpha i*|) < \varepsilon$ and $\max(|\delta\alpha i|) < \varepsilon$.

3.6 Input Quantity Testing

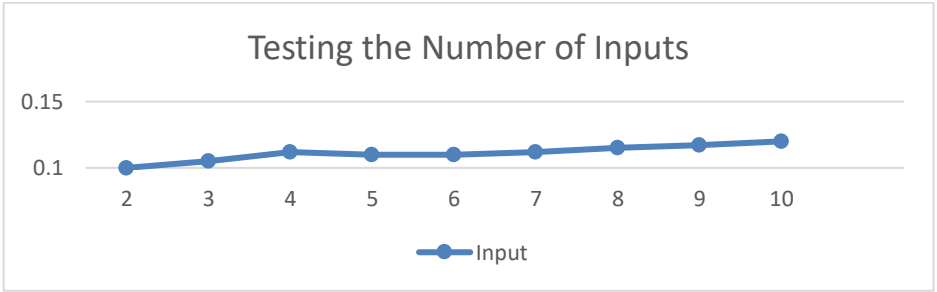


Fig 7. Testing the number of inputs

Based on the graph of the number of inputs tested, many of the inputs used have little effect on the error rate produced.

Overall, the results indicate that the linear kernel produces the lowest MAPE value (0.0985%), significantly outperforming the Gaussian RBF kernel (0.38192%). This suggests that the pattern of population changes in Lampung Province follows a relatively linear trend over time, making the linear

kernel more suitable for capturing such relationships. The RBF kernel, which is typically more effective for non-linear and highly complex datasets, may not provide additional benefits when the underlying data behavior is simple and stable.Although the numerical difference between the two MAPE values may appear small, the improvement is practically meaningful for demographic forecasting, where even slight inaccuracies can influence policy decisions, budget allocation, and population service planning. Accurate short-term predictions are crucial for agencies such as BKKBN and Dispendukcapil, particularly in designing targeted programs, estimating service needs, and projecting administrative workloads.

These findings are consistent with previous studies (Alshammari & Hussain, 2023; Zhou et al., 2021), which reported that linear kernels often outperform non-linear kernels when the dataset exhibits limited fluctuation and low noise. This alignment with earlier research strengthens the argument that SVR with a linear kernel offers a robust and computationally efficient approach for demographic forecasting tasks. The empirical results of this study therefore contribute to improving evidence-based population planning and support the use of SVR models in regional development strategies.

CONCLUSIONS

This study demonstrates that the Support Vector Regression (SVR) method is effective in predicting population growth rates using time series data from Lampung Province. The comparison between kernel functions shows that the linear kernel provides the most accurate prediction with a MAPE value of 0.0985%, outperforming the Gaussian RBF kernel. These findings indicate that the population growth pattern during the observed period is relatively linear, making the linear kernel more suitable for this dataset.

The high level of prediction accuracy suggests that SVR can serve as a valuable tool to support government agencies in population planning, resource allocation, and demographic policy formulation. Future studies may expand the dataset, incorporate additional socio-demographic variables, or compare SVR with other machine learning techniques to further enhance the robustness of population forecasting models.

ACKNOWLEDGMENT

Praise be to the author to the presence of God Almighty for all His abundant graces, gifts, and guidance so that the research entitled "Prediction of Population Growth Rate Using the Support Vector Regression Method" can be completed properly.

The author expresses his highest gratitude and appreciation to the Aisyah Lampung Foundation and Aisyah Pringsewu University for funding this research. And also to BPS Lampung who has provided data for this research.

REFERENCES

Bappenas. (2021). Indonesia vision 2045. Ministry of National Development Planning (Bappenas).

BPS. (2020a). Employment statistics 2020. Statistics Indonesia (BPS).

BPS. (2020b). Employment statistics of Lampung Province 2020. Statistics Indonesia (BPS).

Ahmed, R., Sultana, M., & Rahman, M. (2021). Evaluation of forecasting models: An empirical comparison using MAPE, RMSE, and MAE. Journal of Forecasting and Modeling, 8(1), 42–55. https://doi.org/10.1016/j.jfmodel.2021.02.003

Alshammari, R., & Hussain, M. (2023). Comparative analysis of kernel functions in support vector regression. Applied Artificial Intelligence, 37(3), 199–214. https://doi.org/10.1080/08839514.2023.2168802

Jain, R., & Kumar, D. (2021). Forecasting COVID-19 spread using SVR and polynomial regression. Materials Today: Proceedings, 46, 2897–2903. https://doi.org/10.1016/j.matpr.2021.01.356

Kaur, P., & Singh, M. (2021). Review on data normalization techniques in data mining. International Journal of Engineering Research & Technology (IJERT), 10(6), 1–5.

Lampung Provincial Government. (2021). Lampung Province regional medium-term development plan (RPJMD) 2021–2026. Bappeda Lampung Province.

Pham, T. M., & Tran, D. M. (2022). Enhanced SVR models for non-linear regression in financial time series. Journal of Computational and Applied Mathematics, 405, 113870. https://doi.org/10.1016/j.cam.2021.113870

Rahman, M. M., & Kim, J. M. (2021). Analysis of regularization parameters in support vector regression for time series forecasting. Expert Systems with Applications, 167, 114147. https://doi.org/10.1016/j.eswa.2020.114147

Ramzan, S., Zubair, S., & Khan, M. A. (2023). Performance comparison of machine learning models using error metrics for time series prediction. Procedia Computer Science, 219, 905–912. https://doi.org/10.1016/j.procs.2023.01.120

Rathore, A., Sharma, M., & Bhardwaj, P. (2022). Comparative analysis of normalization methods in machine learning. International Journal of Computer Applications, 184(16), 14–18. https://doi.org/10.5120/ijca2022912515

Tang, Y., Lin, C., & Zhang, H. (2021). An improved SVR model with adaptive regularization parameters for non-linear time series forecasting. Neural Computing and Applications, 33(14), 8443–8455. https://doi.org/10.1007/s00521-021-05987-0

Wang, Y., & Liu, H. (2021). A robust inverse normalization method for recovering original data scale in predictive modeling. Procedia Computer Science, 192, 3701–3709. https://doi.org/10.1016/j.procs.2021.09.145

Widiastuti, R., & Asnawi, R. (2020). Factors that affect the participation of the female labor force in Indonesia. Journal of Economics and Development, 21(1), 1–12.

Zhang, L., Chen, W., & Zhao, Y. (2023). Post-prediction rescaling techniques in time-series machine learning models. Journal of Big Data Analytics, 5(2), 55–68. https://doi.org/10.1186/s41044-023-00229-1

Zhou, T., Wang, L., & Ding, S. (2021). Support vector regression with missing data imputation for time series prediction. Expert Systems with Applications, 167, 114160. https://doi.org/10.1016/j.eswa.2020.114160