Vol. 7, No. 3, June 2026, Pages 2876-2890
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5509

Comparison of SVR Parameter Optimization Using Particle Swarm Optimization (PSO) and Random Search for Rice Harvest Yield Prediction

Narlin Yumeivia*[1], Farid Wajidi[2], Wawan Firgiawan[3]

[1,2]Informatics, Faculty Engineering Department, University from the West West Sulawesi, Indonesia
[3]Affiliates Middle For False Intelligence Studies, Faculty from Engineering, University from West Sulawesi, Indonesia

Email: [1]narlinyumeivia@gmail.com, [2]faridwajidi@unsulbar.ac.id, [3]wawanfirgiawan@unsulbar.ac.id

Received: May 12, 2026; Revised: May 25, 2026; Accepted: May 25, 2026; Published: Jun 15, 2026

Abstract

Rice yield is an important part in a precision agriculture system that can support farmers' decision-making in a more targeted manner. The author's research aims to help farmers and stakeholders in Bambang Village predict crop yields accurately to overcome production fluctuations. Through appropriate efforts and strategies, this technology is expected to improve food security and farmer welfare. The research method uses the *Support Vector Regression* (SVR) algorithm for the modeling process, with the help of *Particle Swarm Optimization* (PSO) and *Random Search optimization* in finding the best parameters. The research dataset includes 1,120 historical data of rice harvests in Bambang Village for the 2022–2023 period tested through 70:30 and 60:40 data sharing scenarios. Model performance is evaluated using the Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and coefficient of determination (R^2) metrics. The MAPE metric is used as the main indicator of relative accuracy by measuring the average percentage deviation between predicted values and actual values; a low MAPE value is very significant because it reflects the model has a minimal error rate on a percentage scale, thus providing more precise estimates for farmers. The results showed that both optimization methods successfully identified SVR parameters (C, gamma, epsilon) that followed the data trend. Random Search produced slightly superior R^2 performance (reaching 82.20% at a 60:40 ratio), while PSO showed more consistent parameter exploration stability. These findings demonstrate that the integration of machine learning and optimization techniques has great potential in strengthening data-driven agricultural systems to improve food security and farmer welfare.

***Keywords:** Padi, Particle flock o Roptimization, Agriculture, Support Vector Regression, Produce.*

1. INTRODUCTION

Indonesia is famous for its rich natural resources [1] . Indonesia is a country rich in natural resource potential, so that sectors, especially agriculture such as food crops, plantations, forestry, fisheries and animal husbandry, are widely developed to increase agricultural production and farmer incomes. The abundant and sustainable agricultural potential in Indonesia is one of the efforts to maintain geostrategy, where a country that has strong food security and is able to meet national food needs and reserves independently will have strong national resilience. [2] .

The agricultural sector has played a vital role in supporting communities in areas with vast tracts of land and fertile soil. Agriculture is crucial for improving a region and the regional economy itself. [3] . Indonesia, which is known as an agricultural country [4] , the majority of the population lives from the agricultural sector, one type of food produced is rice which comes from rice plants which is the main

P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 3, June 2026, Pages 2876-2890
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5509

source of calories and plays a very important role in the economic life of the community, because it is a daily staple food. [5] .

Rice (Oryza sativa) is one of the most important crops in Indonesia. Rice is a plant that produces rice and is a staple food for the Indonesian people. Indonesia is ranked third as the largest rice producing country in the world, after China and India. [6] . Rice is a food commodity that plays an important role in the economy, because it is a material that provides basic needs, as well as a source of livelihood [7] .

In Mamasa Regency, West Sulawesi Province, Bambang Village is known as an agricultural village where most of the population works as farmers. Rice is the main food crop in the village because besides being a staple carbohydrate, rice is also the main livelihood for most of the population. However, rice yields in Bambang Village often experience *unstable fluctuations*, with uncertain decreases or increases each season. Uncertainty in harvest results can impact the community's economy, especially rice farmers, such as shortages of food products that are the staple food of farmers. Uncertainty in harvest results also causes unstable food prices that can be detrimental to consumers and farmers. However, rice yields in Bambang Village often experience unstable fluctuations from one planting season to the next. These fluctuations create uncertainty for farmers in estimating production levels and planning agricultural activities. The lack of an accurate prediction system for rice production makes it difficult for farmers and local stakeholders to anticipate potential declines in harvest yields and manage food supplies effectively. Therefore, the development of accurate prediction models is important to support decision-making in agricultural planning.

Based on the description above, the author will conduct research aimed at helping farmers and related stakeholders predict crop yields, better prepare for yield fluctuations, and plan appropriate mitigation measures. With a reliable prediction system, farmers in Bambang Village can better estimate their harvest yields. Through accurate predictions, appropriate adaptation strategies can be developed to improve food security and farmer welfare in Bambang Village.

Prediction is the process of systematically describing something that is likely to happen in the future based on available past and present information, so that the difference between what happened and the predicted results can be minimized as much as possible [8] . To predict rice yields, one method that can be used is the Support Vector Regression (SVR) method and parameter optimization using Particle Swarm Optimization (PSO) and random search methods. Support Vector Regression (SVR) is a regression method based on Support Vector Machines that is used to model the relationship between input and output variables continuously in the prediction process [9] . This method works by finding the optimal regression function that can minimize prediction errors through the concept of maximum margin and kernel functions to handle non-linear relationships in the data. SVR is widely used in various fields such as commodity price prediction, house prices, agricultural production, and time series data because of its ability to produce predictive models with a high level of accuracy [10] . In 1995, Kennedy and Eberhart developed PSO based on a stochastic optimization technique known as social simulation mode. PSO was developed to solve optimization problems using swarm behavior. [11] . *Particle Swarm Optimization (PSO)* is an optimization method that has been proven effective in solving non-linear optimization problems [12] . *Particle Swarm Optimization* (PSO) is used for *hyperparameter tuning* because of its ability to efficiently explore the hyperparameter space in search of optimal values. PSO utilizes a set of particles, each representing a potential solution in the hyperparameter space. The particles iteratively adjust their positions based on their own experience and the collective knowledge of the swarm, allowing for a dynamic and collaborative search process. This approach is particularly effective for hyperparameter tuning because it can navigate complex multidimensional spaces to find configurations that improve the performance of machine learning models. [13] . Hyperparameter tuning is a crucial step in the training process of a machine learning model, aiming to find the best parameter combination for optimal model performance [14] . Random Search is a random search-based

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5509

hyperparameter optimization method developed to improve the efficiency of solution search in a large parameter space [15] . Experts propagate all combinations as in Grid Search, Random Search, randomly selecting hyperparameter values from a predefined space. In the tuning process, Random Search relies on random sampling from hyperparameter distributions such as uniform or log-uniform, which facilitates the exploration of high-dimensional parameter spaces [16] . main The advantage of Random Search is its clumsiness and ease application to many kinds of types of machine learning algorithms, both simple and complex [17] . Random Search is a hyperparameter optimization technique that randomly samples parameter combinations from a predefined search space to identify configurations that improve model performance. This approach allows for efficient exploration of large parameter spaces without exhaustively evaluating all possible combinations [18] . Random Search is commonly used in machine learning to find optimal hyperparameters by randomly exploring parameter combinations within a defined range to improve predictive performance. Random Search can provide competitive accuracy while significantly reducing computational time compared to exhaustive search methods such as Grid Search. [19] .

Research conducted by Sephia Pratista et al. [20] , "Comparison of Drug Use Prediction Techniques Using Simple Linear Regression Algorithms and Support Vector Regression." From the results of research conducted by comparing the Simple Linear Regression and Support Vector Regression algorithm methods, based on experiments that have been conducted on the drug Paracetamol 500 mg using the Simple Linear Regression algorithm, the MAPE results were 20.85% in the "Quite Good" category and using the Support Vector Regression algorithm produced a MAPE of 18.39% in the "Good" category. Experiments on the drug Cetirizine using the Simple Linear Regression algorithm produced a MAPE result of 18.39% in the "Good" category and using the *Support Vector Regression algorithm* produced a MAPE of 17.14% in the "Good" category. Based on the MAPE results obtained, the *Support Vector Regression algorithm* has better prediction results than the Simple Linear Regression algorithm. Furthermore, research by Rizky [15] , showed that the use of Random Search for hyperparameter tuning in tree-based algorithms, such as Decision Trees, Random Forests, and Deep Forests, can improve the accuracy in predicting software defects. As a result, the optimized Random Forest and Deep Forest models achieved an AUC of 0.79, higher than that of Decision Trees. This proves that *Random Search* is an effective tuning method to significantly improve the performance of predictive models. Furthermore, the results of research conducted by [21] This study is entitled "Application of *Support Vector Regression* (SVR) and *Particle Swarm Optimization* (PSO) Methods for Rice Production Forecasting". This study uses data from January 2013 to December 2023 obtained from the Food Crops and Horticulture Service of Lima Puluh Kota Regency. The method used in this study is *Support Vector Regression* (SVR) which is optimized using *Particle Swarm Optimization* (PSO). The data used in this study is rice production data in the form of a time series. The optimal parameters used are the C parameter limit of 0.001-100, the γ parameter limit of 0.0001-0.5, the parameter limit of 0.001-0.5, 250 iterations, 150 particles, C1 and C2 values of 2 and 2 respectively, W value of 0.7, and using non-differentiated data. The test results show a MAPE value of 4.699388% which is included in the very good category.

Based on research related to the application of machine learning algorithms such as *Support Vector Regression* (SVR). SVR is a very relevant and effective method for solving rice yield prediction problems due to its ability to handle non-linear and multivariate data. However, SVR performance is greatly influenced by the choice of core parameters such as C value, ε, and kernel parameters. Therefore, the application of optimization methods such as *Particle Swarm Optimization* (PSO) is very important to identify optimal parameters that improve the accuracy and stability of the prediction model, as shown in previous research [21] . Therefore, this study aims to compare the effectiveness of Particle Swarm

Vol. 7, No. 3, June 2026, Pages 2876-2890
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5509

Optimization (PSO) and Random Search in optimizing Support Vector Regression (SVR) parameters for rice yield prediction using historical agricultural data from Bambang Village.

2. METHOD

This study uses the *Support Vector Regression* (SVR) algorithm for the modeling process, with the help of Particle Swarm Optimization (PSO) and Random Search optimization in finding the best parameters. Both optimizations are used to find the combination of SVR parameters that produce the most optimal model performance. This study was conducted using the Python programming language. Several libraries were used, including Scikit-Learn for machine learning implementation, Pandas and NumPy for data manipulation, and Matplotlib and Seaborn for data visualization. The model development process and experiments were carried out using the Jupyter Notebook and Google Colab environments to ensure computational reproducibility. The steps of the research method to be carried out are shown in the figure below:

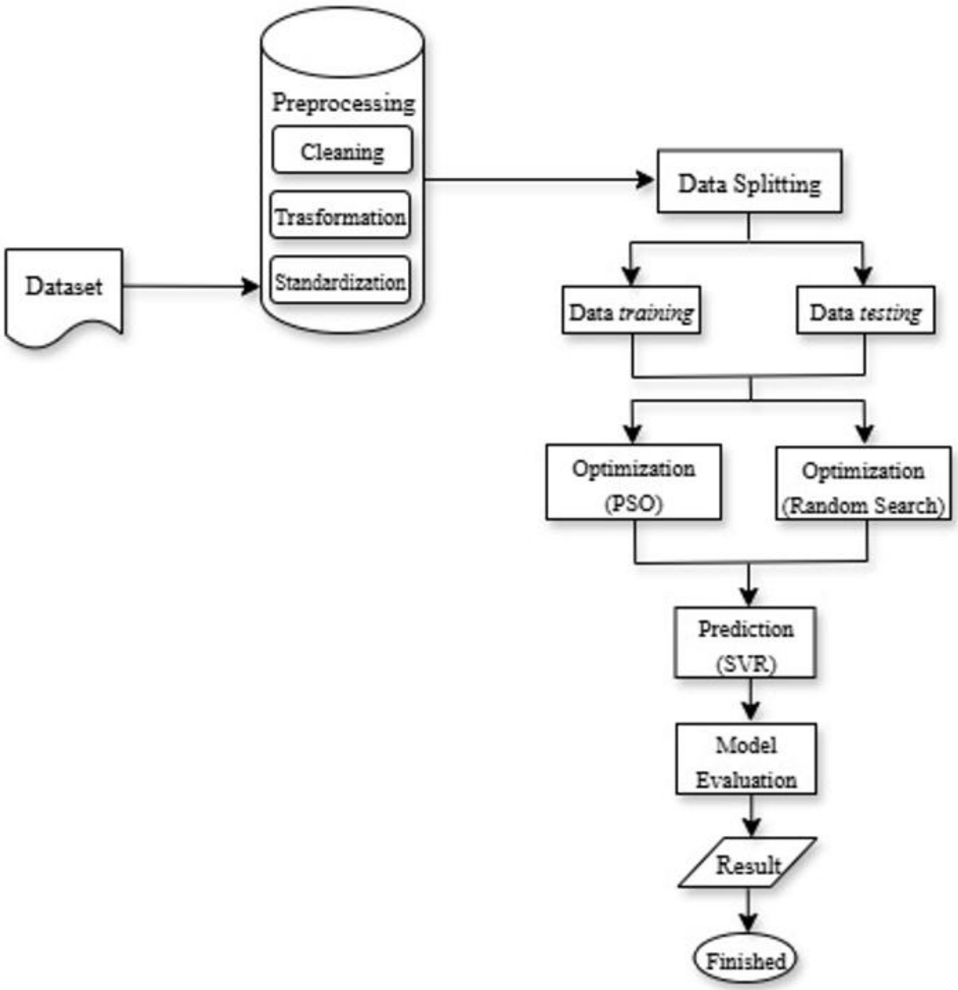

Figure 1. System Flowchart

2.1. Data set

The data used in this study comes from the 2022-2023 rice harvest data of Bambang Village, totaling 1120 data consisting of several variables used as a reference to determine the rice harvest yield as can be seen in Table 1.

Vol. 7, No. 3, June 2026, Pages 2876-2890
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5509

Table 1. Data set Example

land area (Ha)	Number of Seeds (L)	types of fertilizer	It's raining all over	pests and weeds	rice varieties	pest and weed control	planting techniques	Harvest yield d (kg)
0.5	30	Urea and NPK	Normal	0.013	Ciherang	pesticides and weeding	The Legowo Line	1880
0.9	50	Urea and NPK	Normal	0.179	Cibogo	pesticide	The Legowo Line	3465
0.76	45	Urea and NPK	Normal	0.036	Ciherang	pesticides and weeding	Traditional System	2967
1	60	Urea and NPK	Tall	0.07	Ciherang	pesticides and weeding	Traditional System	3734
0.82	49	Urea and NPK	Tall	0.256	Ciherang	pesticides and weeding	The Legowo Line	2982

2.2. Data Preprocessing

The pre-processing stage has a very important role in supporting the improvement of the performance of machine learning algorithms. [22] . After going through the pre-processing stage, the data is expected to have an optimal structure and representation, so that it is ready to be further processed by the machine learning algorithm. [23] . This process begins with Exploratory Data Analysis (EDA) to understand the initial characteristics of the dataset, detect patterns, and identify inconsistent data. Next, the data cleaning stage is carried out to address missing data, duplicates, and extreme values that can interfere with the analysis results. The process continues with data transformation to change the format and structure of the data to suit the needs of the analysis method used [24] .

In this study, before the data obtained is processed using the *Support Vector Regression algorithm*, preprocessing is carried out first. This process aims to ensure that the available data is in an appropriate format so that it can be processed by the system. Each variable is reviewed based on its data type, considering that the algorithm used can only handle data in numeric form [25] . There are several steps carried out in data preprocessing, which include data cleaning, data standardization, data transformation.

2.2.1. Exploratory Data Analysis (EDA) Process

The Exploratory Data Analysis (EDA) process in the data preprocessing stage is the first step in data analysis by checking for empty data or missing values. If there is empty data in the dataset, data cleaning or deletion of empty data can be performed [26] . This process aims to understand the characteristics of the data before further modeling or analysis is carried out. The initial data used in the Exploratory Data Analysis (EDA) process is secondary data sourced from the Bambang Village Office, covering the time span of 2022 to 2023. Before the EDA process, the dataset consisted of 1120 rows. After going through the data cleaning stage, the number of usable data was reduced to 1119 rows.

Vol. 7, No. 3, June 2026, Pages 2876-2890
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5509

2.2.2. Data Cleansing

The data cleaning process is carried out by deleting irrelevant or unnecessary information, such as farmer name columns, number columns, and additional information in the form of units such as (Ha), (Tree), (Kg), and percentage symbols (%).

2.2.3. Data Transformation

The data transformation stage includes converting percentage values to decimal form and applying label coding to convert string data into numeric data. The attributes subjected to the label coding process include fertilizer type, rainfall, pest and weed control, rice variety, and planting technique, where each attribute value is coded into the numbers 0 and 1. Meanwhile, pest and disease attributes, which were originally in percentage form, were converted to decimal values to facilitate further analysis.

2.3. Particle Swarm Optimization (PSO)

In 1995, Kennedy and Eberhart developed PSO based on a stochastic optimization technique known as social simulation mode. PSO was developed to solve optimization problems using swarm behavior. [11] . *Particle Swarm Optimization (PSO)* is an optimization method that has been proven effective in solving non-linear optimization problems [12] .

Particle Swarm Optimization (PSO) is used for *hyperparameter tuning* due to its ability to efficiently explore hyperparameter space in search of optimal values. PSO utilizes a set of particles, each representing a potential solution in the hyperparameter space. The particles iteratively adjust their positions based on their own experience and the collective knowledge of the swarm, enabling a dynamic and collaborative search process. This approach is particularly effective for hyperparameter tuning because it can navigate complex multidimensional spaces to find configurations that improve the performance of machine learning models. [13] **.**

2.4. Hyperparameter Tuning

The hyperparameter tuning process is carried out to find the best parameters for the SVR model in predicting rice yields. In this study, the Random Search method is used, a tuning approach that randomly selects hyperparameter combinations from a predetermined distribution of values. Unlike Grid Search, which exhaustively tests all possible combinations, Random Search offers time and resource efficiency while remaining close to optimal performance. By randomly testing a number of combinations, this method enables faster and more effective modeling, and helps obtain the best SVR parameters, which can improve the overall accuracy of crop yield predictions.

2.4.1. Random Search

Hyperparameter tuning is a crucial step in the training process of a machine learning model, aiming to find the best parameter combination for optimal model performance [14] . Random Search is a random search-based hyperparameter optimization method developed to improve the efficiency of solution search in a large parameter space [15] . In contrast to Grid Search, which exhaustively evaluates all hyperparameter combinations, Random Search randomly samples hyperparameter values from a predefined search space. This approach relies on sampling from a probability distribution, such as uniform or log-uniform, which allows for more efficient exploration of high-dimensional hyperparameter spaces. [16] .

That main The advantage of Random Search is its clumsiness and ease application to many kinds of types of machine learning algorithms, both simple and complex [17] .

Vol. 7, No. 3, June 2026, Pages 2876-2890
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5509

2.5. Support Vector Regression (SVR)

SVR is an extension of SVM for regression purposes. The goal of SVR is to find a hyperplane (separating line) in the form of a regression function that fits all input data containing errors and minimizes it as much as possible [27] . Support Vector Regression (SVR) is one implementation of Support Vector Machine (SVM) where the approach uses a regression method, thus producing predicted values in the form of real or continuous numbers [27] .

In this study, the RBF (Radial Basis Function) kernel is used. The Radial Basis Function kernel, also known as the Gaussian kernel, is used in SVR to address the non-linear relationship between features and the target variable. The RBF kernel function is defined as: [28]

$$K(\mathbf{x}, \mathbf{x}_i) = \exp(-\gamma \|\mathbf{x} - \mathbf{x}_i\|^2) \qquad (1)$$

Where $\mathbf{x}$, and $\mathbf{x}_i$am input data or vector, ||.|| describes the Euclidean distance between x and xi, while $\boldsymbol{\gamma}$ is a parameter that controls the width of the Gaussian curve, thus determining the complexity of the model and its ability to capture local patterns in nonlinear data.

nonlinear data. The characteristics of the RBF kernel that consider the Euclidean distance between data allows SVR to capture local variations that ordinary linear approaches cannot accommodate, making it particularly suitable for regression on data with irregular patterns [27] .

2.6. Data Division

After data preprocessing, the next step is data partitioning. The dataset is divided into two parts: training data and testing data. The model is trained using the dimensionally reduced training data and then evaluated on the testing data to measure the model's generalization ability. This approach provides a more objective assessment of the model's predictive performance. The data is divided into two ratios, as shown in the following table:

Table 2. Composition of training - testing

NO	Scheme	Training test composition
1	I	60:40
2	II	70:30

2.7. Evaluation

This stage is carried out to ensure that the algorithm used is running as planned. Several methods are used in testing. These include the Root Mean Squared Error (RMSE) and the Mean Absolute Percentage Error (MAPE), which are used to assess the algorithm's error rate. The final stage is calculating the accuracy of the MAPE error value to determine the algorithm's accuracy.

The Root Mean Squared Error (RMSE) method is a widely used method for evaluating forecasting performance. By applying RMSE, the resulting error indicates the extent of the difference between the estimated and predicted results. The formula for calculating the RMSE value is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2} \qquad (2)$$

Where:
RMSE : *Mean Square Error*
n : Data loss
Y_t : Actual value
Y_{t+1} : Predicted value

Vol. 7, No. 3, June 2026, Pages 2876-2890
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5509

Mean Absolute Percentage Error (MAPE) is a measure of relative error. MAPE is typically a relative measure, calculating the percentage error in forecast results. MAPE is the average of the absolute errors. MAPE is used to test accuracy due to its ability to provide precise and accurate results across a wide range of variables. The MAPE value can be calculated using the following formula:

$$MAPE = \frac{100\%}{N} \sum_{t=1}^{N} \left| \frac{X_t - F_t}{X_t} \right| \quad (3)$$

Where:
MAPE : *Mean Absolute Percentage Error*
N : Number of data
X_t : Actual value
F_t : Prediction value

The lower the MAPE value, the better the model, as the resulting accuracy level will be high. The accuracy level is obtained based on the equation:

$$Accuracy = 100\% - MAPE \quad (4)$$

3. RESULTS

This study aims to compare the effectiveness of two parameter optimization methods, namely Particle Swarm Optimization (PSO) and Random Search, in optimizing the Support Vector Regression (SVR) algorithm to obtain the best parameter configuration. The use of the PSO optimization method and the Random Search tuning method in this study is intended to ensure the model can perform optimally and produce more accurate predictions, by minimizing prediction errors during the training process.

The dataset used is rice harvest data that has been collected and preprocessed, including data cleaning, normalization, and target feature separation. The data is then divided into two parts: training data and testing data, to ensure that the model performance evaluation process is carried out systematically, objectively, and without bias towards any particular data. This approach is expected to provide a comprehensive understanding of the advantages of each algorithm in predicting crop yields based on available historical data patterns.

3.1. SVR Parameter Optimization Using PSO

In the optimization stage, the Particle Swarm Optimization (PSO) algorithm is used to determine the optimal values for the C, gamma, and epsilon parameters in the Support Vector Regression (SVR) model. The goal of this process is to minimize the Mean Squared Error (MSE) so that the model can produce more accurate predictions. Evaluation during the optimization process is carried out using a 5-fold cross-validation technique to ensure that the obtained parameters are stable and do not experience overfitting.

The optimal parameter values obtained through PSO can vary depending on the training and testing data split ratio used in the testing scenario. This is due to the different distribution structures of the training data used in the learning and validation processes, which affect the PSO algorithm's search for the best solution. Therefore, in each 70:30 and 60:40 scenario, the parameter optimization process is performed separately. The following parameter optimization results with the 70:30 scenario are shown in Table 2-5 below:

The results of SVR parameter optimization using PSO with a data split of 70:30 are presented in Table 3.

Vol. 7, No. 3, June 2026, Pages 2876-2890
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5509

Table 3. Optimization Results with a 70:30 Ratio

Parameter	Optimal Value
C	2209.5985
gamma	0.0076501
epsilon	0.9192301
MSE CV	277300.9698

Table 3 shows that based on the results of the *Support Vector Regression* (SVR) model parameter optimization using the *Particle Swarm Optimization* (PSO) algorithm with a ratio of 70:30, the best parameter configuration was obtained, namely C of 2209.5985, gamma of 0.0076501, and epsilon of 0.9192301. This parameter combination resulted in a cross-validation *Mean Squared Error* (MSE) value of 277,300.97, which reflects the average squared error of the model's predictions on the validation data. A high C value indicates that the model imposes a large penalty on prediction errors, thus tending to produce more accurate predictions on the training data. However, this value must also be balanced to avoid overfitting. On the other hand, a low gamma value indicates that the model captures patterns globally and is less sensitive to local variations or noise, which helps maintain model stability. Meanwhile, a relatively large epsilon value provides a wider tolerance for small errors, so the model is less rigid to small deviations and focuses more on general trends in the data. Overall, this configuration demonstrates a balance between model accuracy and generalizability. The use of PSO has proven effective in finding optimal SVR parameters that can improve predictive performance, particularly in the context of data-driven crop yield estimation.

The results of the SVR model evaluation using a 70:30 data split are presented in Table 4**.**

Table 4. Results of the 70:30 Ratio Evaluation

Matrix Evaluation	Yield Value
MAP	7391.28%
RMSE	581.4686
R^2	79.66% (Accuracy)

Table 4 shows the model evaluation results, which indicate a Mean Absolute Percentage Error (MAPE) of 7391.28%, a Root Mean Square Error (RMSE) of 581.4686, and an R-squared (R^2) of 79.66%. A very high MAPE value indicates that the model's prediction error in percentage terms relative to the actual value is relatively large. This indicates that the model is less accurate in estimating crop yields from a relative perspective. Meanwhile, the RMSE of 581.4686 indicates that the average deviation of predictions from actual values is quite large in kilograms, indicating a significant difference between actual and predicted values. The low R^2 value of 79.66% indicates that the model is only able to explain some of the variation in the target data. The closer to zero, the smaller the contribution of input features to the output variation that the model is able to explain. In this context, a low R^2 value indicates that the model is not yet able to capture a strong relationship between input and output variables.

The results of parameter optimization using the PSO algorithm with a data split of 60:40 are presented in Table 5.

Table 5. Optimization results with a ratio of 60:40:

Parameter	Optimal Value
C	4003.5916
gamma	0.0036580
epsilon	0.3277900
MSE CV	284443.5736

Vol. 7, No. 3, June 2026, Pages 2876-2890
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5509

Table 5 shows the test results with a 60:40 data ratio that produced the best parameters for the SVR model, namely C of 4003.5916, gamma of 0.0036580, and epsilon of 0.3277900. The Mean Squared Error (MSE) value from the cross-validation obtained was 284,443.57. A large C value indicates that the model imposes a high penalty on prediction errors, thus attempting to minimize errors. A low gamma value allows the model to recognize common patterns in the data and reduces the risk of overfitting. Meanwhile, a moderate epsilon provides room for small error tolerance, allowing the model to be more flexible in the prediction process following data variations. This parameter combination indicates that the model can still adapt and perform quite well even though the portion of the training data is smaller.

The results of the SVR model evaluation using a 60:40 data split are presented in Table 6.

Table 6. Results of the 60:40 Ratio Evaluation

Matrix Evaluation	Result value
MAPE	8973.03%
RMSE	556.8594
R^2	81.88% (accuracy)

Table 6 shows the results of the model evaluation with a data ratio of 60:40, obtaining a MAPE value of 8973.03%, an RMSE of 556.8594, and an R^2 of 81.88%. Although the MAPE and RMSE values are relatively high, the R^2 value of 81.88% indicates that the model has a fairly good ability to explain data variation.

When compared, the test results with the 70:30 and 60:40 ratios show differences in both optimal parameters and evaluation performance. At the 70:30 ratio, the model obtained a lower cross-validation MSE value than the 60:40 ratio, although the MAPE and R^2 values both indicated poor predictive performance. Meanwhile, at the 60:40 ratio, the model used less training data, which can limit the space for learning patterns. This results in a decrease in prediction quality, reflected in an increase in the MAPE value and a better R^2 value.

3.1.1. Visualization of Prediction Results

Graphic visualizations were used to demonstrate the degree of agreement between model predictions and actual data. This graph compares actual rice yields with those predicted by the SVR model, optimized using Particle Swarm Optimization (PSO).

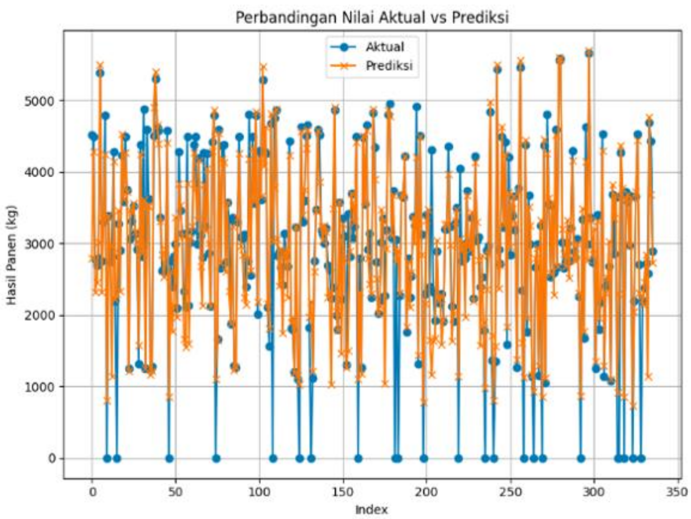

Figure 2. Actual vs Predicted Values

Vol. 7, No. 3, June 2026, Pages 2876-2890
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5509

The graph in Figure 1 shows a visualization of the comparison between the actual yield values. And That predicted mark produced by That model. From that chart, he Can become observed that in general, the predicted line (orange) follows the pattern of the actual values (blue), although at some points there are significant differences. This indicates that the model is able to capture the general pattern. However, there is still a number of deviation that can become caused by noise or variables That own No has been included in the model.

3.2. SVR Using Random Search

In the optimization stage, the Random Search method is used to determine the most optimal values for the C, gamma, and epsilon parameters in the Support Vector Regression (SVR) model. The goal of this process is to minimize the Mean Squared Error (MSE) value so that the model can produce optimal results again appropriate prediction. Evaluation during That optimization process is has brought go out use A 5-fold cross-validation technique was used to ensure the obtained parameters were stable and did not experience overfitting. The optimal parameter values obtained through Random Search may differ depending on the training and testing data split ratio used in the testing scenario. This is due to the different distribution structures of the training data used in the learning and validation processes, thus affecting the process of finding the best solution with the Random Search method. Therefore, in each data split scenario of 70:30, 80:20, and 90:10, the parameter optimization process was carried out separately. The following parameter optimization results with the 70:30 scenario are shown in the following table:

The results of parameter optimization using the Random Search algorithm with a 70:30 data split are presented in Table 7.

Table 7 Optimization Results with 70:30 Comparison

Parameter	Optimal Value
C	8342.9880
gamma	0.0001557
epsilon	0.7724725
MSE CV	299135.4744

Table 7 shows the results of the optimization of the Support Vector Regression (SVR) model parameters using the Random Search method with a data sharing ratio of 70:30. From the optimization process, the best parameter configuration was obtained, namely a C value of 8342.9880, a gamma of 0.0001557, and an epsilon of 0.7724725. This parameter combination produced a cross-validation Mean Squared Error (MSE) value of 299,135.47, which describes the average squared error of the model's predictions against the validation data.

The evaluation results of the optimized SVR model using the Random Search method with a ratio of 70:30 are shown in Table 8.

Table 8. Evaluation Results with a ratio of 70:30

Matrix Evaluation	Yield Value
MAP	8677.94%
RMSE	574.8736
R^2	80.05% (Accuracy)

Table 8 shows the results of the SVR model evaluation with optimized parameters using the Random Search method at a data ratio of 70:30. The evaluation produced a Mean Absolute Percentage

Vol. 7, No. 3, June 2026, Pages 2876-2890
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5509

Error (MAPE) value of 8677.94%, a Root Mean Square Error (RMSE) of 574.8736, and a coefficient of determination R-squared (R^2) of 80.05%.

The results of parameter optimization using the Random Search algorithm with a data split of 60:40 are presented in Table 9.

Table 9. Optimization results with a ratio of 60:40:

Parameter	Optimal Value
C	8591.501
gamma	0.0004882
epsilon	0.5499951
MSE CV	285866.1091

Table 9 shows the results of parameter optimization using the *Random Search method* with a data ratio of 60:40, which produced the best parameter combination for the Support Vector Regression (SVR) model, as shown in Table 4. The optimal parameter values obtained were C of 8591.501, gamma of 0.0004882, and epsilon of 0.5499951. The model with this parameter combination produced a cross-validation Mean Squared Error (MSE) value of 285866.1091. A high C value indicates that the model imposes a large penalty on prediction errors, thus attempting to minimize errors. A very small gamma value indicates that the model focuses more on general patterns than small details, which helps avoid overfitting. Meanwhile, a relatively large epsilon value provides tolerance for small deviations between predictions and actual values, making the model more stable in the face of data fluctuations. These results indicate that even with less training data, the model is still able to recognize patterns and provide fairly good prediction performance.

The evaluation results of the optimized SVR model using Random Search with a ratio of 60:40 are presented in Table 10**.**

Table 10. Results of the 60:40 Ratio Evaluation

Matrix Evaluation	Yield Value
MAPE	9540.69%
RMSE	551.9380
R^2	82.20% (accuracy)

Table 10 shows the results of the model evaluation with a data ratio of 60:40. By applying the Random Search technique, the MAPE value was obtained at 9540.69%, RMSE at 551.9380, and R^2 at 0.8220%.

4. DISCUSSION

Based on the research results, the Support Vector Regression (SVR) model optimized using the Particle Swarm Optimization (PSO) algorithm shows that this hybrid approach is capable of producing optimal parameter configurations through an efficient search process in the solution space. This can be observed from the best parameter values obtained for C, gamma, and epsilon at various data sharing ratios, which indicates that PSO is adaptive to the data distribution structure.

Model evaluation under two data partitioning schemes (70:30 and 60:40) showed moderate predictive performance. This is reflected in the coefficient of determination (R^2), which was around 0.80, indicating that the model can capture most of the relationship between input features and rice yield. However, the Mean Absolute Percentage Error (MAPE) remained relatively large, exceeding 70%, indicating that substantial prediction errors still occur at the individual observation level. This situation may be due to several factors, such as the limited number of predictor variables, the presence of outliers in the dataset, and potential inconsistencies in agricultural data across growing seasons. These findings

Vol. 7, No. 3, June 2026, Pages 2876-2890
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5509

are generally consistent with previous research showing that optimization algorithms can improve the performance of machine learning models in agricultural prediction tasks. For example, several studies have reported that integrating PSO with SVR can improve prediction accuracy by automatically identifying optimal hyperparameters. However, the accuracy obtained in this study was slightly lower than some previous studies, possibly due to the limited dataset size and the lack of additional explanatory variables such as weather conditions, soil properties, and irrigation factors.

The results obtained from the SVR model optimized using the Random Search method indicate that this approach is also capable of generating relatively effective parameter configurations through simple stochastic exploration of the parameter space. Interestingly, the SVR–Random Search model yields slightly better prediction performance compared to the SVR–PSO model based on the R^2 values obtained in the experiments. This suggests that Random Search can be an efficient alternative optimization method, especially when computational simplicity and implementation efficiency are desired.

However, several limitations of this study should be acknowledged. First, the dataset used in this study only includes rice yield records from a limited geographic area and time period, which may reduce the generalizability of the developed model. Second, this study only considers historical yield data without incorporating other potentially influential variables such as rainfall, temperature, soil fertility, and agricultural practices. These limitations may have contributed to the relatively high MAPE values observed in the study results.

Despite these limitations, this study provides valuable insights into the potential use of machine learning and optimization techniques in agricultural yield prediction. From a practical perspective, the proposed approach can serve as an initial framework for developing data-driven agricultural decision support systems that can assist farmers and agricultural stakeholders in more effectively estimating rice production.

Future research should consider incorporating more diverse datasets, additional environmental variables, and alternative machine learning algorithms to improve predictive accuracy and model robustness.

5. CONCLUSION

This research inspect that application from Support Vector Regression (SVR) optimized This study uses Particle Swarm Optimization (PSO) and Random Search to build a data-based rice yield prediction system. This study shows that both optimization methods are capable of producing stable and consistent model parameters. In a 60:40 data split scenario, the model optimized with PSO achieved a coefficient of determination (R^2) of 0.8188%, an RMSE of 556.86, and a MAPE of 8973.03%. Then, optimization using Random Search produced results with an R^2 of 0.8220%, an RMSE of 551.9380, and a MAPE of 9540.69%.

Both optimization approaches have their own characteristics. PSO excels in collaboration. exploration from global parameter. When Random Look for offer efficiency in the setup. This research provides a strong initial foundation for developing a machine learning-based crop yield prediction system. This research makes an initial contribution to the application of machine learning to rice produce forecast. To repair accuracy in the That future, he is recommended to add outside by leveraging features such as weather data, as well as applying more complex prediction models such as *Random Forest*, *XGBoost*, or *LSTM, and with continued development, this approach* has the potential to improve the accuracy of rice yield predictions. potential in the support technology-based food security.

For future research, it is recommended to incorporate additional relevant variables such as rainfall, temperature, soil conditions, and cultivation practices to improve prediction accuracy. Furthermore, future studies could explore other machine learning or deep learning algorithms and apply

Vol. 7, No. 3, June 2026, Pages 2876-2890
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5509

larger datasets from different regions to improve model generalization and robustness in agricultural prediction systems.

CONFLICT OF INTEREST

There is NO conflict from interest between That author or That object from research in the This paper.

THANK-YOU NOTE

We would like to express our deepest gratitude to Mr. Farid Wajidi, S. Kom., MT, and Mr. Wawan Firgiawan, ST, MT, as our first and second supervisors, for their valuable guidance, direction, and input throughout the research and writing process of this article. We would also like to express our gratitude to the Faculty of Engineering and the University of West Sulawesi for their support and contributions to the smooth and successful implementation of this research. Finally, we would like to express our appreciation to all parties who have assisted directly or indirectly in the completion of this research. We hope that the results of this research will be beneficial. research Can provide benefit and positive contribution to That development from science in related fields.

REFERENCE

[1] F. Esfandiari *et al.* , "Assistance of MUI Halal Contract and Certification and Education on Halal Product Assurance for Cangloh Beverages in Mergosono, Malang City," *J. Dedik. huh.* , vol. 1, no. 2, pp. 87–99, 2021, doi: 10.22219/jdh.v1i2.17607.

[2] EIR Rhofita, "Optimizing Indonesian Agricultural Resources to Support the National Food and Energy Security Program," *J. National Resilience* , vol. 28, no. 1, p. 82, 2022, doi: 10.22146/jkn.71642.

[3] ICR Drajana and A. Bode, "Support Vector Machine for Food Crop Production Prediction in Gorontalo Province," *Nusant. English* , vol. 4, no. 2, pp. 104–111, 2021, doi: 10.29407/noe.v4i2.16757.

[4] Q. Ayun, S. Kurniawan, and WA Saputro, "The Development of Agricultural Land Conversion in Agrarian Countries," *Vigor J. Agricultural Science. Trop. And Subtrop.* , vol. 5, no. 2, pp. 38–44, 2020, doi: 10.31002/vigor.v5i2.3040.

[5] S. Nurul, F. Harahap, J. Nasution, and K. Tambunan, "Rice supplies, rice prices, rice consumption against inflation in Medan city," vol. 13, no. 02, pp. 625–638, 2024.

[6] A. Basit, "IMPLEMENTATION OF NAIVE BAYES ALGORITHM TO PREDICTE RICE HARVEST RESULTS," *J. Tech. Inform. Kaputama* , vol. 4, no. 2, pp. 208–213, 2020.

[7] AAP Jusnawati, Arifin, "Jusnawati, Arifin, Abd.Asis Pata: Contribution of Paddy Production in SIPILU Center Areas (Sidrap, Pinrang, Luwu) to Paddy Production in South Sulawesi," vol. 12, no. 2, pp. 46–55, 2020.

[8] R. Rustam, S. Rahmatullah, S. Supriyato, and S. Wahyuni, "Application of Data Mining for Predicting Plywood Product Sales at PT Puncak Menara Hijau Mas," *J. Inf. and Computer.* , vol. 8, no. 2, pp. 75–86, 2020, doi: 10.35959/jik.v8i2.186.

[9] Mulyawan, R. Subagja, D. Rohman, and DI Efendi, "Implementation of Support Vector Regression for House Price Prediction with Grid Search Optimization," *J. Electron. Comput. Sci. Udayana* , vol. 12, no. 3, pp. 687–696, 2024.

[10] MA Adam, HY Chrisnanto, and RF Umbara, "Predicting Financial Patterns in the Indonesian Stock Exchange Stock Market Using the Support Vector Machine for Regression (SVR) Algorithm," *J. Education and Research. Inform.* , vol. 10, no. 3, pp. 476–484, 2024.

[11] AM Ahmed, "Scientific Journal of Cihan University – Sulaimaniya," *Sci. J. Cihan Univ. – Sulaimaniya* , vol. 6, no. 1, pp. 145–156, 2022.

[12] K. Watanabe and X. Xu, "Particle swarm optimization based on nonlinear crossing strategy with time-varying acceleration coefficient," *Appl. Intell.* , vol. 54, no. 13–14, pp. 7229–7277, 2024, doi: 10.1007/s10489-024-05502-1.

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5509

[13] D. Pebrianti, H. Kurniawan, L. Bayuaji, and R. Rusdah, "XgBoost Hyper-Parameter Tuning Using Particle Swarm Optimization for Stock Price Forecasting," *J. Ilm. Tech. Elektro Comput. dan Inform.* , vol. 9, no. 4, pp. 1179–1195, 2023, doi: 10.26555/jiteki.v9i4.27712.

[14] E. Bayraktar, I. Ekren, and X. Zhang, "Prediction against finite opponents," *J. Mach. Learn. Res.* , vol. 22, pp. 1–33, 2021.

[15] MH Rizky, MR Faisal, I. Budiman, D. Kartini, and F. Abadi, "The Effect of Hyperparameter Tuning Using Random Search on Tree-Based Classification Algorithms for Software Defect Prediction," *IJCCS (Indonesian J. Comput. Cybern. Syst.* , vol. 18, no. 1, p. 95, 2024, doi: 10.22146/ijccs.90437.

[16] ARM Rom, N. Jamil, and S. Ibrahim, "Multi-objective hyperparameter tuning via random search on deep learning models," *Telkomnika (Telecommunication Comput. Electron. Control.* , vol. 22, no. 4, pp. 956–968, 2024, doi: 10.12928/TELKOMNIKA.v22i4.25847.

[17] M. Arifin, W. Widowati, and F. Farikhin, "Hyperparameter Optimization in Machine Learning to Improve Student Academic Performance Prediction," *Ing. des Syst. d'Information* , vol. 28, no. 3, pp. 575–582, 2023, doi: 10.18280/isi.280305.

[18] M. Sholeh, U. Lestari, and D. Andayati, "Hyperparameter Optimization Using Grid Search and Random Search to Improve the Performance of Prediction Models with Decision Trees," *J. Ris. Multidisciplinary and Inov. Technol.* , vol. 3, no. 03, pp. 453–464, 2025, doi: 10.59653/jimat.v3i03.2025.

[19] SFA Khansa, N. Ulinnuha, and WD Utami, "Optimization of Hyperparameter Tuning of Grid Search and Random Search on Xgboost Algorithm for Parkinson's Disease Classification," *Barekeng* , vol. 19, no. 3, pp. 1609–1624, 2025, doi: 10.30598/barekengvol19iss3pp1609-1624.

[20] S. Pratista, A. Nazir, I. Iskandar, E. Budianita, and I. Afrianty, "Comparison of Drug Use Prediction Techniques Using Simple Linear Regression Algorithms and Support Vector Regression," vol. 5, no. 2, August 2023, doi: 10.47065/bits.v5i2.4260.

[21] T. Aprilia, *APPLICATION OF SUPPORT VECTOR REGRESSION (SVR) AND PARTICLE SWARM OPTIMIZATION (PSO) METHODS FOR FORECASTING RICE PRODUCTION RESULTS* .

[22] S. Erçel and S. Akyol, "Impact of Data Preprocessing Techniques on the Performance of Machine Learning Models for Drought Prediction," *Acadlore Trans. AI Mach. Learn.* , vol. 4, no. 1, pp. 14–24, 2025, doi: 10.56578/ataiml040102.

[23] BM Akbar, AT Akbar, and R. Husaini, "Sentiment and Emotion Analysis of Sinovac Vaccine on Twitter using Naive Bayes and Valence Shifter," *J. Integrated Technol.* , vol. 7, no. 2, pp. 83–92, 2021, doi: 10.54914/jtt.v7i2.433.

[24] K. Purohit, "Separation of Data Cleaning Concepts from EDA," *Int. J. Data Sci. Anal.* , vol. 7, no. 3, p. 89, 2021, doi: 10.11648/j.ijdsa.20210703.16.

[25] S. Saadah and H. Salsabila, "Bitcoin Price Prediction Using the Random Forest Method," *J. Comput. Terap.* , vol. 7, no. 1, pp. 24–32, 2021, doi: 10.35143/jkt.v7i1.4618.

[26] C. Okpara, C. Edokwe, G. Ioannidis, A. Papaioannou, J.D. Adachi, and L. Thabane, "Reporting and handling of missing data in longitudinal studies of older adults remains suboptimal: a methodological survey of geriatrics journals," *BMC Med. Res. Methodol.* , vol. 22, no. 1, pp. 1–9, 2022, doi: 10.1186/s12874-022-01605-w.

[27] Z. Rais, "Support Vector Regression (SVR) Analysis with Kernel Radial Basis Function (Rbf) to Predict Inflation Rate in Indonesia," *VARIANCE J. Stat. Appl. Teach. Res.* , vol. 4, no. 1, pp. 30–38, 2022, doi: 10.35580/variansiunm13.

[28] N. Aldhafferi, "Android Malware Detection Using Support Vector Regression for Dynamic Feature Analysis," *Inf.* , vol. 15, no. 10, 2024, doi: 10.3390/info15100658.