
Application of Support Vector Regression for Rice Production Prediction in Manokwari Regency

Sri Fitrawati^{1*}, Marlinda Sanglise¹, Christian Dwi Suhendra¹

¹University of Papua, Faculty of Engineering, Informatics Engineering, Jl. Gunung Salju, Amban, Manokwari, Papua Barat 98314, Indonesia

Article History

Submitted: 27-06-2026

Revised: 15-07-2026

Accepted: 20-07-2026

Published: 19-09-2026

DOI: 10.32664/j-intech.v14i03.2439

Keywords

Machine Learning; Manokwari Regency; Prediction; Rice Production; Support Vector Regression

*Corresponding Author:

fitrawati.1910@email.com

Abstract

Annual fluctuations in rice production present significant challenges for agricultural planning and regional food security, particularly in areas where agricultural productivity is strongly influenced by environmental and production-related factors. Developing an accurate forecasting model is therefore essential to support evidence-based decision-making and improve agricultural resource management. Although Support Vector Regression (SVR) has demonstrated promising performance in agricultural prediction, previous studies have primarily focused on large-scale datasets or relied on default model parameters, limiting their applicability to local agricultural conditions. Therefore, this study aims to develop and evaluate a GridSearchCV-optimized Support Vector Regression model for predicting annual rice production in Manokwari Regency. Secondary data were obtained from the Agricultural Statistics Database maintained by the Indonesian Ministry of Agriculture, rainfall records from Rendani Meteorological Station, and statistical publications issued by the Central Statistics Agency of Manokwari Regency. The proposed framework integrates data preprocessing, systematic hyperparameter optimization using GridSearchCV, and Support Vector Regression modelling. The predictive capability of the model was assessed using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). The results demonstrate that the selected model employed the Radial Basis Function (RBF) kernel with $C = 1000$, $\gamma = 0.01$, and $\epsilon = 0.001$, achieving an MAE of 460.15, RMSE of 718.86, MAPE of 5.36%, and an R^2 value of 0.9578. These findings demonstrate satisfactory predictive performance and suggest that the proposed model has the potential to serve as a decision-support tool for agricultural production planning and regional food security management in Manokwari Regency.

1. Introduction

Rice is one of the most essential agricultural commodities in Indonesia and serves as the primary staple food for the majority of the population. In addition to ensuring national food security, rice production contributes significantly to regional economic development, employment opportunities, and the sustainability of agricultural systems. Consequently, maintaining stable rice production has become a national priority to support food availability and satisfy the continuously increasing demand driven by population growth.

Despite its strategic importance, annual rice production frequently experiences considerable fluctuations resulting from the interaction of climatic conditions, land availability, and agricultural management practices. These fluctuations create substantial challenges for governments and agricultural stakeholders in estimating future production, allocating resources efficiently, and formulating evidence-based agricultural policies. Reliable production forecasting is therefore essential to support agricultural planning, reduce production uncertainty, and strengthen regional food security through data-driven decision-making.

Manokwari Regency is one of the agricultural regions in West Papua Province with considerable potential for rice cultivation. However, historical production records from 2010 to 2025 indicate substantial year-to-year fluctuations in rice production. Such variability suggests that annual production is influenced by multiple interacting factors rather than a single determinant. Harvested area, productivity, and rainfall are among the primary variables affecting rice production, while climatic variables such as rainfall and temperature have also been reported to influence agricultural output across different regions. Nevertheless, the magnitude of these effects depends on local environmental and geographical characteristics [1]. As a result, prediction models developed using broader geographical datasets may not adequately represent the production characteristics of individual regions such as Manokwari Regency. Moreover, because the relationships among these variables are often nonlinear and highly complex, conventional statistical approaches may not accurately capture agricultural production patterns.

Recent advances in machine learning have significantly improved predictive modelling in the agricultural sector. Compared with conventional statistical approaches, machine learning algorithms are capable of learning complex relationships directly from historical observations without requiring strict assumptions regarding data distribution. This capability makes machine learning particularly suitable for agricultural datasets, where production is influenced by multiple environmental, climatic, and management-related factors. Several machine learning techniques, including Artificial Neural Networks (ANN), Decision Trees, Random Forest (RF), Support Vector Regression (SVR), and Extreme Gradient Boosting (XGBoost), have been successfully implemented for agricultural prediction. Previous studies have demonstrated that these approaches generally outperform traditional linear regression models when handling nonlinear agricultural data [2], [3], [4]. Likewise, Random Forest has been reported to effectively model complex interactions among agricultural variables and achieve satisfactory prediction performance for rice production [5] while comparative analyses indicate that the predictive performance of machine learning algorithms depends on dataset characteristics, feature selection, and model parameter configuration [6].

Among the available machine learning algorithms, Support Vector Regression (SVR), which was developed from the Support Vector Machine framework introduced by Cortes and Vapnik (1995), has attracted considerable attention because of its strong generalization capability and its effectiveness in modelling nonlinear relationships through kernel functions. In particular, the Radial Basis Function (RBF) kernel is widely adopted because it effectively captures nonlinear interactions while maintaining computational efficiency. These characteristics make SVR particularly suitable for agricultural prediction, where relationships among production variables rarely follow simple linear patterns. Previous studies have demonstrated that SVR performs well in agricultural prediction tasks, including paddy growth modelling using Sentinel-1 imagery [7], rice production forecasting under varying climatic conditions [8], provincial-scale rice production prediction [9], and modelling the relationship between climatic variables and rice production [10]. Collectively, these studies indicate that SVR is a promising approach for modelling agricultural datasets with complex and fluctuating patterns.

Although SVR has demonstrated encouraging predictive capability, its performance is highly dependent on appropriate hyperparameter selection. The regularization parameter (C), kernel coefficient (γ), and epsilon-insensitive loss parameter (ϵ) directly influence model complexity, prediction accuracy, and generalization capability. Consequently, inappropriate parameter settings may lead to underfitting or overfitting. To overcome this limitation, GridSearchCV has been widely adopted as a systematic optimization technique that evaluates predefined combinations of hyperparameters using cross-validation to identify the configuration with the best predictive performance. Compared with manual parameter tuning, GridSearchCV provides a more objective, reproducible, and reliable optimization process while reducing the possibility of selecting suboptimal parameter values [11], [12], [13].

Despite the encouraging results reported in previous studies, several limitations in the existing literature remain. Most existing studies have been conducted using provincial or national datasets, which may not adequately capture the production characteristics at the regency level. In addition, differences in environmental conditions, cultivation practices, and resource availability across regions limit the direct applicability of broader-scale prediction models to local agricultural systems. Furthermore, many previous studies employing SVR relied on default hyperparameter settings or limited parameter tuning, even though prediction performance is strongly influenced by hyperparameter configuration. Existing studies also differ considerably in terms of study area, predictor variables, dataset characteristics, and optimization strategies, making their findings difficult to generalize to local agricultural conditions in Manokwari Regency. Moreover, limited studies have simultaneously incorporated harvested area, productivity, and rainfall as predictor variables within a systematically optimized SVR framework. Consequently, the effectiveness of GridSearchCV-based hyperparameter optimization for annual rice production prediction using locally observed agricultural data in Manokwari Regency remains insufficiently investigated.

Motivated by these research gaps, this study develops a Support Vector Regression model optimized using GridSearchCV for annual rice production prediction in Manokwari Regency. Rather than proposing a new prediction algorithm, this study evaluates the applicability of a systematically optimized SVR model using locally observed agricultural data to support regional rice production forecasting in an area that has received limited attention in previous research. The proposed model is intended to improve prediction accuracy and provide a practical decision-support tool for agricultural planning and regional food security management. Accordingly, the primary contribution of this study lies in demonstrating the application and evaluation of an optimized SVR model for rice production prediction under the specific agricultural conditions of Manokwari Regency, thereby providing empirical evidence for the use of machine learning in regional agricultural forecasting.

Based on these considerations, the objective of this study is to develop and evaluate a GridSearchCV-optimized Support Vector Regression model for annual rice production prediction in Manokwari Regency using harvested area, productivity, and rainfall as predictor variables. The predictive performance of the proposed model is evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). The findings are expected to contribute to the advancement of agricultural prediction research while providing practical insights for developing data-driven decision-support systems that support sustainable regional agricultural planning and food security management.

2. Research Method

This study employed a machine learning-based framework for constructing a predictive model for annual rice production in Manokwari Regency. The proposed methodology consisted of six sequential stages: dataset acquisition, data preprocessing, feature normalization, model development, hyperparameter optimization, and model evaluation. The complete research procedure used in this study is shown in Figure 1.

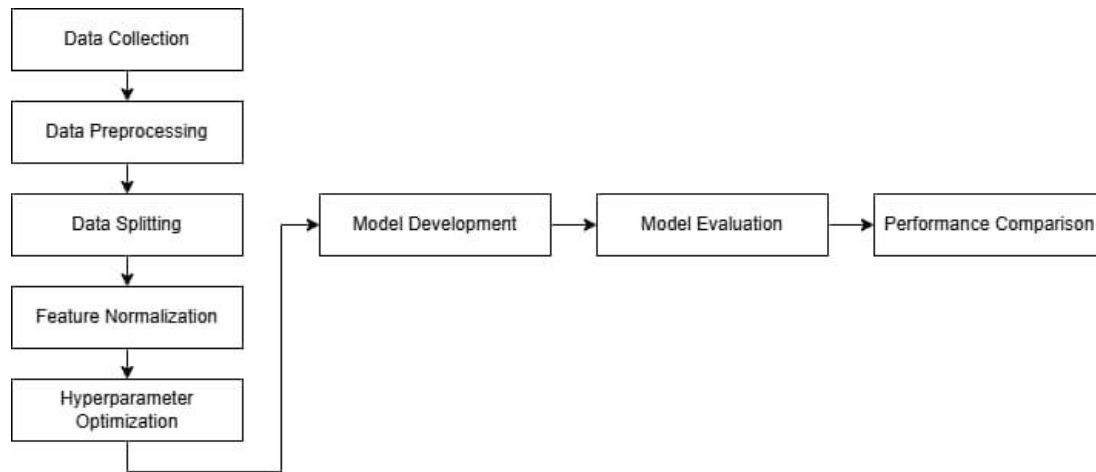


Fig 1. Research Method

2.1 Data Collection

Secondary data were compiled from three official sources: the Agricultural Statistics Database (BDSP) of the Indonesian Ministry of Agriculture [14], the Central Statistics Agency (BPS) of Manokwari Regency [15], and meteorological observations from Rendani Station obtained through the Meteostat database. Annual observations covering the period 2010–2025 were used in this study, resulting in 16 observations. All data used in this study, including the 2025 records, were obtained as officially finalized figures from the respective institutions rather than provisional or preliminary releases. Harvested area (ha), productivity (ku/ha), and rainfall (mm) were considered predictor variables, whereas annual rice production (tons) was defined as the target variable.

Table 1. Research Variables

Variables	Information
Harvested Area (ha)	Input Variables
Productivity (ku/ha)	Input Variables
Rainfall (mm)	Input Variables
Rice Production (tons)	Output Variable

2.2 Data Preprocessing

To ensure that the dataset was suitable for model development, a series of preprocessing procedures was performed prior to model training. First, the dataset was examined for missing values, and no missing observations were identified; therefore, no imputation procedure was required. Potential outliers were then assessed using the Interquartile Range (IQR) method for reporting purposes only, and no observations were removed because the dataset consisted of only sixteen annual records. Pearson correlation analysis was subsequently conducted to examine the relationships among the study variables. Finally, Min–Max normalization was applied to rescale the predictor and target variables into a consistent range. To prevent data leakage, the dataset was first divided into training and testing sets, after which the normalization parameters were fitted exclusively on the training data and subsequently applied to the testing data.

This correlation pattern, examined in detail in Section 3.1, also served as a basis for evaluating variable retention [9], [10], [16]. Pearson correlation coefficients range from -1 to 1 , where coefficients approaching 1 indicate a strong positive relationship, coefficients approaching -1 indicate a strong negative relationship, and values close to 0 indicate a weak or negligible linear association. In this study, the correlation analysis was used as an exploratory analysis rather than as a feature selection technique. Therefore, all predictor variables

harvested area, productivity, and rainfall were retained because they represent important agronomic factors reported in previous studies. Moreover, since Support Vector Regression with an RBF kernel is capable of modeling nonlinear relationships, variables with weak linear correlations may still contribute to predictive performance. To further assess the contribution of each predictor, an ablation experiment using harvested area as the sole predictor was conducted and compared with the full-feature model.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

2.3 Data Splitting

Following the preprocessing stage, the dataset was divided into training and testing subsets following an 80:20 ratio. This partitioning strategy was implemented to provide sufficient observations for model training while maintaining an independent dataset for model performance evaluation, consistent with previous studies on Support Vector Regression [11].

Because the dataset consists of annual time-series observations, the partitioning process preserved the chronological order of the data by disabling random shuffling (`shuffle=False`). Maintaining the temporal sequence prevents information leakage from future observations into the training stage and provides a more realistic assessment of predictive performance under practical forecasting conditions.

The training subset was used to learn the nonlinear relationships between harvested area, productivity, rainfall, and rice production, whereas the testing subset was reserved exclusively for evaluating the predictive capability of the optimized SVR model on previously unseen observations. Furthermore, hyperparameter optimization was performed using GridSearchCV in conjunction with TimeSeriesSplit (`n_splits = 3`), ensuring that each cross-validation fold respected the temporal ordering of the observations and preventing information from future data from being used during model selection. This evaluation strategy provides a more reliable assessment of the model's generalization capability for time-series prediction.

2.4 Feature Normalization

Since the predictor variables have different measurement scales, Min-Max Scaling was applied before model training. This normalization technique transforms each variable into the interval $[0,1]$, preventing variables with wider numerical ranges from exerting excessive influence on the learning process.

Feature normalization also improves numerical stability during optimization and accelerates model convergence. The transformation is expressed in Equation (2). Consistent with the data splitting stage described above, the normalization parameters (minimum and maximum values) were fitted exclusively on the training subset and subsequently applied to transform the testing subset, thereby preventing information from the testing data from influencing the scaling process.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2)$$

2.5 Hyperparameter Optimization

The predictive performance of Support Vector Regression is strongly influenced by the selection of its hyperparameters. Therefore, a systematic hyperparameter optimization procedure was performed using GridSearchCV to identify the most appropriate configuration for the penalty parameter (C), kernel coefficient (γ), and epsilon-insensitive loss parameter (ϵ) within the Radial Basis Function (RBF) kernel.

GridSearchCV was chosen because it performs an exhaustive search across predefined hyperparameter combinations while simultaneously applying cross-validation to determine the parameter configuration producing the highest predictive performance. Compared with manual parameter tuning, this approach

provides a more objective, reproducible, and reliable optimization process. Evidence from previous studies suggests that the predictive capability of Support Vector Regression can be enhanced through systematic hyperparameter optimization in agricultural forecasting tasks [17]

The predefined search space consisted of C values of {1, 10, 100, 1000}, gamma values of {0.001, 0.01, 0.1, 1}, and epsilon values of {0.001, 0.01, 0.1}. Given the limited sample size (12–13 training observations), standard k-fold cross-validation with random fold assignment was considered inappropriate, as it would both produce very small validation folds and disregard the temporal ordering of the data. While Leave-One-Out Cross-Validation (LOOCV) maximizes data usage, its standard implementation does not preserve temporal order, which risks validating on past observations using information from future ones a form of temporal data leakage that is particularly problematic for time-series forecasting tasks such as this one. To address both concerns simultaneously, hyperparameter optimization was instead performed using TimeSeriesSplit, a cross-validation scheme that partitions the data into sequential, non-overlapping folds while strictly preserving chronological order, ensuring that each validation fold is evaluated only on observations that occur after its corresponding training data. Three splits were used, consistent with the original design, while eliminating the risk of temporal leakage. The parameter combination with the best validation performance under this scheme was selected as the final SVR model. Notably, re-running the optimization under this stricter, leakage-free validation scheme yielded the same optimal configuration (C = 1000, gamma = 0.01, epsilon = 0.001) and identical test-set performance as originally reported, indicating that the model's hyperparameter selection is stable and not an artifact of the validation strategy [11], [12], [13].

2.6 Model Development

Support Vector Regression (SVR) was employed as the primary prediction algorithm in this study because of its capability to model complex nonlinear relationships while maintaining strong generalization performance, particularly when applied to relatively small datasets [18]. Unlike conventional regression techniques, SVR estimates an optimal regression function by balancing prediction accuracy and model complexity, making it well suited for agricultural prediction problems involving nonlinear interactions among variables. The regression function of SVR is expressed as Equation (3).

$$f(x) = w^T \phi(x) + b \quad (3)$$

where ($f(x)$) represents the predicted output, (w) denotes a weight vector, ($\phi(x)$) is a nonlinear mapping function that projects the input data into a higher-dimensional feature space, while (b) represents the bias term.

To capture the nonlinear relationships between harvested area, productivity, rainfall, and rice production, this study employed the Radial Basis Function (RBF) kernel. The RBF kernel was selected because it has demonstrated excellent performance in modelling nonlinear agricultural datasets while providing good generalization capability without requiring prior assumptions regarding the functional relationship between variables. The RBF kernel is defined as Equation (4).

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right) \quad (4)$$

Before model training, all predictor variables were scaled using the Min–Max Scaling approach and divided into training and testing datasets following an 80:20 ratio. Hyperparameter optimization was subsequently performed using GridSearchCV to identify the most suitable values of C, gamma, and epsilon. The optimized SVR model obtained from this process was then applied to predict annual rice production using the testing dataset.

2.7 Model Evaluation

The predictive capability of the optimized Support Vector Regression (SVR) model was assessed using the testing dataset, which was not involved during the training process. Evaluating the model on unseen data provides an objective assessment of its generalization capability and predictive reliability.

Four complementary evaluation metrics were utilized: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2) [19]. These metrics collectively evaluate prediction accuracy, sensitivity to large prediction errors, relative percentage error, and the proportion of variance explained by the model. The mathematical formulations of the evaluation metrics are presented in Equations (5)–(8).

Mean Absolute Error (MAE) can be calculated using Equation (5).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

Root Mean Squared Error (RMSE) can be calculated using Equation (6).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

Mean Absolute Percentage Error (MAPE) is calculated using Equation (7).

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (7)$$

The coefficient of determination (R^2) is calculated using Equation (8).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (8)$$

Lower values of MAE, RMSE, and MAPE indicate better prediction accuracy, whereas an R^2 value closer to one reflects a stronger agreement between predicted and observed rice production values, indicating better explanatory capability of the developed model [19].

2.8 Performance Comparison

To evaluate the added value of the GridSearchCV-based hyperparameter optimization and to assess the relative contribution of the predictor variables, the optimized SVR model was benchmarked against four additional configurations using the same training-testing split and evaluation metrics: (1) a simple linear regression model, serving as a baseline representing a conventional statistical approach; (2) an SVR model using default hyperparameters (scikit-learn defaults), to isolate the specific contribution of hyperparameter tuning; (3) a Random Forest Regressor, representing an alternative ensemble-based machine learning algorithm; and (4) an ablation model, consisting of an SVR model tuned with the same hyperparameter configuration but using harvested area as the sole predictor, to examine whether productivity and rainfall meaningfully improve predictive performance beyond what harvested area alone can explain. All comparison models were trained and evaluated using the same MAE, RMSE, MAPE, and R^2 metrics described above, enabling a direct and consistent performance comparison.

3. Result and Discussions

3.1 Correlation Analysis Results

Pearson correlation analysis was performed to investigate the predictor variables and annual rice production prior to model development. This analysis provides preliminary insight into the relative contribution of harvested area, productivity, and rainfall to rice production. The correlation coefficients are presented in Table 2.

Table 2. Pearson Correlation Coefficients between Predictor Variables and Rice Production

Variables	Correlation Value
Harvested Area	0.895
Productivity	-0.205
Rainfall	-0.111

The correlation analysis revealed that harvested area exhibited a strong positive correlation with rice production, with a Pearson correlation coefficient of 0.895. This result indicates that increases in harvested area are generally associated with higher annual rice production. Among the variables considered in this study, harvested area appears to be the dominant predictor influencing production variability, highlighting the importance of cultivated land availability in determining total rice output.

In contrast, productivity showed a weak negative correlation ($r = -0.205$) with rice production. This zero-order correlation, however, does not reflect productivity's true contribution once the influence of harvested area is accounted for. Harvested area itself was found to correlate negatively with productivity ($r = -0.595$) and strongly positively with rice production ($r = 0.895$), indicating a multicollinearity relationship among the predictors. Controlling for this relationship, the partial correlation between productivity and rice production is strongly positive ($r = 0.913$), consistent with agronomic expectations that higher yield per unit area contributes to higher total production. This pattern is characteristic of a suppressor effect, in which a predictor's zero-order correlation with the outcome is masked or reversed by its correlation with another predictor. This explanation is further supported by an ablation analysis comparing the full three-predictor SVR model against a single-predictor model using harvested area alone: the single-predictor model achieved a substantially lower predictive accuracy ($R^2 = 0.890$, $MAE = 1,125.65$) compared to the full model ($R^2 = 0.958$, $MAE = 460.15$), confirming that productivity meaningfully improves predictive performance despite its weak zero-order correlation.

Rainfall similarly exhibited a very weak negative correlation ($r = -0.111$) with rice production. Applying the same partial correlation approach, controlling for harvested area yields a partial correlation of $r = 0.209$ between rainfall and rice production a modest but positive relationship. Unlike productivity, this partial correlation remains weak in magnitude, suggesting that rainfall's limited zero-order correlation is only partly attributable to multicollinearity with harvested area, and that annual rainfall alone genuinely has limited explanatory power for production variability within this dataset. This is plausible given that rainfall effects on rice yield are typically mediated by timing and distribution across the growing season rather than annual totals, a temporal resolution not captured by the present dataset. Nevertheless, its retention as a predictor is supported by the ablation results above, which show that the full model incorporating both productivity and rainfall outperforms the single-predictor alternative. The Pearson correlation coefficients are visualized using the correlation heatmap presented in Figure 2.

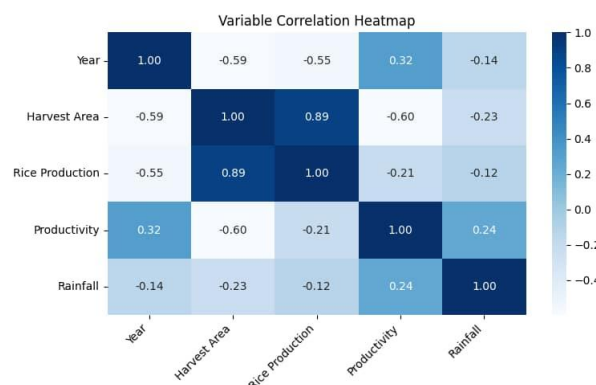


Fig 2. Pearson Correlation Heatmap of Predictor Variables and Rice Production

3.2 SVR Parameter Optimization Results

Hyperparameter optimization was conducted using the GridSearchCV algorithm to identify the optimal parameter configuration for the SVR model. The optimization procedure evaluated different combinations of the penalty parameter C , kernel coefficient (gamma), and epsilon-insensitive loss parameter (epsilon) while employing the Radial Basis Function (RBF) kernel. The optimal hyperparameters obtained through GridSearchCV are presented in Table 3.

Table 3. Optimal Hyperparameter Configuration of the SVR Model

Parameter	Optimal Value
Kernel	RBF
C	1000
Epsilon	0.001
Gamma	0.01

The GridSearchCV optimization identified the RBF kernel with $C = 1000$, $\gamma = 0.01$, and $\epsilon = 0.001$ as the optimal hyperparameter configuration. This combination achieved the best cross-validation performance among all evaluated parameter sets and was therefore selected for the final prediction model.

The relatively large value of C indicates that the model places greater emphasis on minimizing prediction errors during training while maintaining an appropriate balance between model complexity and generalization capability. Meanwhile, the selected gamma value (0.01) allows the RBF kernel to effectively capture nonlinear relationships among the predictor variables without producing an excessively complex decision function. Furthermore, the small epsilon value (0.001) reduces the tolerance for prediction errors within the insensitive loss region, enabling the regression function to fit the observed data more precisely.

Overall, the optimized hyperparameter configuration demonstrates that systematic parameter tuning using GridSearchCV substantially improves the capability of the SVR model to represent nonlinear relationships in annual rice production data. The selected parameter combination was therefore adopted for all subsequent prediction and evaluation experiments.

3.3 Model Evaluation Results

The predictive capability of the optimized Support Vector Regression (SVR) model was evaluated using four complementary evaluation metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). These metrics collectively assess prediction accuracy, sensitivity to large prediction errors, relative percentage error, and the model's explanatory capability. The assessment results are summarized in Table 4.

Table 4. SVR Model Evaluation Results

Evaluation Metric	Value
MAE	460.15
RMSE	718.886
MAPE	5.36%
R^2	0.9578

The optimized SVR model produced an MAE of 460.15 along with an RMSE of 718.886, indicating relatively small prediction errors between actual and predicted rice production values. The low magnitude of these errors suggests that the model is capable of accurately estimating annual rice production while maintaining stable predictive performance.

The model also produced a MAPE of 5.36%, indicating a very low average percentage error. In forecasting studies, a MAPE value below 10% is generally considered to represent highly accurate prediction performance. Therefore, the obtained MAPE demonstrates that the proposed SVR model provides reliable forecasts for annual rice production in Manokwari Regency.

Furthermore, the coefficient of determination ($R^2 = 0.9578$) suggest that approximately 95.78% of the variability in annual rice production can be explained by harvested area, productivity, and rainfall. This high explanatory power confirms that the selected predictor variables effectively capture the dominant production patterns represented in the dataset. Beyond the numerical evaluation metrics, model performance was further examined by comparing the predicted values with the corresponding observed rice production values, as illustrated in Figure 3.

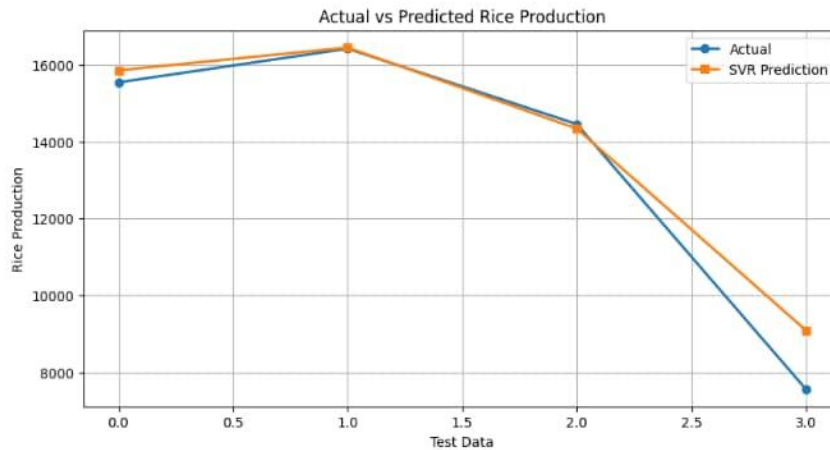


Fig 3. Comparison of Actual and Predicted Values of Rice Production Using SVR

Figure 3 demonstrates that the predicted values closely follow the overall trend of the observed production data throughout the testing period. Although slight discrepancies are visible in several observations, these deviations remain relatively small and do not substantially affect the overall prediction trend. This result indicates that the optimized SVR model successfully captures the underlying nonlinear relationship between the predictor variables and annual rice production while maintaining good generalization capability on previously unseen data.

3.4 Discussion

The results demonstrate that the GridSearchCV-optimized Support Vector Regression (SVR) model provides robust predictive performance for annual rice production forecasting in Manokwari Regency. The combination of a high coefficient of determination ($R^2 = 0.9578$) and low prediction error values (MAE = 460.15, RMSE = 718.886, and MAPE = 5.36%) indicates that the developed model effectively captures the nonlinear relationships among harvested area, productivity, rainfall, and rice production. These findings confirm that SVR, when combined with systematic hyperparameter optimization, is well suited for modeling agricultural datasets characterized by complex and nonlinear interactions.

The correlation analysis further revealed that harvested area is the most influential predictor of annual rice production, exhibiting a strong positive correlation ($r = 0.895$). This finding is consistent with the fundamental principle that the total cultivated area directly determines the potential production capacity of a region. In contrast, productivity and rainfall exhibited weak negative zero-order correlations with rice production. As shown in the correlation analysis, this pattern is largely explained by multicollinearity with harvested area rather than a genuine inverse relationship: the partial correlation between productivity and production after controlling for harvested area was strongly positive ($r = 0.913$), consistent with agronomic expectations, whereas rainfall's partial correlation remained comparatively weak ($r = 0.209$). This distinction indicates that productivity's apparent negative association is a suppressor effect driven by its relationship with harvested

area, while rainfall's limited explanatory power is more likely genuine, potentially because annual rainfall totals do not capture the seasonal timing and distribution that more directly affect rice yield.

The optimization process played a significant role in improving model performance. Rather than relying on default parameter values, GridSearchCV systematically evaluated multiple hyperparameter combinations to identify the configuration that achieved the best validation performance using TimeSeriesSplit. The optimal configuration ($C = 1000$, $\gamma = 0.01$, and $\epsilon = 0.001$) produced the highest predictive accuracy among the evaluated candidates. Notably, the optimal value of C was located at the upper boundary of the predefined search space, suggesting that the true optimum may lie beyond the evaluated range. Consequently, future studies should consider exploring a wider range of C values to determine whether additional performance improvements can be achieved. Furthermore, although the optimized model achieved a relatively low prediction error, its practical significance can be illustrated by expressing the MAE relative to the average annual rice production in the dataset (mean = 15,356.02 tons): an MAE of 460.15 tons corresponds to approximately 3.0% of average annual production, indicating that prediction errors are practically small relative to typical production volumes in Manokwari Regency. As discussed in the correlation analysis above, the apparent weak negative correlations of productivity and rainfall are largely attributable to multicollinearity with harvested area rather than a genuine lack of predictive contribution. Finally, because the model was developed using historical observations from 2010–2025, its predictive capability for future production beyond the observation period should be interpreted cautiously and requires further validation using newly available data.

The evaluation results also demonstrate that the optimized model possesses strong predictive capability. The low values of MAE, RMSE, and MAPE indicate that prediction errors remain relatively small, while the high R^2 value suggests that the selected predictor variables explain most of the observed variation in annual rice production. Collectively, these findings indicate that the proposed SVR model can provide reliable production forecasts and may serve as an effective decision-support tool for agricultural planning and regional food security management.

The results of this study are in line with earlier studies reporting the effectiveness of Support Vector Regression for agricultural prediction [7], [8], [10], [20]. To substantiate this claim, the optimized SVR model was benchmarked against three alternative approaches using the same dataset and evaluation scheme: simple linear regression, an SVR model using default hyperparameters, and Random Forest.

Table 5. Performance Comparison of SVR (Tuned) Against Baseline and Ablation Models

Model	MAE	RMSE	MAPE (%)	R^2
SVR + GridSearchCV (tuned, TimeSeriesSplit)	460.15	718.86	5.36	0.9578
Linear Regression (baseline)	724.02	1096.83	8.26	0.9018
SVR default hyperparameter (baseline)	2028.82	3653.67	25.37	-0.0896
Random Forest (baseline)	1861.14	2306.64	19.04	0.5657
SVR tuned - harvested area only (ablation)	1125.65	1159.77	9.72	0.8902

As summarized in Table 5, the optimized SVR model achieved the lowest prediction error (MAE = 460.15, RMSE = 718.86, $R^2 = 0.958$), outperforming linear regression (MAE = 724.02, $R^2 = 0.902$), Random Forest (MAE = 1,861.14, $R^2 = 0.566$), and notably, SVR with default hyperparameters (MAE = 2,028.82, $R^2 = -0.090$). The markedly poor performance of the default-parameter SVR — performing worse than a naïve mean-based prediction — demonstrates that the predictive gains reported in this study are attributable specifically to the GridSearchCV-based hyperparameter optimization procedure, rather than to the SVR algorithm alone. This finding quantifies the specific contribution of systematic parameter optimization and confirms that, unlike earlier studies that primarily employed default parameters or focused on broader geographic regions, adapting model configuration to local, district-level agricultural data through systematic tuning substantially improves prediction performance. This contribution highlights the importance of adapting machine learning models to regional agricultural characteristics rather than relying solely on generic model configurations.

From a practical perspective, the proposed prediction model has the potential to support local governments and agricultural agencies in estimating annual rice production before harvest. Reliable production forecasts can assist decision-makers in planning food distribution, allocating agricultural resources, anticipating production shortages, and formulating evidence-based food security policies. Therefore, the implementation of machine learning approaches such as optimized SVR into agricultural decision-support systems may contribute to more effective and sustainable agricultural management.

4. Conclusions and Future Works

This study successfully developed and evaluated a GridSearchCV-optimized Support Vector Regression (SVR) model for annual rice production prediction in Manokwari Regency using harvested area, productivity, and rainfall as predictor variables. The optimization process identified the optimal hyperparameter configuration consisting of an RBF kernel, $C = 1000$, $\gamma = 0.01$, and $\epsilon = 0.001$, resulting in satisfactory predictive performance with an MAE of 460.15, an RMSE of 718.86, a MAPE of 5.36%, and an R^2 value of 0.9578. These findings demonstrate that combining Support Vector Regression with systematic hyperparameter optimization provides an effective approach for modeling nonlinear agricultural data and generating reliable annual rice production forecasts.

The findings of this study are significant because they demonstrate that systematic hyperparameter optimization can substantially improve the predictive capability of Support Vector Regression for agricultural forecasting, particularly when applied to relatively limited local datasets. Rather than proposing a new algorithmic technique, this study contributes empirical evidence on the applicability of a systematically optimized SVR model for district-level agricultural forecasting, an area that has received limited attention in previous research. The proposed gap in performance between the tuned and default-parameter SVR models further underscores the practical importance of systematic hyperparameter optimization in applied agricultural prediction settings. Practically, the proposed model provides a reliable decision-support tool that can assist local governments and agricultural agencies in production planning, resource allocation, food distribution, and the formulation of evidence-based food security policies.

Despite these promising findings, several methodological limitations should be acknowledged. The study utilized annual rice production data from Manokwari Regency covering the period from 2010 to 2025, resulting in a dataset of only sixteen observations. With an 80:20 training-testing split, the model was trained on approximately 12–13 observations and evaluated on only 3–4 test observations. Performance metrics derived from such a limited test set, including the high coefficient of determination ($R^2 = 0.958$), should therefore be interpreted with caution rather than as a definitive measure of generalization performance, as a test set of this size cannot support statistically reliable point estimates. While a formal confidence interval could not be constructed given the limited number of independent observations, the stability of the reported results was partially supported by the TimeSeriesSplit cross-validation procedure and by comparison against multiple alternative models, both of which yielded consistent findings. Furthermore, the model incorporated only three predictor variables harvested area, productivity, and rainfall while other influential factors, such as soil characteristics, temperature, irrigation systems, fertilizer application, pest infestation, and agricultural management practices, were not considered. Finally, because the model was developed using data from a single geographical region, the findings cannot yet be generalized to other agricultural areas.

Future research should focus on addressing the identified limitations in this study. First, longer observation periods and additional environmental and agricultural variables should be incorporated to improve model robustness and generalizability. Second, given the dominant correlation of harvested area with rice production, future studies should further investigate whether a simplified prediction model using harvested area as the primary predictor can achieve performance comparable to that of the full-feature model. Third, because the optimal hyperparameter value for C reached the upper boundary of the predefined search space ($C = 1000$), future experiments should evaluate a broader search range to determine whether further performance improvements can be achieved. In addition, approaches for increasing the effective amount of training data, such as integrating spatially distributed agricultural data from neighboring regions, may help improve model

stability and predictive reliability. Overall, this study demonstrates the practical applicability of a GridSearchCV-optimized Support Vector Regression model for annual rice production prediction in Manokwari Regency and provides empirical evidence supporting the use of machine learning as a decision-support tool for regional agricultural planning and food security.

5. Declaration

Author Contributions

The first and second supervisors contributed to the conceptualization, methodology, and research design of the study, as well as providing feedback and revisions prior to manuscript submission. The author was responsible for data collection, data analysis, implementation of the SVR model, interpretation of the results, manuscript preparation, and subsequent revisions following submission.

Funding Statement

This research received no external funding.

Conflict of Interest

The author declares no conflict of interest.

Data Availability Statement

The data used in this study were obtained from the official websites of the relevant government institutions and are publicly available.

Use of Artificial Intelligence (AI)

AI tools were used to assist with language editing, improve the clarity of the manuscript, and understand the journal formatting requirements.

Ethical Clearance

Ethical clearance was not required for this study because it used secondary data and did not involve human participants or animal subjects.

6. References

- [1] M. I. Kharisma S, A. I. Hadiana, and E. Ramadhan, "Model Prediksi Produksi Padi Berdasarkan Curah Hujan dan Suhu Menggunakan Regresi Linier Berganda," *Jurnal Algoritma*, vol. 22, no. 2, pp. 704–716, 2025, doi: 10.33364/algoritma/v.22-2.2793.
- [2] L. Fatimah, Martanto, A. R. Dikananda, and A. Rifa'i, "Algoritma Regresi Linear untuk Prediksi Hasil Panen dan Strategi Produksi Padi di Kabupaten Cirebon," *Jurnal Informatika Teknologi dan Sains (JINTEKS)*, vol. 7, no. 2, pp. 464–472, 2025.
- [3] H. Putra and N. Ulfa, "Jurnal Nasional Teknologi dan Sistem Informasi Penerapan Prediksi Produksi Padi Menggunakan Artificial Neural Network Algoritma Backpropagation," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 06, no. 02, pp. 100–107, 2020.
- [4] Setiawan Cahyono and Muhammad Imron Rosadi, "Penerapan Artificial Neural Network untuk Prediksi Produksi Padi di Sumatera," *Jurnal Informatika Polinema*, vol. 11, no. 4, pp. 487–494, 2025, doi: 10.33795/jip.v11i4.7727.
- [5] D. Manurung, B. Zealtiel, and A. H. Lubis, "Prediksi Produksi Tanaman Padi di Indonesia dengan Menggunakan Algoritma Random Forest Regressor," *Journal of Computing and Informatics Research*, vol. 4, no. 3, pp. 345–345, 2025, doi: 10.47065/comforch.v4i3.2125.

- [6] F. Yasin, M. R. Firmansyah, D. Aldo, and M. A. Amrustian, "Multivariate Forecasting of Paddy Production: A Comparative Study of Machine Learning Models," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 3, pp. 1431–1442, 2025, doi: 10.52436/1.jutif.2025.6.3.4681.
- [7] H. Muradi, A. Saefuddin, I. M. Sumertajaya, A. M. Soleh, and D. D. Domiri, "Support Vector Regression (Svr) Method for Paddy Growth Phase Modeling Using Sentinel-1 Image Data," *Media Statistika*, vol. 16, no. 1, pp. 25–36, 2023, doi: 10.14710/medstat.16.1.25-36.
- [8] R. Yunis, Sudarto, and I. Adiputra Pardosi, "Enhancing Rice Production Prediction: A Comparative Machine Learning Analysis of Climate Variables," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 13, no. 1, pp. 91–104, 2024, doi: 10.23887/janapati.v13i1.71527.
- [9] E. E. Sijabat, "Support Vector Regression-Based Prediction of Rice Production Across Provinces in Sumatra Island," *Brilliance: Research of Artificial Intelligence*, vol. 5, no. 2, pp. 1199–1206, 2025, doi: 10.47709/brilliance.v5i2.7429.
- [10] H. Purnomo, R. Estian Pambudi, O. Arifin, F. Kurniawan Ikhsan, P. Negeri Lampung, and I. Darmajaya Bandar Lampung, "Analisis Prediksi Produksi Tanaman Padi Berdasarkan Variabel Iklim Menggunakan Support Vector Regression," *Aisyah Journal of Informatics and Electrical Engineering*, vol. 07, no. 02, pp. 10–16, 2025.
- [11] A. W. Ishlah, S. Sudarno, and P. Kartikasari, "Implementasi Gridsearchcv Pada Support Vector Regression (Svr) Untuk Peramalan Harga Saham," *Jurnal Gaussian*, vol. 12, no. 2, pp. 276–286, 2023, doi: 10.14710/j.gauss.12.2.276-286.
- [12] I. Muhamad Malik Matin, "Hyperparameter Tuning Menggunakan GridsearchCV pada Random Forest untuk Deteksi Malware," *Multinetics*, vol. 9, no. 1, pp. 43–50, 2023, doi: 10.32722/multinetics.v9i1.5578.
- [13] G. H. Saputra, A. H. Wigena, and B. Sartono, "Penggunaan Support Vector Regression Dalam Pemodelan Indeks Saham Syariah Indonesia Dengan Algoritme Grid Search," *Indonesian Journal of Statistics and Its Applications*, vol. 3, no. 2, pp. 148–160, 2019, doi: 10.29244/ijsa.v3i2.172.
- [14] Kementerian Pertanian Republik Indonesia, "Basis Data Statistik Pertanian Indonesia." [Online]. Available: <https://bdsp2.pertanian.go.id/bdsp/id/indikator>
- [15] Badan Pusat Statistik Kabupaten Manokwari, "Jumlah Curah Hujan (mm), 2022." Accessed: Apr. 21, 2026. [Online]. Available: <https://manokwarikab.bps.go.id/id/statistics-table/2/NDEjMg==/jumlah-curah-hujan.html>
- [16] D. I. P. Desy, A. F. Riza Kholdani, Tri Wahyu Qur'ana, and A. Dharmawati, "Pemodelan Spasial untuk Analisa Produksi Padi Integrasi Machine Learning," *Digital Zone: Jurnal Teknologi Informasi dan Komunikasi*, vol. 14, no. 2, pp. 128–137, 2023, doi: 10.31849/digitalzone.v14i2.16256.
- [17] S. Rahmah, R. Suhendra, H. Maghfirah, Sanusi, and I. Raziah, "IMPLEMENTASI ALGORITMA SUPPORT VECTOR REGRESSION DALAM MEMPREDIKSI HASIL PRODUKSI TANAMAN PANGAN PADA KABUPATEN ACEH BARAT DAYA," pp. 19–29, 2025.
- [18] C. Cortes and V. Vapnik, "Support-Vector Networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1023/A:1022627411411.
- [19] F. Wahyu, D. Wicaksono, and A. B. Rahmat, "Optimasi Peramalan Tingkat Laju Inflasi Indonesia Melalui Pendekatan Ensemble Bagging Pada Algoritma Support Vector Regression," vol. 12, no. 2, 2024.
- [20] D. R. Anamisa, B. D. Satoto, M. K. Sophan, and M. Yusuf, "Mathematical Modelling of Engineering Problems Development of the Support Vector Regression and Genetic Algorithm Hybrid Models for Forecasting of Madura Rice Yield," vol. 13, no. 1, pp. 181–198, 2026.