RESEARCH ARTICLE • OPEN ACCESS

Forecasting of Rice Harvest Results Using SVR Modeling Techniques

Devie Rosa Anamisa et al.

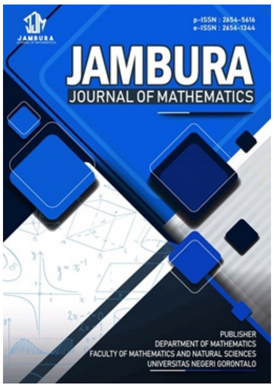
JAMBURA
JOURNAL OF MATHEMATICS

Volume 7, Issue 1, Pages 84–91, February 2025

Received 25 October 2024, Revised 20 February 2025, Accepted 24 February 2025, Published 27 February 2025

To Cite this Article : D. R. Anamisa et al.,"Forecasting of Rice Harvest Results Using SVR Modeling Techniques", *Jambura J. Math*, vol. 7, no. 1, pp. 84–91, 2025, *https://doi.org/10.37905/jjom.v7i1.30592*

JOURNAL INFO • JAMBURA JOURNAL OF MATHEMATICS



Homepage	:	http://ejurnal.ung.ac.id/index.php/jjom/index
Journal Abbreviation	:	Jambura J. Math.
Frequency	:	Biannual (February and August)
Publication Language	:	English (preferable), Indonesia
DOI	:	https://doi.org/10.37905/jjom
Online ISSN	:	2656-1344
Editor-in-Chief	:	Hasan S. Panigoro
Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
Country	:	Indonesia
OAI Address	:	http://ejurnal.ung.ac.id/index.php/jjom/oai
Google Scholar ID	:	iWLjgaUAAAAJ
Email	:	info.jjom@ung.ac.id

Jambura Journal of Mathematics, Volume 7, Issue 1, Pages 84–91, February 2025 *https://doi.org/10.37905/jjom.v7i1.30592*

Research Article

Check for updates

Forecasting of Rice Harvest Results Using SVR Modeling Techniques

Devie Rosa Anamisa[1,*], Bain Khusnul Khotimah[1], Mohammad Yanuar Hariyawan[2], Firli Irhamni[1], Achmad Jauhari[1], Fifin Ayu Mufarroha[1], Dina Violina[1], and Dinah Nuraini[1]

[1] *Informatics Engineering Department, University of Trunojoyo Madura, Bangkalan, Indonesia*
[2] *Electrical Engineering Department, Telkom University, Surabaya, Indonesia*

ARTICLE HISTORY
Received 25 October 2024
Revised 20 February 2025
Accepted 24 February 2025
Published 27 February 2025

KEYWORDS
Forecasting
Time Series Data
Rice Harvest Results
Modeling
SVR

ABSTRACT. *Forecasting is an activity that predicts future values by utilizing existing track record data. The object of this study is rice plants because they are the primary food source for the Indonesian people. Every year, the Government strives for rice farmers throughout Indonesia to produce abundant rice harvests to meet the community's food needs. Therefore, rice farmers need a system that can predict their rice harvests to obtain information about future harvests to find out whether their harvests have decreased or increased so that they can determine efforts that can be made in the future and can be used as a policy maker for the Government in maintaining the national food security chain. This study uses time series data on rice harvests in Pamekasan, Madura, for 2007-2023 using the Support Vector Regression (SVR) model. The results of several trials have shown that the application of the SVR model for forecasting rice harvests in 2024 has produced good accuracy with a relatively low MAPE error rate of 3.97%, and the rice harvest has reached an average prediction of 15470.08 tons with an average actual data of 7937.884 tons. Therefore, applying this SVR model can be recommended for predicting future rice harvests.*

1. Introduction

Forecasting is the activity of predicting future results by utilizing existing track record data [1]. In a company usually do forecasting to predict the future conditions that are still unknown for the progress of the company. Indonesia is one of the agricultural countries with most of its population being farmers. And most of the staple food of the Indonesian population is rice [2]. Therefore, rice production is a benchmark for food availability in Indonesia, and efforts are needed to increase rice production in various regions, one of which is Madura. In 2022, rice production in Madura reached an average of 19215.12 tons of dry milled grain (DMG), while in 2023, it decreased by 1782.25 tons of DMG [3]. The decline in rice production that year consecutively raised concerns among farmers in Madura about their production results in the coming year. This is because the natural phenomenon currently occurring in the region has had an impact of prolonged drought, so rice farmers in the Madura area almost experienced crop failure. Therefore, this study attempts to help overcome concerns about harvest results in the Madura region in 2024 by designing a forecasting system to provide a picture of predicted rice production results in 2024 using data from previous years.

Several previous studies, forecasting systems have applied several forecasting methods, such as: research conducted by Fendiyanto [4] regarding the forecasting of rice production results in Mulawarman Village using the trend moment method and processing the prediction using several influencing factors such as land area, climate, labor, fertilizer, and seeds used so that it reaches a Mean Absolute Percentage Error (MAPE) value of 26.20%. In addition, the research conducted by Prasetyo et al. [5] regarding the prediction of soybean sales using a Double Moving Average (DMA) has produced a MAPE of 14.67%. Meanwhile, in the research conducted by Mardhika et al. [6] regarding forecasting rice harvest results in Malang Regency, East Java, applying SVR has produced a smaller MAPE than the trend moment method and DMA by 10.133%. In the research conducted by Suyono et al. [7] regarding consumer index price predictions [8] for food commodities in Surabaya City, the SVR kernel polynomial has produced the best error rate, with a MAPE value of 4.31%.

Based on the application of forecasting methods that several previous studies have applied, it has been shown that applying the SVR method for forecasting has produced a reasonably small MAPE. Therefore, this study developed a forecasting system for rice production results using the SVR method. The purpose of this study is to be able to solve forecasting problems with prediction results that are almost accurate with a small error rate so that it can improve strategies both in planning and developing future rice production by anticipating, preparing, and estimating all supporting factors for harvest results so that rice harvest results do not change or decrease.

2. Methods

Forecasting plays a vital role in effective planning and targeted policy-making. The forecasting process involves collecting historical data and processing the data using forecasting methods. This study uses rice harvest data from the Pamekasan Re-

*Corresponding Author.

Email : *devros_gress@trunojoyo.ac.id* (D. R. Anamisa)
Homepage : http://ejurnal.ung.ac.id/index.php/jjom/index / E-ISSN : 2656-1344

Table 1. Dataset of rice harvest

Years	Month (Tons)											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
2007	87587	10216	68573	64441	99134	92981	9195	8716	9253	14805	5645	2382
2008	68098	13279	45002	94039	41325	82608	8493	8886	8322	10139	10275	6908
2009	41467	20201	72605	30171	51146	52604	4282	8649	8254	11862	12878	4512
2010	86067	75732	97678	34906	23519	27760	9569	2859	5721	92735	13921	2334
…	…	…	…	…	…	…	…	…	…	…	…	…
2020	448	1618	14486	112783	32976	3565	18292	25742	14946	1537	1984	596
2021	290	11135	99717	45945	3305	10242	18650	22013	6489	3735	453	438
2022	3403	20342	110793	32538	2510	16358	17739	11320	4999	6529	3554	491
2023	2574	18790	19743	39509	11847	18697	14172	19121	15379	17894	11894	19571

gency Agriculture Service, Madura, from January 2007 to December 2023. The stages of research in this study are data collection, data processing, and the development of techniques with prediction models using SVR [9], and evaluation of the prediction model that has been built. Visualization of the research method is shown in Figure 1.

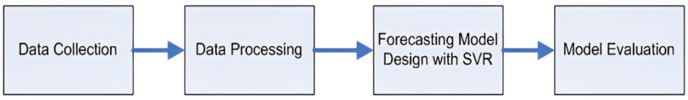


Figure 1. Research methods

In this study, the data collection process was carried out by conducting observations at the Department of Agriculture to collect time series data from rice harvests in the Pamekasan area, Madura, as in Table 1. Then, data was processed with a data normalization process. Normalization functions to change raw data into valuable and efficient data because raw data is often incomplete and inconsistent [10]. Several processes have been carried out in data normalization, such as data cleaning, transforming, and reducing data. In this study, the data normalization process is carried out by applying the Z-Score method. The Z-Score Normalization method is used to change a data set into a standard scale so that it has a mean of zero and a standard deviation of one [11, 12]. The Z-Score method formula can be seen in eq. (1) with the aim of comparing the quality of achievement or target of data with the average distribution of data in a group based on the standard deviation value [13]. In the final process of this forecast, the Z-Score method is also used for the denormalization process. This denormalization process is to return the value to the beginning of the data before data normalization is carried out [14]. The data denormalization formula can be seen in eq. (2). This denormalization process is carried out to see the prediction results.

$$z = \frac{x-\mu}{\sigma}, \quad (1)$$

$$x = (z * \sigma) + \mu, \quad (2)$$

with

$$\sigma = \frac{1}{n-1}\sum_{i=1}^{n}\left(x-\overline{x}^2\right)^{\frac{1}{2}},$$

where μ is the mean of the data set, σ is the standard deviation of the data set with n is lots of data, $\overline{x}$ is average of x data, and z is the normalized Z-Score data.

In this research, the Z-Score method has been chosen for data normalization because there are often different ranges of values in each feature attribute in the dataset. The difference in the range of values that is quite far from the existing attributes causes the role of the attributes in the dataset not to function optimally. Z-Score uses the mean and standard deviation for each feature attribute by changing the value scale of the data for the normalization process. In addition, Z-Score is able to reduce the effect of outliers. The process of raw rice harvest dataset for forecasting by applying Z-Score normalization involves several steps, including:

1. Preparing the rice harvest dataset in a structured form such as a table then separating the features and targets, as in table 1.
2. Calculating the Mean and Standard Deviation (Std) and applying eq. (1) and eq. (2).
3. Carrying out the process of separating Training and Testing Data, where training data is used to train the forecasting model and testing data is used to test model performance.
4. Implementing the forecasting model using the machine learning method.

Meanwhile, in the modeling process for this forecasting, it is done by applying the SVR method. The SVR method is a development method of Support Vector Machine (SVM) with the best kernel as an aid in solving production forecasting problems and the forecast results will be the basis for planning in the coming year by dividing data and finding the best model [15, 16]. The process description of the SVR method for forecasting rice harvest yields can be seen in Figure 2.

In this description, there are several steps in the harvest yield forecasting process, including [17]:

1. Determine the data set which is divided into training data and testing data.
2. Calculating the distance between training data, which is used for calculating the Radial Basis Function (RBF) kernel. Based on the application of the RBF kernel to SVR, this kernel is able to produce a model from training data that is finite in the spatial dimension [18]. The distance calculation in this study uses the Euclidean distance formula, as in eq. (3).

$$d = \|x_i - x_j\|^2, \quad (3)$$

where d is the distance between the data, $x_i$is the ith test data and x_j is the jth train data.

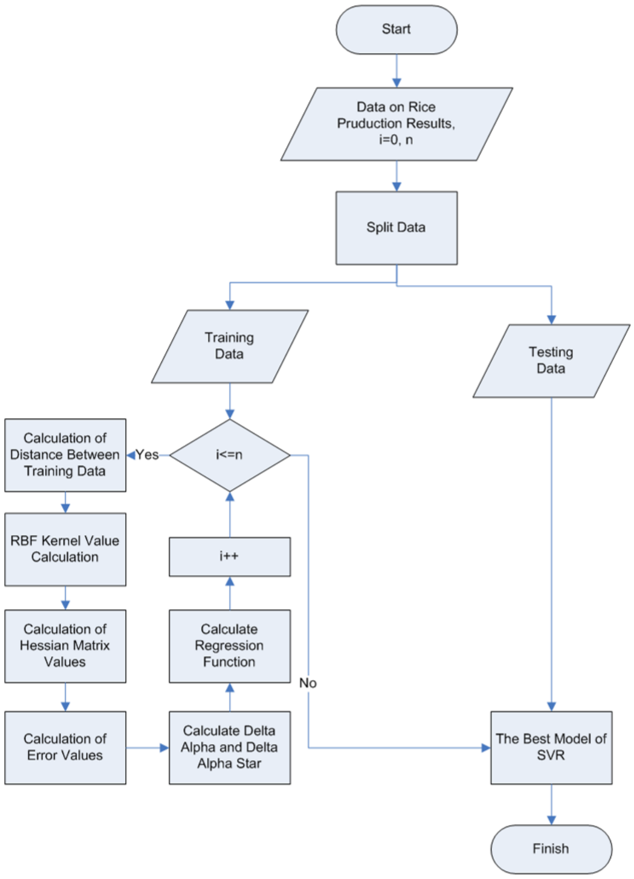


Figure 2. Process of SVR to forecasting of rice harvest

3. Calculating the RBF kernel value, where the kernel function is the most important part of the SVR method. The kernel itself is an algorithm used for pattern analysis and recognition. The Gaussian RBF kernel is calculated using eq. (4), where the sigma parameter (σ) can be adjusted and functions in the performance of the kernel itself so that it must be initialized according to the problem to be solved. If it is too high, the exponential will behave almost linearly and the higher dimensional projection will start to lose its non-linear power.

$$K\left(x_i, x_j\right) = \exp\left(-\frac{d}{2\sigma^2}\right), \tag{4}$$

where $K\left(x_i, x_j\right)$ is how similar two data points $x_i$and x_j are and σ is the sigma value. In the RBF kernel that uses the γ (gamma) function that determines how far the influence of one data point on another. If γ is large, each data point only affects its very close neighbors, while if γ is small, the influence is wider.
Calculate the Hessian Matrix using eq. (5). The output of the Hessian matrix is the gamma parameter value which is used in the next stage, namely the sequential learning process. The gamma parameter value can be calculated using eq. (6).

$$R_{ij} = (K\left(x_i, x_j\right) + \alpha^2), \tag{5}$$

$$\gamma = \frac{Constanta\ Learning\ Rate}{\max(Matrix\ Hessian)}, \tag{6}$$

where R_{ij} is the Hessian matrix of row i and column j, γ is the Learning Rate to control the speed of the learning process, and α is lambda. The sequential learning process is a process that exists in every calculation of the SVR function which aims to obtain an optimal dividing line or hyperplane.

4. On the training data, the error value is also calculated using eq. (7), the change in the Lagrange Multiplier value is calculated using eq. (8) and the new Lagrange Multiplier value, namely α_i and α_i^* the updated value and using eq. (9).

$$E_i = y_i = \sum_{i=1}^{n} (\alpha_i^* - \alpha_i) R_{ij}, \tag{7}$$

$$\delta_{ai*} = MIN\{MAX\left(\gamma\left(E_i - \varepsilon\right), -\alpha_i^*\right), C - \alpha_i^*\}, \tag{8}$$

$$\alpha_i^*\left(baru\right) = \delta_{\alpha i*} + \alpha_{i*}, \tag{9}$$

where E_i is the i-th error value, y_i is the value of the actual data, α_i^* is the upper limit of the langrange multipliers, α_i is the lower limit of the langrange multipliers and R_{ij} is the hessian matrix of the i-th row and the j-th column. And then δ_{ai*} is the change in the upper bound value, and C is the complexity.

5. Calculate the regression function with eq. (10). The regression function is the function that has the largest deviation from the actual target, for all training data. The regression function is used to find a function as a hyperplane (separating line). If the value is equal to 0, then a perfect regression equation is obtained [19].

$$f\left(x\right) = \sum_{i=1}^{l} (\alpha_{i*} - \alpha_i)(K\left(x_i, x_j\right) + \propto^2, \tag{10}$$

$$MAX\left(|\delta\alpha_i^*|\right) < \varepsilon\ and\ MAX\left(|\delta\alpha_{i|} < \varepsilon\right),$$

where $f\left(x\right)$ is a regression function, α_i^* is the upper limit of the Lagrange multiplier, α_i is the lower bound of the Lagrange multiplier, $\delta\alpha_{i*}$ is the change in the upper limit value,$\delta\alpha_i$ is the change in the lower limit value, and ε is epsilon. MAX shows that sequential learning calculations are carried out iteratively until convergence is achieved.

The model evaluation process in this study is used to measure how well the SVR prediction method performs in predicting data. One type of measurement of the accuracy of this forecasting method is the Mean Absolute Percentage Error (MAPE). MAPE is a method of measuring relative error values using absolute value measurements [20, 21]. The use of absolute values in MAPE calculations has two advantages, namely absolute values keep the calculation results positive and MAPE allows for comparing accuracy results between time series data of different scales because MAPE does not depend on the dependent variable. The calculation of MAPE values is shown in eq. (11).

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} |\frac{y_i - \hat{y}_i}{y_i}| \times 100\%, \tag{11}$$

where y_i is actual data, $\hat{y}_i$ is the target data and n is the amount of data. The MAPE value can be interpreted into 4 categories, namely if MAPE $\leq$ 10% then the forecasting model performance is accurate, if 10% $<$ MAPE $\leq$ 20% then the forecasting model performance is good, if 20% $<$ MAPE $\leq$ 50% then the forecasting model performance is decent and if MAPE $>$ 50% then the forecasting model performance is poor.

Table 2. Normalization results of rice harvest data with Z-score

Data	X1	X2	X3	X4	Z	Data	X1	X2	X3	X4	Z
1	1.22	-0.74	0.74	0.63	1.52	185	-0.55	-0.094	-0.048	-0.26	-0.46
2	-0.74	0.74	0.63	1.52	1.36	186	-0.094	-0.048	-0.26	-0.46	-0.41
3	0.74	0.63	1.52	1.36	-0.76	187	-0.048	-0.26	-0.46	-0.41	-0.51
4	0.63	1.52	1.36	-0.76	-0.78	188	-0.26	-0.46	-0.41	-0.51	-0.61
5	1.52	1.36	-0.76	-0.78	-0.76	189	-0.46	-0.41	-0.51	-0.61	-1.74
6	1.36	-0.76	-0.78	-0.76	-0.62	190	-0.41	-0.51	-0.61	-1.74	0.15
7	-0.76	-0.78	-0.76	-0.62	-0.85	191	-0.51	-0.61	-1.74	0.15	0.27
8	-0.78	-0.76	-0.62	-0.85	-0.94	192	-0.61	-1.74	0.15	0.27	2.58
9	-0.76	-0.62	-0.85	-0.94	1.08	193	-1.74	0.15	0.27	2.58	-0,65
10	-0.62	-0.85	-0.94	1.08	-0.61	194	0.15	0.27	2.58	-0,65	0,14
11	-0.85	-0.94	1.08	-0.61	0.36	195	0.27	2.58	-0,65	0,14	-0.38
12	-0.94	1.08	-0.61	0.36	1.88	196	2.58	-0,65	0,14	-0.38	0.19
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

3. Results and Discussion

At this stage, several steps have been taken to predict future production results. The first step is to look at the development of rice harvest data in 2007-2023 in Pamekasan Regency, Madura, as in Table 1. From the results of observations of rice harvests in Madura, especially Pamekasan, the data has been classified as a time series with an increase almost every year, where a series of observations are sorted based on a fluctuating time series every year with the same distance. Based on descriptive analysis to see a general picture of the distribution of research data, as in Figure 3. From the data graph, it can be concluded that over the past 16 years from 2007 - 2023, rice harvests in Pamekasan have decreased with very diverse distribution characteristics, this is indicated by a large standard deviation value of 25724.41 and the Median of the data obtained a value of 14565.705.

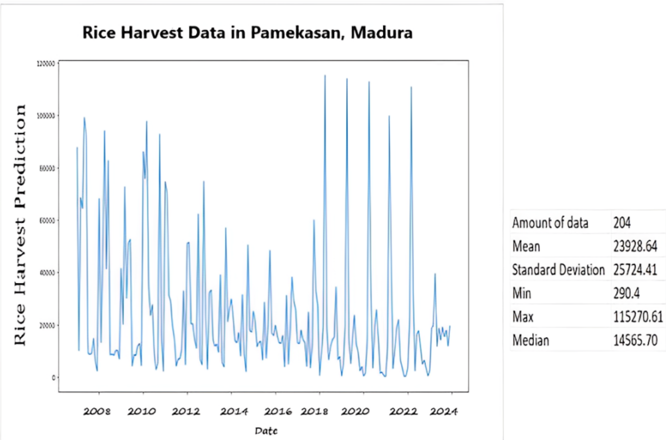


Figure 3. Time series plot of rice harvest data

The next step is the data normalization process. In the process of normalizing rice harvest data, this is done by forming a database structure so that most of the ambiguity can be eliminated. In this study, Z-Score was used according to eq. (1). The results of normalization using z-score in this study are in the range of -1 to 3, as can be seen in Table 2. The results of normalization using z-score are often in the range of -1 to 3. The normalization value can exceed this range. Values that exceed this range are usually values that rarely appear or are too far from most values. This range of values is formed because it looks at the average relationship of all values to their standard deviations with the aim of being able to produce a balance of comparative values between data before and after the process.

Before the forecasting process is carried out with the SVR model, the data is divided or grid searched into 90% training data and 10% as testing data, where the train data is used to build a model with the SVR method. While the test data is used to predict with the SVR method based on the model that has been obtained previously. SVR training involves several stages, such as calculating the distance between training data, calculating the RBF kernel, calculating the hessian matrix, sequential learning, calculating the $f(x)$ value to determine the forecast value, denormalization to return the value to its original range, and testing the training data. The results of the SVR training process can be seen in Table 3. Table 3 shows the prediction results with parameters sigma $(\sigma) = 0.1$, Alpha $(\alpha) = 0.412$, Gamma $(\gamma) = 4.784$, Cost $(C) = 0.082$, and Epsilon $(\varepsilon) = 0.083$ which shows that model optimization has obtained maximum results by looking at the best parameters and the most minimal error. These results indicate that optimizing the SVR model with the RBF kernel type requires parameters according to its kernel.

After the best prediction model has been generated from the training process that has been carried out previously, the model will be used to validate the model with the existing validation data to see the learning performance of the training process that has been carried out. If the model obtained produces a poor forecast of its validation data, then the learning process is repeated until an appropriate model is obtained. Overfitting in SVR occurs when the model learns too much detail from the training data, such as in the parameter C that is too large, because the parameter C in SVR is used to control the trade-off between minimizing the error in the training data and keeping the model simple (regularization). In addition, the gamma parameter that is too large for the Gaussian kernel can also affect the distance between data points and other data points so that it tends to be overfitting. To overcome this, it can be handled when the machine learning model provides accurate predictions on the training data but not for the testing data. In this study, the evaluation process is used to determine the accuracy of the fore-

Table 3. Result of SVR method training process

Step by Step	Training Data										
	Data	1	2	3	4	5	···	182	183	184	185
Distance Calculation between Training Data	1	0	6.87	3,27	7,83	8,82	···	9.81	18.5	7.55	1.82
	2	6.87	0	3,02	8,28	12,79	···	8.35	11.0	19.7	6.06
	3	3.27	3.02	0	5,36	10,98	···	5.68	11.9	12.5	6.10
	4	7.83	8.28	5,36	0	5,35	···	8.42	4.13	11.9	6.97
	5	8.82	12.7	10,9	5,35	0	···	23.2	7.13	4.24	5.82
	⋮	⋮	⋮	⋮	⋮	⋮	⋱	⋮	⋮	⋮	⋮
	180	10.6	4.83	8.84	23.4	24.7	···	18.1	28.7	26.9	11.8
	181	9.81	8.35	5.68	8.42	23.2	···	0	19.0	28.5	12.4
	182	18.5	11.0	11.9	4.13	7.13	···	19.0	0	18.9	14.8
	183	7.55	19.7	12.5	11.9	4.24	···	28.5	18.9	0	8.89
	184	1.82	6.06	6.10	6.97	5.82	···	12.4	14.8	8.8	0
	Data	1	2	3	4	5	···	182	183	184	185
RBF Kernel Value Calculation	1	1	0.96	0.98	0.96	0.96	···	0.95	0.91	0.96	0.99
	2	0.96	1	0.99	0.95	0.93	···	0.95	0.94	0.90	0.97
	3	0.98	0.98	1	0.97	0.94	···	0.97	0.94	0.93	0.96
	4	0.96	0.95	0.97	1	0.97	···	0.95	0.97	0.94	0.96
	5	0.95	0.93	0.94	0.97	1	···	0.89	0.96	0.97	0.97
	⋮	⋮	⋮	⋮	⋮	⋮	⋱	⋮	⋮	⋮	⋮
	180	0.94	0.97	0.95	0.88	0.88	···	0.91	0.86	0.87	0.94
	181	0.95	0.95	0.97	0.95	0.89	···	1	0.90	0.86	0.93
	182	0.91	0.94	0.94	0.97	0.96	···	0.90	1	0.90	0.92
	183	0.96	0.90	0.93	0.94	0.97	···	0.86	0.90	1	0.95
	184	0.99	0.97	0.96	0.96	0.97	···	0.93	0.92	0.95	1
	Data	1	2	3	4	5	···	182	183	184	185
Hessian Matrix Value Calculation	1	1.17	1.14	1.15	1.13	1.12	···	1.12	1.08	1.13	1.16
	2	1.13	1.17	1.15	1.12	1.10	···	1.12	1.11	1.07	1.14
	3	1.15	1.15	1.17	1.14	1.11	···	1.14	1.11	1.10	1.13
	4	1.13	1.13	1.14	1.17	1.14	···	1.12	1.14	1.11	1.13
	5	1.12	1.11	1.11	1.14	1.17	···	1.06	1.13	1.14	1.14
	⋮	⋮	⋮	⋮	⋮	⋮	⋱	⋮	⋮	⋮	⋮
	180	1.11	1.14	1.12	1.05	1.05	···	1.08	1.03	1.04	1.11
	181	1.12	1.12	1.14	1.12	1.06	···	1.17	1.07	1.03	1.10
	182	1.08	1.11	1.11	1.14	1.13	···	1.07	1.17	1.07	1.09
	183	1.13	1.07	1.10	1.11	1.14	···	1.03	1.07	1.17	1.12
	184	1.16	1.14	1.13	1.13	1.14	···	1.10	1.09	1.12	1.17
	Data	E_i	$\delta^*_{\alpha i}$	$\delta_{\alpha i}$	$f(x)$	Data		E_i	$\delta^*_{\alpha i}$	$\delta_{\alpha i}$	$f(x)$
Calculating Error Value, $\delta\alpha^*$ Value, $\delta\alpha$ Value, and $f(x)$	1	1.52	0.08	0	1.34	40		-0.97	0	0.07	1.35
	2	1.36	0.08	0	1.29	41		-0.89	0	0.06	1.35
	3	-0.76	0	0.05	1.30	42		1.41	0.08	0	1.35
	4	-0.78	0	0.05	1.31	43		-0.67	0	0.05	1.31
	5	-0.76	0	0.05	1.34	44		-0.92	0	0.07	1.31
	⋮	⋮	⋮	⋮	⋮	⋮		⋮	⋮	⋮	⋮
	35	1.55	0.08	0	1.31	180		0.43	0.02	0	1.25
	36	-0.11	0	0.003	1.27	181		-0.55	0	0.0393	1.25
	37	-0.42	0	0.028	1.32	182		-0.09	0	0.0008	1.26
	38	-0.30	0	0.019	1.33	183		-0.04	0	0	1.34
	39	-0.79	0	0.059	1.36	184		-0.26	0	0.0148	1.36

cast results that have been carried out on the actual data with the influence of the SVR parameters. Training data and testing data are commonly used in machine learning. The machine is given a group of datasets to learn and is called training data, then the learning results will be used to process a new dataset called testing data. This study conducted experiments with five different percentages of training data and testing data, namely with percentages of 50%:50%, 60%:40%, 70%:30%, 80%:20%, 90%:10%. The

Table 4. Grid search result parameters with RBF kernel type

Data Split		Best Parameters			MAPE	Data Split		Best Parameters			MAPE
Training Data (%)	Testing Data (%)	γ	C	ε	(%)	Training Data (%)	Testing Data (%)	γ	C	ε	(%)
50	50	0.084	0.0826	0.083	12.52	50	50	4.784	2.457	0.083	16.98
60	40	0.084	0.0826	0.083	13.45	60	40	4.784	2.457	0.083	16.53
70	30	0.084	0.0826	0.083	13.33	70	30	4.784	2.457	0.083	17.08
80	20	0.084	0.0826	0.083	7.84	80	20	4.784	2.457	0.083	13.89
90	10	0.084	0.0826	0.083	7.01	90	10	4.784	2.457	0.083	21.66
50	50	0.784	0.0826	0.083	11.20	50	50	4.784	2.782	0.083	17.42
60	40	0.784	0.0826	0.083	11.81	60	40	4.784	2.782	0.083	16.87
70	30	0.784	0.0826	0.083	12.64	70	30	4.784	2.782	0.083	17.49
80	20	0.784	0.0826	0.083	7.80	80	20	4.784	2.782	0.083	14.73
90	10	0.784	0.0826	0.083	7.04	90	10	4.784	2.782	0.083	21.99
50	50	1.284	0.0826	0.083	11.03	50	50	4.784	3.107	0.083	17.64
60	40	1.284	0.0826	0.083	11.64	60	40	4.784	3.107	0.083	17.05
70	30	1.284	0.0826	0.083	12.66	70	30	4.784	3.107	0.083	17.87
80	20	1.284	0.0826	0.083	7.63	80	20	4.784	3.107	0.083	15.41
90	10	1.284	0.0826	0.083	6.90	90	10	4.784	3.107	0.083	22.37
50	50	1.784	0.0826	0.083	10.85	50	50	4.784	3.432	0.083	17.93
60	40	1.784	0.0826	0.083	11.46	60	40	4.784	3.432	0.083	17.24
70	30	1.784	0.0826	0.083	12.61	70	30	4.784	3.432	0.083	18.28
80	20	1.784	0.0826	0.083	7.52	80	20	4.784	3.432	0.083	15.70
90	10	1.784	0.0826	0.083	6.73	90	10	4.784	3.432	0.083	22.80
50	50	2.284	0.0826	0.083	10.81	50	50	4.784	3.757	0.083	18.24
60	40	2.284	0.0826	0.083	11.39	60	40	4.784	3.757	0.083	17.48
70	30	2.284	0.0826	0.083	12.59	70	30	4.784	3.757	0.083	18.74
80	20	2.284	0.0826	0.083	7.36	80	20	4.784	3.757	0.083	15.99
90	10	2.284	0.0826	0.083	6.51	90	10	4.784	3.757	0.083	23.36
50	50	2.784	0.0826	0.083	10.85	50	50	4.784	4.082	0.083	18.54
60	40	2.784	0.0826	0.083	11.38	60	40	4.784	4.082	0.083	17.72
70	30	2.784	0.0826	0.083	12.54	70	30	4.784	4.082	0.083	19.11
80	20	2.784	0.0826	0.083	7.27	80	20	4.784	4.082	0.083	16.23
90	10	2.784	0.0826	0.083	6.46	90	10	4.784	4.082	0.083	23.75
50	50	3.284	0.0826	0.083	10.86	50	50	4.784	0.0826	0.083	10.79
60	40	3.284	0.0826	0.083	11.34	60	40	4.784	0.0826	0.083	11.26
70	30	3.284	0.0826	0.083	12.50	70	30	4.784	0.0826	0.083	12.33
80	20	3.284	0.0826	0.083	7.21	80	20	4.784	0.0826	0.083	7.07
90	10	3.284	0.0826	0.083	6.25	90	10	4.784	0.0826	0.083	5.98
50	50	3.784	0.0826	0.083	10.85	50	50	4.784	0.0826	2.183	49.86
60	40	3.784	0.0826	0.083	11.30	60	40	4.784	0.0826	2.183	51.74
70	30	3.784	0.0826	0.083	12.45	70	30	4.784	0.0826	2.183	67.24
80	20	3.784	0.0826	0.083	7.15	80	20	4.784	0.0826	2.183	44.47
90	10	3.784	0.0826	0.083	6.11	90	10	4.784	0.0826	2.183	48.08
50	50	4.284	0.0826	0.083	10.83	50	50	4.784	0.0826	4.283	49.86
60	40	4.284	0.0826	0.083	11.28	60	40	4.784	0.0826	4.283	51.74
70	30	4.284	0.0826	0.083	12.39	70	30	4.784	0.0826	4.283	67.20
80	20	4.284	0.0826	0.083	7.10	80	20	4.784	0.0826	4.283	45.05
90	10	4.284	0.0826	0.083	6.02	90	10	4.784	0.0826	4.283	48.76
50	50	4.784	0.0826	0.083	10.79	50	50	4.784	0.0826	6.383	49.86
60	40	4.784	0.0826	0.083	11.26	60	40	4.784	0.0826	6.383	49.86
70	30	4.784	0.0826	0.083	12.33	70	30	4.784	0.0826	6.383	51.74
80	20	4.784	0.0826	0.083	7.07	80	20	4.784	0.0826	6.383	67.20
90	10	4.784	0.0826	0.083	**3.97**	90	10	4.784	0.0826	6.383	45.05
50	50	4.784	0.0826	0.083	10.79	50	50	4.784	0.0826	8.483	51.74
60	40	4.784	0.0826	0.083	11.26	60	40	4.784	0.0826	8.483	67.20

Table 4. Grid search result parameters with RBF kernel type (Continued)

Data Split		Best Parameters			MAPE (%)	Data Split		Best Parameters			MAPE (%)
Training Data (%)	Testing Data (%)	γ	C	ε		Training Data (%)	Testing Data (%)	γ	C	ε	
70	30	4.784	0.0826	0.083	12.33	70	30	4.784	0.0826	8.483	45.05
80	20	4.784	0.0826	0.083	7.07	80	20	4.784	0.0826	8.483	48.76
90	10	4.784	0.0826	0.083	5.97	90	10	4.784	0.0826	8.483	49.86
50	50	4.784	1.482	0.083	13.92	50	50	4.784	0.0826	10.58	49.86
60	40	4.784	1.482	0.083	13.07	60	40	4.784	0.0826	10.58	51.74
70	30	4.784	1.482	0.083	14.92	70	30	4.784	0.0826	10.58	67.20
80	20	4.784	1.482	0.083	10.33	80	20	4.784	0.0826	10.58	45.05
90	10	4.784	1.482	0.083	19.99	90	10	4.784	0.0826	10.58	48.76
50	50	4.784	1.807	0.083	15.13	50	50	4.784	0.0826	12.68	49.86
60	40	4.784	1.807	0.083	14.53	60	40	4.784	0.0826	12.68	51.74
70	30	4.784	1.807	0.083	15.71	70	30	4.784	0.0826	12.68	67.20
80	20	4.784	1.807	0.083	11.63	80	20	4.784	0.0826	12.68	45.05
90	10	4.784	1.807	0.083	20.98	90	10	4.784	0.0826	12.68	48.76
50	50	4.784	2.132	0.083	16.15	50	50	4.784	0.0826	14.78	49.86
60	40	4.784	2.132	0.083	15.71	60	40	4.784	0.0826	14.78	51.74
70	30	4.784	2.132	0.083	16.53	70	30	4.784	0.0826	14.78	67.20
80	20	4.784	2.132	0.083	12.87	80	20	4.784	0.0826	14.78	45.05
90	10	4.784	2.132	0.083	21.35	90	10	4.784	0.0826	14.78	48.76

results of the SVR process with several of the best parameters and with the division of data can be seen in Table 4. From these results, it shows that the best model in the division of 90% training data and 10% testing data with γ of 4.784, $C = 0.0826$ and $\varepsilon = 0.083$ with the kernel parameters used are the RBF kernel and the MAPE (Mean Absolute Percentage Error) value obtained is 3.97%. These results indicate that the prediction value follows the existing data pattern so that the model in RBF is suitable for making predictions on testing data. Based on Table 4 shows that the value of ε is used as the limit of the error value. The smaller the value of ε used, the longer the learning process will be to find the right model, and the larger the value of the parameter γ used, the smaller the resulting MAPE value.

Table 5. 2024 rice harvest results in Pamekasan, Madura

Month	Prediction (Tons)	Actual (Tons)
January	11299.93	2646.46
February	17234.48	5148.68
March	15324.7	2842.69
April	11054.51	7152.78
May	17799.44	21776.89
June	17477.81	2306.46
July	16434.32	6413.58
August	18411.79	10423.38
September	14060.42	6946.56
October	16369.72	10828.64
November	15341.04	6897.15
December	14832.79	11871.34

The results of the rice harvest prediction in Pamekasan, Madura from January-December 2024 with the best model can be seen in Table 5. It can be seen that the rice harvest prediction results in January-December 2024 increased every month, although it was constrained in the 5th month by a decrease of 3977.45 tons. This shows that the SVR model is still categorized as being able to follow the data pattern well so that it is able to predict the rice harvest significantly. The process that will be tested on the dataset, namely testing before normalization with Z-Score and also testing after data normalization with the best data split, can be seen in Figure 4. In the experiment, data that had been normalized with z-score had a small error rate in applying the SVR method for forecasting rice harvest yields.

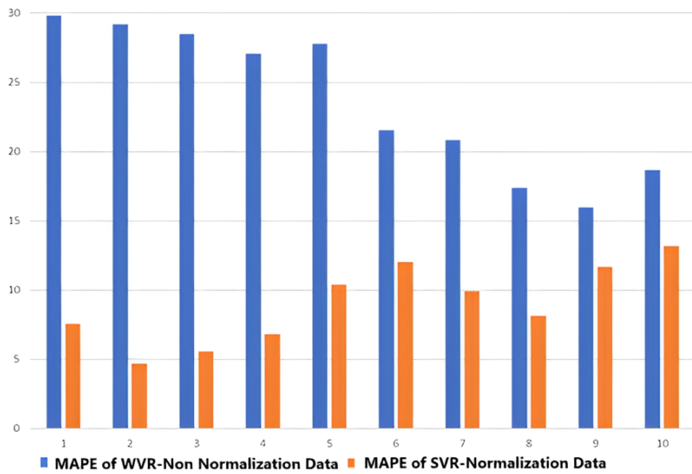


Figure 4. Test results on comparison of data normalization models

4. Conclusion

Based on the tests that have been carried out, in predicting rice harvest results with SVR, the best results were obtained on training data of 90% of the total data, γ of 4.784, $C = 0.0826$ and $\varepsilon = 0.083$ and the kernel parameters are the RBF kernel which has produced a MAPE value of 3.97%. Based on this, it can be concluded that the parameters in the SVR can affect the model data with a large γ value can provide the best model impact, a small C makes it more focused on the tolerance of errors that occur in data training, and ε can improve model performance and the RBF kernel can also handle problems with complex data. The forecasting results using the SVR model with the best combination of parameters by performing data normalization have produced the smallest MAPE. This proves that the SVR model can predict rice harvests much better than forecasting without using the data normalization model. In addition, forecasting rice harvest results in Pamekasan Regency in 2024 using the SVR method has achieved an average prediction of 15,470.08 tons with an average actual data of 7,937,884 tons. This shows that SVR has the ability to model that can follow data patterns well even though in the 5th month there has been a decline in rice production. However, in other months there has been an increase from the previous data so that SVR can be recommended for predicting rice harvest results in the future.

Author Contributions. **Devie Rosa Anamisa**: Conceptualization, methodology, analysis, writing and drafting the original draft. **Bain Khusnul Khotimah, Mohammad Yanuar Hariyawan, and Dina Violina**: Conceptualization, methodology, analysis, writing—review and editing, supervision. **Firli Irhamni, Achmad Jauhari, Fifin Ayu Mufarroha, and Dinah Nuraini**: Conceptualization, methodology, writing—review and editing, supervision. All authors have read and agreed to the published version of the manuscript.

Acknowledgement. We want to thank the Department of Agriculture and Food Security in Pamekasan, Madura, for sharing primary data to be processed in this research. We also acknowledge the organizers of the Brawijaya International Conference on Pure and Applied Mathematics (BICoPAM) 2024 for providing a platform to present and discuss this research. Additionally, we express our heartfelt thanks to the editors and reviewers for their meticulous review, insightful feedback, and constructive suggestions, which have greatly enhanced the quality of this paper.

Funding. We thank the campus, especially the research and community service institutions at Trunojoyo University, Madura, for all their support for this research.

Conflict of interest. The authors declare no conflict of interest related to this article.

Data availability. Not applicable.

References

[1] M. P. Erlin, A. T. Putra, and M. A. Muslim, "Forecasting Inflation Rate Using Support Vector Regression (SVR) Based Weight Attribute Particle Swarm Optimization (WAPSO)," *Sci. J. Informatics*, vol. 5, no. 2, pp. 118–127, 2018.

[2] L. Lindawati, I. Zulfida, S. K. H. Nasution, and S. Handayani, "Multi-aspect Analysis of Rice Sustainability in the Improvement of Rice Production in North Sumatra Province, Indonesia," *Agro Bali Agric. J.*, vol. 7, no. 2, pp. 390–398, 2024, doi: 10.37637/ab.v7i2.1741.

[3] S. Dewi and R. P. Destiarni, "Technical Efficiency of Ciherang Rice Farming in East Lenteng Village, Lenteng District, Sumenep Regency," *Agrisocionomics*, vol. 7, no. 2, pp. 349–362, 2023, [Online]. Available: http://ejournal2.undip.ac.id/index.php/agrisocionomics.

[4] P. Fendiyanto, "Peramalan Hasil Produksi Padi di Desa Mulawarman dengan Metode Trend Moment," *AKSIOMA (J. Mat. dan Pendidik. Mat.)*, vol. 15, no. 1, pp. 105–113, 2024.

[5] G. A. Prasetyo, S. A. Wibowo, and N. Vendyansyah, "Penerapan Data Mining dengan Metode Double Moving Average untuk Memprediksi Penjualan Kedelai," *JATI (J. Mhs. Tek. Inform.)*, vol. 7, no. 4, pp. 2303–2310, 2023, doi: 10.36040/jati.v7i4.7490.

[6] D. A. Mardhika, B. D. Setiawan, and R. C. Wihandika, "Penerapan Algoritma Support Vector Regression pada Peramalan Hasil Panen Padi Studi Kasus Kabupaten Malang," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 10, pp. 9402–9412, 2019, [Online]. Available: http://j-ptiik.ub.ac.id.

[7] A. A. Suyono, K. Kusrini, and M. R. Arief, "Prediksi Indeks Harga Konsumen Komoditas Makanan di Kota Surabaya menggunakan Support Vector Regression," *Metik J.*, vol. 6, no. 1, pp. 45–51, 2022, doi: 10.47002/metik.v6i1.339.

[8] M. F. Rohmah, I. K. G. D. Putra, R. S. Hartati, and L. Ardiantoro, "Comparison of Four Kernels of SVR to Predict Consumer Price Index," *J. Phys. Conf. Ser.*, vol. 1737, no. 1, 2021, doi: 10.1088/1742-6596/1737/1/012018.

[9] H. Khan, T. J. Esau, A. A. Farooque, and F. Abbas, "Wild Blueberry Harvesting Losses Predicted with Selective Machine Learning Algorithms," *Agric.*, vol. 12, no. 10, pp. 1–16, 2022, doi: 10.3390/agriculture12101657.

[10] R. J. D. Simamora, Tibyani, and Sutrisno, "Peramalan Curah Hujan Menggunakan Metode Extreme Learning Machine," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 10, pp. 9670–9676, 2019, [Online]. Available: http://j-ptiik.ub.ac.id.

[11] I. Permana, "The Effect of Data Normalization on the Performance of the Classification Results of the Backpropagation Algorithm," *Indones. J. Inform. Res. Softw. Eng.*, vol. 2, no. 1, pp. 67–72, 2022, [Online]. Available: https://media.neliti.com/media/publications/485639.pdf.

[12] T. N. Tran, B. M. Lam, A. T. Nguyen, and Q. B. Le, "Load forecasting with support vector regression: influence of data normalization on grid search algorithm," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 4, pp. 3410–3420, 2022, doi: 10.11591/ijece.v12i4.pp3410-3420.

[13] M. L. Radhitya and G. I. Sudipa, "Pendekatan Z-Score dan Fuzzy dalam Pengujian Akurasi Peramalan Curah Hujan," *SINTECH (Sci. Inf. Technol. J.)*, vol. 3, no. 2, pp. 149–156, 2020, doi: 10.31598/sintechjournal.v3i2.567.

[14] Y. H. Chen, W. C. Hong, W. Shen, and N. N. Huang, "Electric load forecasting based on a least squares support vector machine with fuzzy time series and global harmony search algorithm," *Energies*, vol. 9, no. 2, 2016, doi: 10.3390/en9020070.

[15] R. Dini, B. D. Setiawan, and C. Dewi, "Prediksi Hasil Panen Benih Tanaman Kenaf Menggunakan Metode Support Vector Regression (SVR) pada Balai Penelitian Tanaman Pemanis dan Serat (Balittas)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 12, pp. 6519–6526, 2018, [Online]. Available: http://j-ptiik.ub.ac.id.

[16] H. H. Mahmood and Z. Hamodat, "Intelligent Systems and Applications in Enhancing Maximum Power Point Tracking through Ensemble Techniques," vol. 12, no. Ic, pp. 1072–1084, 2024.

[17] A. M. Priyatno, A. Wiratmo, F. Syuhada, and P. Cholidhazia, "Perbandingan Imputasi dan Parameter Support Vector Regression untuk Peramalan Cuaca," *J. SIMETRIS*, vol. 10, no. 2, pp. 651–660, 2019, doi: 10.24176/simet.v10i2.3402.

[18] S. Saadah, F. Z. Z, and H. H. Z, "Support Vector Regression (SVR) dalam Memprediksi Harga Minyak Kelapa Sawit di Indonesia dan Nilai Tukar Mata Uang EUR/USD," *J. Comput. Sci. Informatics Eng.*, vol. 5, no. 1, pp. 85–92, 2021, doi: 10.29303/jcosine.v5i1.403.

[19] Z. Rais, "Analisis Support Vector Regression (SVR) dengan Kernel Radial Basis Function (RBF) untuk Memprediksi Laju Inflasi di Indonesia," *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 4, no. 1, pp. 30–38, 2022, doi: 10.35580/variansiunm13.

[20] M. R. Rifqi, B. Darma Setiawan, and F. A. Bacthiar, "Support Vector Regression Untuk Peramalan Permintaan Darah: Studi Kasus Unit Transfusi Darah Cabang-PMI Kota Malang," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 10, pp. 3332–3342, 2018.

[21] S. Amar, A. Sudiarso, and M. K. Herliansyah, "The Accuracy Measurement of Stock Price Numerical Prediction," *J. Phys. Conf. Ser.*, vol. 1569, no. 3, 2020, doi: 10.1088/1742-6596/1569/3/032027.