BAREKENG: Journal of Mathematics and Its Applications
March 2026 Volume 20 Issue 1 Page 0367–0380
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss1pp0367-0380

PERFORMANCE EVALUATION OF SEASONAL ARIMA-SVR AND SEASONAL ARIMAX-SVR HYBRID METHODS ON FORECASTING PADDY PRODUCTION

I'lmisukma Risnawati [1*], Farit Mochamad Afendi [2], I Made Sumertajaya [3]

[1,2,3]*School of Data Science, Mathematics and Informatics, IPB University*
Jln. Dramaga, Bogor, Jawa Barat, 16680, Indonesia

*Corresponding author's e-mail: * **rwati992017@gmail.com***

Article Info

Article History:
Received: 25th March 2025
Revised: 27th April 2025
Accepted: 20th July 2025
Available online: 24th November 2025

Keywords:
ARIMA;
ARIMAX;
Hybrid;
Non-linear;
SVR.

ABSTRACT

This study explores advances in forecasting time series data by combining linear and non-linear models. Traditional methods such as ARIMA and its variant ARIMAX are effective for linear data but have limitations when dealing with non-linearity. Support Vector Regression (SVR), a non-linear method, complements these weaknesses. Hybrid models such as ARIMA-SVR and ARIMAX-SVR synergize ARIMA or ARIMAX for linear components and SVR for non-linear components, improving accuracy. The purpose of this study is to evaluate the performance of hybrid ARIMA-SVR and ARIMAX-SVR methods on Indonesian paddy production data. The data analyzed is national-level data per sub-round (i.e., three sub-rounds per year) from sub-round 1 (January-April) of 1992 to sub-round 3 (September-December) of 2024, obtained from the Indonesian Central Statistics Agency and the Indonesian Ministry of Agriculture.Forecasting accuracy is measured using Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The results show that the best model is the Seasonal ARIMAX (1,1,1)(0,1,1)[3]-SVR ($C = 2^3$; $\gamma = 2^2$; $\varepsilon = 0.05$) hybrid model, with the smallest RMSE and MAPE values of 0.304 and 1.473%. The addition of the harvested area variable and the ASF dummy improved the accuracy of the ARIMAX model prediction, while the application of SVR to ARIMAX residuals successfully captured previously undetected linear patterns. Based on these considerations, the Seasonal ARIMAX(1,1,1)(0,1,1)[3]-SVR ($C = 2^3$; $\gamma = 2^2$; $\varepsilon = 0.05$) hybrid model was selected as the model with the best performance.

How to cite this article:

I. Risnawati, F. M. Afendi, I Made Sumertajaya., " PERFORMANCE EVALUATION OF SEASONAL ARIMA-SVR AND SEASONAL ARIMAX-SVR HYBRID METHODS ON FORECASTING PADDY PRODUCTION," *BAREKENG: J. Math. & App.*, vol. 20, iss. 1, pp. 0367-0380, Mar, 2026.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Autoregressive Integrated Moving Average (ARIMA) is the most commonly used traditional technique for forecasting univariate time series data. ARIMA is considered a simple model because it uses only historical data of the variable itself. The popularity of ARIMA is due to its good statistical properties and ability to build reliable models [1]. However, this approach is less effective when dealing with extreme noise or fluctuations in data. A variant of ARIMA, the Autoregressive Integrated Moving Average with Exogenous (ARIMAX), incorporates additional variables that significantly influence the data to improve accuracy [2]. Li et al. employed the ARIMAX method to analyze the impact of Internet search data on suspected case trends in the COVID-19 surveillance system [3]. Similarly, Susila explored how Idul Fitri affects inflation using ARIMAX [4]. Aprilianto et al. utilized the ARIMAX model to predict stock prices in the healthcare industry by integrating exogenous factors such as opening price, highest and lowest prices, and stock volume [5]. Despite their advantages, ARIMA and ARIMAX have weaknesses in handling non-linear problems, as their accuracy decreases.

The Support Vector Regression (SVR) method is appropriate for forecasting time series data with non-linear patterns. SVR is an adaptation of the Support Vector Machine (SVM) method designed for regression tasks [6]. While ARIMA, ARIMAX, and SVR individually offer different advantages in addressing linear and non-linear model challenges, combining these methods enhances predictive capabilities [7].

Since time series data often contains both linear and non-linear patterns, hybrid approaches integrating linear and non-linear models are necessary. These methods are expected to improve forecasting performance by reducing errors. The ARIMA-SVR and ARIMAX-SVR hybrid methods combine ARIMA and ARIMAX for linear components and SVR for non-linear components to generate more accurate forecasts. Previous studies have explored hybrid methods, including forecasting closing stock prices using ARIMA-SVR [8] and predicting short-term electricity network maximum demand with ARIMAX-SVR [9].

Paddy is one of the most critical and strategic commodities for the Indonesian population. Before the introduction of the Area Sampling Framework (ASF) method, data collection on paddy-harvested areas was still carried out traditionally through the reporting of the Agricultural Statistics Report (ASR). Since 2018, the ASF method has been implemented to improve accuracy [10]. This change in calculating the size of harvested areas is considered to be one of the causes of the changes in paddy production patterns. Additionally, the harvested area factor is also thought to significantly impact the production levels.

The quality of paddy production is reported in terms of dry unhusked paddy. In 2024, Indonesia's paddy production reached 53.14 million tons, a decrease of 838.27 thousand tons (1.55%) compared to 2023, which amounted to 53.98 million tons [11]. The production decline in 2024 occurred in several potential areas, including West Java, East Java, and Central Java.

Numerous studies have explored the forecasting of paddy production through time series analysis. Munasingha and Napagoda used the ARIMA method for forecasting rice production [12]. while Nurviana et al. predicted paddy production in Aceh Province using ARIMA alongside Exponential Smoothing techniques [13]. Andita and Sulistijanti improved paddy production forecasting in Kendal Regency using the Support Vector Machine method [14]. However, these studies only use one linear or non-linear approach, which stands alone on univariate time series data. In contrast, this study combines techniques for both processes (linear and non-linear) on univariate and multivariate time series using ARIMA-SVR and ARIMAX-SVR methods.

This study aims to evaluate the performance of ARIMA-SVR and ARIMAX-SVR hybrid methods in forecasting paddy production in Indonesia by combining linear and non-linear approaches alongside the influence of exogenous factors, such as harvested area and changes in data collection methods. These results are expected to provide more accurate forecasting recommendations to support agricultural sector policies. This study used data breakdowns into three periods per year, with each period referred to as a subround. Subround 1 data represented the total for January–April, subround 2 covered May–August, and subround 3 included September–December. The optimal method will be used to forecast paddy production for the upcoming six periods of the sub-round.

2. RESEARCH METHODS

2.1 Research Data

The variables in this study are presented in Table 1. The variables include paddy production and two covariates, namely paddy harvested area and a dummy variable for data collection by the ASF method. The empirical data consists of four monthly figures or three sub-rounds per year, with 99 time series observations from sub-round 1 in 1992 to sub-round 3 in 2024. The data represents national-level data obtained from Statistics Indonesia (BPS) and the Ministry of Agriculture. The paddy harvested area was selected as a covariate due to its direct impact on production volume, enabling the model to capture production trends. Meanwhile, data collection method changes were used as a dummy variable to account for visible differences in production figures before and after 2018.

Table 1. Research Variable

Variable	Variable Name	Units
Response variable		
Y	Paddy production	Million tons of dry unhusked paddy
Covariates		
X_1	Paddy harvested area	Million hectares
X_2	Dummy variable for data collection with ASF	1: Yes, and 0: No

The analysis steps in this research are the following:

1. Analyzing the dataset by visualizing the time series to detect recurring trends or patterns.
2. Data will be separated into two parts: sub-round 1 in 1992 to sub-round 3 in 2019 as the training data, and sub-round 1 in 2020 to sub-round 3 in 2024 as the testing data.
3. Form an ARIMA model using training production data, then perform non-linear testing on the ARIMA residual using the Terasvirta test before constructing the SVR model.
4. Combining the forecast results from the ARIMA model with those generated by the SVR model to obtain the predictions of the ARIMA-SVR hybrid model.
5. Form an ARIMAX model using training production data, then perform non-linear testing on the ARIMAX residual using the Terasvirta test before constructing the SVR model.
6. Combining the forecast results from the ARIMAX model with those generated by the SVR model to obtain the predictions of the ARIMAX-SVR hybrid model.
7. Calculate the Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) for each model using the test data, then compare the results. The minimum RMSE and MAPE metrics determine the optimal model.

2.2 ARIMA

The ARIMA model is a linear time series framework incorporating autoregressive (AR) and moving average (MA) components alongside elements addressing data non-stationarity [15]. The standard structure of the ARIMA (p,d,q) model equation is as follows:

$$\phi_p B(1-B)^d Y_t = \delta + \theta_q(B)\varepsilon_t \tag{1}$$

where (p, d, q) represent the AR, differencing, and MA orders, respectively; δ denotes a constant, B refers to the backshift operators, Y_t indicates the time series variable Y at time t, ϕ_p signifies the AR coefficient at the order p, $(1-B)^d$ corresponds to the differencing series of order d; θ_q represent the MA coefficient at order q, and ε_t stands for the error term at time t.

The development of the ARIMA method for seasonally patterned time series data is the seasonal ARIMA method. The general form of the seasonal ARIMA model is $(p, d, q)(P, D, Q)^S$ is as follows [16]:

$$\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D Y_t = \delta + \theta_q(B)\Theta_Q(B^S)\varepsilon_t \tag{2}$$

where (P, D, Q) represent the AR, differencing, and MA orders, respectively; S denotes a seasonal period, Y_t indicates the time series variable Y at time t, Φ_P signifies the seasonal AR coefficient at the P order, $(1-B^S)^D$ corresponds to the seasonal differencing series of D order; Θ_Q represent the MA coefficient at Q order, and ε_t stands for the error term at time t.

In ARIMA modeling, data is assumed to be stationary or can be made so using techniques such as differencing for the mean and Box-Cox transformation for the variance. Additionally, the residuals of the ARIMA model are expected to follow a normal distribution and remain free of autocorrelation.

The steps in constructing the ARIMA model are as follows:

1. Plotting the data to determine the appropriate transformation to achieve stationarity.
2. Identify the order of the ARIMA model using ACF, PACF plots, and Extended Autocorrelation Function (EACF). In the EACF table, the "X" sign indicates a significant value, while the "O" sign indicates an insignificant value. The appropriate model is determined by finding the empty corner pattern in the EACF table, which indicates the optimal combination of p and q.
3. Predicting parameters and evaluating their significance.
4. Perform model diagnostics and overfitting to choose the optimal model by considering the lowest AIC value. The following formula expresses the method for determining the AIC:
$$AIC = ln\left(\frac{\sum_{t=1}^{n} e_t^2}{n}\right) + \frac{2h}{n} \tag{3}$$
where e_t are the residuals from the model fit in time t, n represents the data within the model, and h denotes the number of parameters.
5. Forecasting with the best ARIMA model, then analyze using RMSE and MAPE to determine the accuracy level.

2.3 ARIMAX

ARIMAX is an improvement of ARIMA by including covariates that are considered significant explanatory variables. Adding other covariates to a time series model is intended to enhance the precision of the prediction. The assumptions in ARIMA also apply to ARIMAX. The basic form of the ARIMAX (*p*,*d*,*q*) model is as follows [17]:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \cdots + \beta_k X_{k,t} + \frac{\theta_q(B)}{\phi_p(B)(1-B)^d}\varepsilon_t \tag{4}$$

The general form of the seasonal ARIMAX model $(p,d,q)(P,D,Q)^s$ is as follows [18]:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \cdots + \beta_k X_{k,t} + \frac{\theta_q(B)\Theta_Q(B^S)}{\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D}\varepsilon_t \tag{5}$$

where Y_t is the time series variable, $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ regression coefficients, and $X_{1,t}, X_{2,t}, \dots, X_{k,t}$ are the covariates.

The steps in constructing the ARIMAX model are as follows:

1. Modeling ARIMAX using the best ARIMA model by adding covariates;
2. Predicting parameters, evaluating their significance, and performing diagnostic examinations;
3. Forecasting the ARIMAX model, then analyze using RMSE and MAPE to determine the accuracy level.

2.4 Support Vector Regression (SVR)

A practical model for forecasting non-linear time series data is SVR. As a machine learning-based model, SVR can recognize patterns in time series data and produce more accurate forecasts [19]. Kernels are a critical element of non-linear mapping in SVR models [20]. The procedure includes separating the dataset into two parts: one for training and the other for testing. This training data establishes a regression function with specific deviation constraints to generate predictions that closely align with the actual values. Suppose $f(x)$ is the function of regression (hyperplane) in the SVR method as follows:

$$f(x) = w.\phi(x) + b \tag{6}$$

where x represents the input, w and b are constant vectors, and $\phi(x)$ denotes a non-linear function. The core concept of the SVR algorithm lies in determining the optimal w and b parameters to solve the optimization problem. This study applies the Radial Basis Function (RBF) kernel to the SVR model. The non-linear SVR regression function with RBF kernel is as follows [21]:

$$f(x_i) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) exp\,(-\gamma\|x_i - x\|^2) + b \tag{7}$$

where α_i, α_i^* is a Lagrange multiplier, $exp(-\gamma\|x_i - x\|^2)$ is a RBF kernel, and b is a constant.

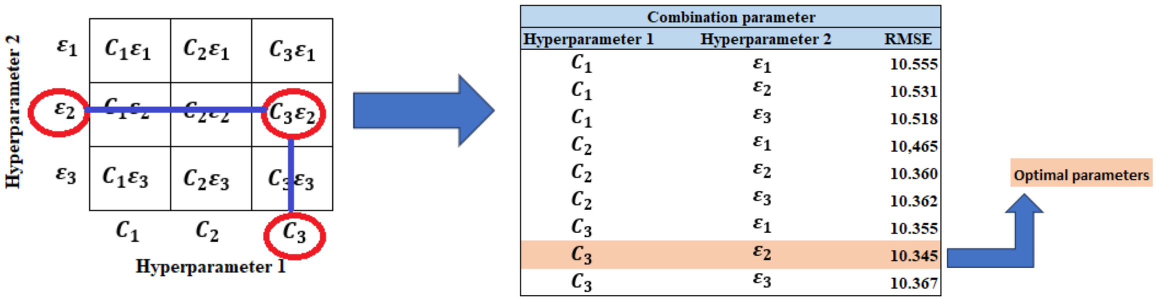


Figure 1. An Illustration of The Grid Search Process

The RBF kernel requires the selection of the regularization parameter (C), kernel coefficients (γ), and the insensitivity parameter (ε). The grid search algorithm systematically explores all possible combinations of predefined hyperparameters. These combinations are first organized into grids, where each pair is used to train the model and evaluated based on the RMSE criterion. Fig. 1 presents a simple illustration of the grid search process with two hyperparameters, C and ε. The combination located in the i^{th} row and j^{th} column that yields the smallest RMSE is selected as the optimal model configuration.

Table 2. Hyperparameter of RBF Kernel

Parameter	Values
C	$2^{-7}, 2^{-6}, \dots, 2^{3}$
γ	$2^{-3}, 2^{-2}, \dots, 2^{7}$
ε	0.05, 0.06, 0.1

The value ranges for the RBF kernel hyperparameters shown in Table 2 are used to identify the optimal SVR model in this research.

2.5 The Hybrid Methods

Paddy production is influenced by various external factors, in this case including harvested area size and changes in data collection methods, both of which often exhibit nonlinear patterns. SVR is well-suited for capturing these complex relationships, outperforming linear models like ARIMA and ARIMAX. This rationale supports the adoption of hybrid methods in the analysis.

The steps in constructing the hybrid model are as follows:

1. Create a linear component using the ARIMA or ARIMAX model. Then, the rest of the linear model is assumed to contain a non-linear relationship. The rest of the linear component, according to Pakrooh and Pishbahar is as follows [22]:

$$e_t = Y_t - \hat{L}_t \tag{8}$$

   where e_t is the residual of the linear model, Y_t is the time series data, and $\hat{L}_t$ represents the prediction generated by the linear model. Eq. (8) shows that e_t is the residual of the linear model, representing the data portion not captured by the ARIMA/ARIMAX model. This residual indicates a non-linear pattern that can be further analyzed using SVR.

2. Create a non-linear component with the SVR model. The input used in the SVR model is the linear (ARIMA or ARIMAX) model residual. The SVR model is as follows:

$$\hat{N}_t = f(e_{t-1}, e_{t-2}, \dots, e_{t-n}) + \varepsilon_t \tag{9}$$

   where f represents a non-linear function in the SVR framework, and ε_t is a random error that represents the uncertainty in the model. Eq. (9) illustrates how SVR transforms historical residuals (e) into a non-linear function (f), generating a predicted non-linear value ($\hat{N}_t$).

3. Create a hybrid model by integrating the forecasting outcomes of the ARIMA or ARIMAX and SVR models to construct the hybrid model. The hybrid model is represented as follows:

$$\hat{Y}_t = \hat{L}_t + \hat{N}_t \tag{10}$$

Eq. (10) indicates that the final prediction ($\hat{Y}_t$) is obtained by combining the linear component from the ARIMA or ARIMAX model ($\hat{L}_t$) with the non-linear component of the SVR model ($\hat{N}_t$).

2.6 Performance Evaluation

The best model is selected based on the results of the forecasting accuracy evaluation, using RMSE and MAPE criteria. RMSE is more sensitive to outliers, while MAPE is more commonly used due to its percentage-based format, which simplifies interpretation. The combination of RMSE and MAPE enables a more comprehensive assessment.

Table 3. Interpretation of MAPE Values

MAPE Value	Interpretation
> 50%	Inaccurate forecasting
20% - 50%	Reasonable forecasting
10% - 20%	Good forecasting
< 10%	High accurate forecasting

Table 3 shows the MAPE value intervals and their interpretation. The best model is obtained when the MAPE and RMSE values are the minimum among other models. Islam and Alam present the formulas for calculating MAPE and RMSE as follows [23]:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \tag{11}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| x\ 100\% \tag{12}$$

where n is the number of data, y_t is the observed value, and $\hat{y}_t$ is the predicted value.

3. RESULTS AND DISCUSSION

3.1 Data Exploration

The graphs shown in Fig.2 below shows data plots related to national paddy production and harvested areas. Data was collected from sub-round 1 in 1992 to sub-round 3 in 2017 using the ASR method. Meanwhile, from sub-round 1 in 2018 to sub-round 3 in 2024, the method used is the ASF. In the ASR, overestimation may occur because data collection often relies on visual estimates, whereas ASF uses standard rice field area maps obtained from remote sensing technology and utilizes Android devices for direct observation. This minimizes the potential for bias in production trends after 2018 so that although production figures appear smaller with the ASF method, the data produced is more accurate.

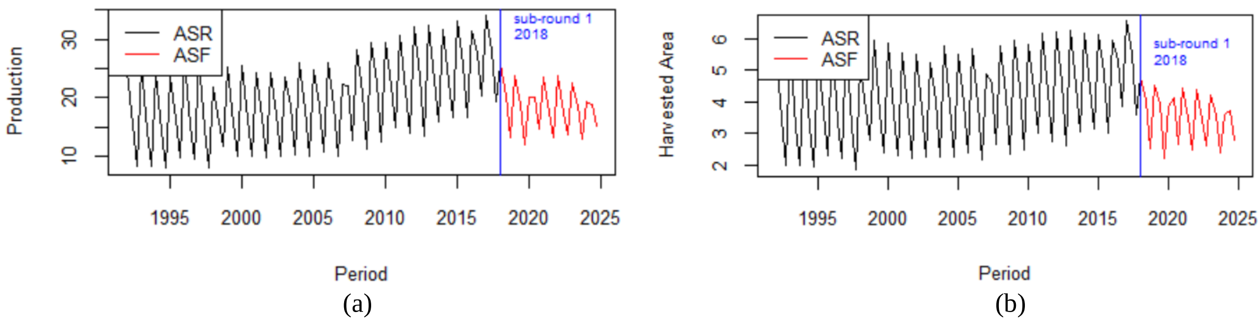


Figure 2. (a) The Production and (b) Harvested Area

The visualization in Fig. 2 illustrates a downward trend in the harvested area following the implementation of KSA in sub-round 1 of 2018. This trend aligns with the decline in production during the same period.

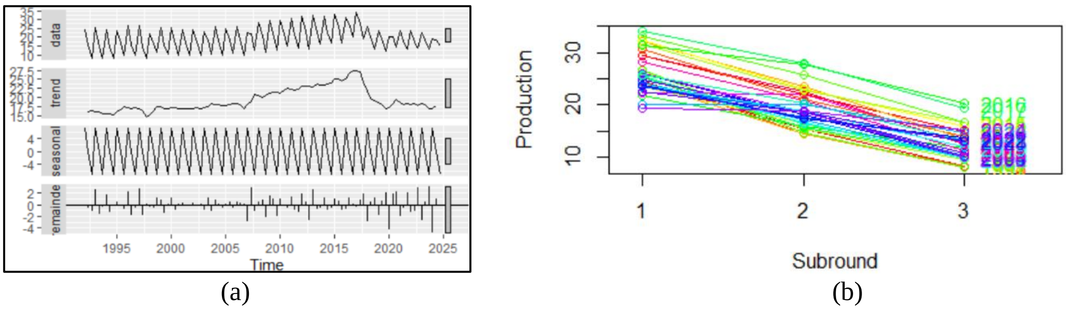


Figure 3. (a) The Decomposition Plot of Production and (b) Seasonal Period of Production

Based on Fig. 3, it can be observed that there is a seasonal period of three (i.e., three sub-rounds per year) in paddy production. This is evident from the recurring pattern every three sub-rounds, indicating a seasonal cycle.

3.2 ARIMA-SVR Hybrid Modeling

The production data are first made stationary before constructing the ARIMA model. After applying the first differencing to the non-seasonal (d = 1) and seasonal components (D = 1), the adjusted data are use are used for analysis. Fig. 4 shows the identification results of the non-seasonal component orders. Lag 3 in the ACF and PACF plots exceeds the significant limit, so the order p = 3 and q = 3 is selected. Based on this evaluation, the identified ARIMA models are ARIMA (0,1,3), ARIMA (3,1,0), and ARIMA (3,1,3). Additionally, the EACF analysis suggests p = 1 and q = 1, resulting in the ARIMA (1,1,1) model.

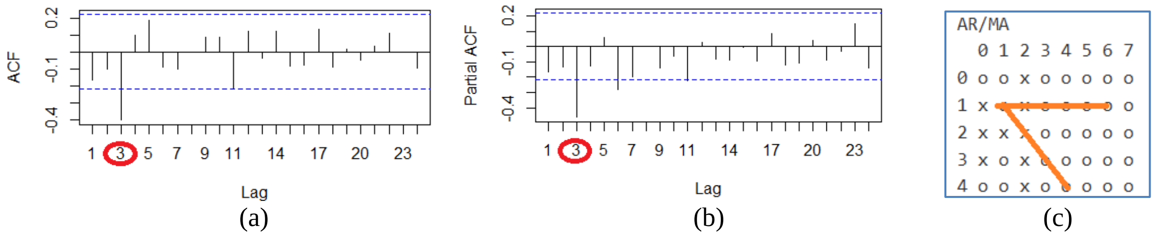


Figure 4. Identification of The ARIMA Order of The Non-Seasonal Component Based on (a) The ACF, (b) PACF, and (c) EACF

The identification results of the seasonal component orders are shown in Fig. 5. From the ACF and PACF plots and the visualization in Figure 3, it can be assumed that the model of the data is seasonal ARIMA with a period of 3. Fig. 5 (a) shows that the first lag is significant beyond the significance limit on the ACF plot, and Fig. 5 (b) shows that the second lag is significant beyond the significance limit on the PACF plot, so the orders P = 2 and Q = 1 are determined. The ARIMA order identification results based on ACF and PACF are ARIMA (2,1,0)[3], ARIMA (0,1,1)[3], and ARIMA (2,1,1)[3]. Based on Figs. 5 and 6 results, 12 tentative models were formed.

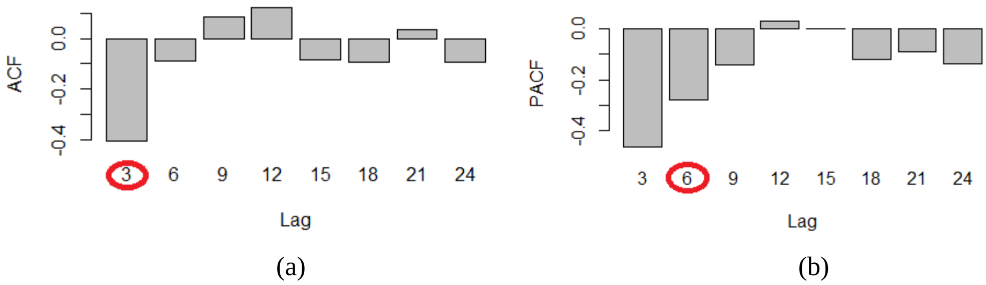


Figure 5. Identification of The ARIMA Order of The Seasonal Component Based on ACF (a) and PACF (b)

Table 4 shows the tentative model with the smallest AIC value compared to several other models among the 12 tentative models that could have been formed. The AIC value comparison results show that the seasonal ARIMA (1,1,1)(0,1,1)[3] is chosen as the best model, because it has the minimum AIC and all parameter estimates are significant.

Table 4. ARIMA Model Identification

Model	AIC	Parameters	Coefficients	p-value	Significance of The Model
Seasonal ARIMA (0,1,3)(0,1,1)[3]	339.72	MA (1)	-0.3920	0.0004	Significant
		MA (2)	-0.2047	0.1153	Not Significant
		MA (3)	-0.0851	0.7599	Not Significant
		SMA (1)	-0.4712	0.0517	Not Significant
Seasonal ARIMA (1,1,1)(0,1,1)[3]	337.63	AR (1)	0.4552	0.02831	Significant
		MA (1)	-0.8530	1.049×10^{-7}	Significant
		SMA (1)	-0.4848	9.118×10^{-5}	Significant
Seasonal ARIMA (1,1,1)(2,1,0)[3]	338.43	AR (1)	0.4612	0.0235	Significant
		MA (1)	-0.8557	7.09×10^{-8}	Significant
		SAR (1)	-0.4599	0.0002	Significant
		SAR (2)	-0.2833	0.0191	Significant

The diagnostic checks for the seasonal ARIMA (1,1,1)(0,1,1)[3] are presented in Table 5. The Ljung-Box test has a p-value more than α = 0.05. This means that the residual white noise. The Jarque Bera test has a p-value less than α = 0.05, which means that the residuals are not normally distributed.

Table 5. ARIMA Diagnostic Check

Test Statistics	p-value
Ljung-Box	0.6098
Jarque Bera	2.22×10^{-16}

The overfitting model was checked with the seasonal ARIMA (2,1,1)(0,1,1)[3]. The AR (1) and AR (2) parameters in the model seasonal ARIMA (2,1,1)(0,1,1)[3] were not significant. Moreover, its AIC value (339.61) was greater than that of the seasonal ARIMA (1,1,1)(0,1,1)[3] model. Therefore, the seasonal ARIMA (1,1,1)(0,1,1)[3] was proposed as the most appropriate model.

Then, the residuals of the seasonal ARIMA (1,1,1)(0,1,1)[3] model are tested for non-linearity. The non-linearity Terasvirta test in Table 6 shows a non-linear pattern because the p-value is $0.04 < \alpha$ (0.05).

Table 6. Non-linearity Test of ARIMA Residuals

Residual	p-value	Pattern
Seasonal ARIMA (1,1,1)(0,1,1)[3]	0.0439	Non-linear

The next step is SVR modeling with the input of residual from the seasonal ARIMA (1,1,1)(0,1,1)[3] model and the hyperparameter from Table 2. The smallest RMSE of SVR is 4.758, obtained when $C = 2^{-7}$; $\gamma = 2^{-3}$; $\varepsilon = 0.08$ (Table 7). Small C and γ values tend to make the model more stable and capable of generalizing data well.

Table 7. Grid Search Process of SVR Using ARIMA Residuals Input

Combination	C	γ	ε	RMSE
1	2^{-7}	2^{-3}	0.05	4.856182
:	:	:	:	:
4	2^{-7}	2^{-3}	0.08	4.758639
:	:	:	:	:
726	2^{3}	2^{7}	0.1	5.182708

The graph in Fig. 6 further confirms that the SVR forecast results significantly deviate from the ARIMA model residual values. This indicates that applying the SVR model to ARIMA residuals may not yield optimal results in this case. The ARIMA-SVR hybrid forecast model is obtained by adding the results of the seasonal ARIMA (1,1,1)(0,1,1)[3] forecast with the results of the SVR forecast.

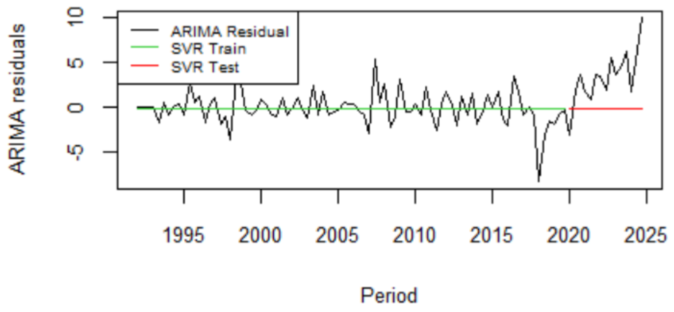


Figure 6. Comparison of ARIMA Residuals and SVR Forecast

3.3 ARIMAX-SVR Hybrid Modeling

The ARIMAX model is built by adding covariates of harvested area (X_1) and the ASF dummy (X_2) when building the model. The regression analysis revealed that the covariates significantly affected paddy production, with an R-squared value of 0.9526. This suggests that 95.26% of the variation in paddy production can be attributed to data on harvested area and the ASF data collection method, while the remaining 4.74% is influenced by other factors. The next step is to build an ARIMAX model based on tentative ARIMA models. The ARIMAX model is presented in Table 8.

Table 8. ARIMAX Model Identification

Model	AIC	Parameters	Coefficients	p-value	Significance of The Model
Seasonal ARIMAX (0,1,3)(0,1,1)[3]	106.73	MA (1)	-0.4148	0.0029	Significant
		MA (2)	-0.3133	0.0127	Significant
		MA (3)	-0.2718	0.0342	Significant
		SMA (1)	-0.2084	0.1019	Not Significant
		X_1	4.8037	$< 2.2 \times 10^{-16}$	Significant
		X_2	-0.4203	0.3286	Not Significant
Seasonal ARIMAX (1,1,1)(0,1,1)[3]	**105.57**	**AR (1)**	**0.5444**	$\mathbf{1.346 \times 10^{-7}}$	**Significant**
		MA (1)	**-0.9999**	$\mathbf{< 2.2 \times 10^{-16}}$	**Significant**
		SMA (1)	**-0.3256**	**0.0007**	**Significant**
		$\mathbf{X_1}$	**4.8018**	$\mathbf{< 2.2 \times 10^{-16}}$	**Significant**
		$\mathbf{X_2}$	**-0.3938**	**0.3613**	**Not Significant**
Seasonal ARIMAX (1,1,1)(2,1,0)[3]	106.18	AR (1)	0.5478	1.735×10^{-7}	Significant
		MA (1)	-1.0000	$< 2.2 \times 10^{-16}$	Significant
		SAR (1)	-0.3957	0.0009	Significant
		SAR (2)	-0.1419	0.2424	Not Significant
		X_1	4.8060	$< 2.2 \times 10^{-16}$	Significant
		X_2	-0.4963	0.2764	Not Significant

The AIC value comparison results show that the seasonal ARIMAX (1,1,1)(0,1,1)[3] model has the minimum AIC value, and all parameter estimates except the coefficient of X_2 are significant. The diagnostic checks for the seasonal ARIMAX (1,1,1)(0,1,1)[3] are presented in Table 9. The Ljung-Box test has a p-value more than $\alpha = 0.05$. This means that the residual white noise. The Jarque Bera test has a p-value less than $\alpha = 0.05$, which means that the residuals are not normally distributed.

Table 9. ARIMA Diagnostic Check

Test Statistics	p-value
Ljung-Box	0.9699
Jarque Bera	0.0005

The overfitting model was checked with the seasonal ARIMAX (2,1,1)(0,1,1)[3]. The AR (2) parameters and the coefficient of X_2 were not significant. Moreover, its AIC value (107.30) was greater than that of the seasonal ARIMAX (1,1,1)(0,1,1)[3] model. Therefore, the seasonal ARIMAX (1,1,1)(0,1,1)[3] was proposed as the best model.

Then, the residuals of the seasonal ARIMAX (1,1,1)(1,1,0)[3] model are tested for non-linearity. The non-linearity test in Table 10 shows a linear pattern because the p-value is $0.41 > \alpha$ (0.05). The findings suggest that the ARIMAX is sufficient to forecast the data, so there is no need for SVR modeling. However, SVR modeling is still carried out to see if there is an improvement in accuracy compared to the findings from the ARIMAX analysis. The input used is the residual of the seasonal ARIMAX (1,1,1)(0,1,1)[3] model.

Table 10. Non-linearity Test of ARIMAX Residuals

Residual	p-value	Pattern
Seasonal ARIMAX (1,1,1)(0,1,1)[3]	0.4461	Linear

Based on Table 11, the smallest RMSE occurs when $C = 2^3$; $\gamma = 2^2$; $\varepsilon = 0.05$. The ARIMAX-SVR hybrid forecast model is obtained by adding the results of the seasonal ARIMAX (1,1,1)(0,1,1)[3] forecast with the results of the SVR forecast.

Table 11. Grid Search Process of SVR Using ARIMAX Residuals Input

Combination	C	γ	ε	RMSE
1	2^{-7}	2^{-3}	0.05	0.399131
:	:	:	:	:
691	2^3	2^2	0.05	0.344228
:	:	:	:	:
726	2^3	2^7	0.1	0.436876

The graph in Fig. 7 shows that the SVR model's forecast results followed the seasonal residual pattern of ARIMAX (1,1,1)(0,1,1)[3]. This indicates that SVR successfully captured the nonlinear pattern that remained after the linear component had been handled by ARIMAX. The ARIMA-SVR hybrid forecast model is obtained by adding the results of the seasonal ARIMA (1,1,1)(0,1,1)[3] forecast with the results of the SVR forecast

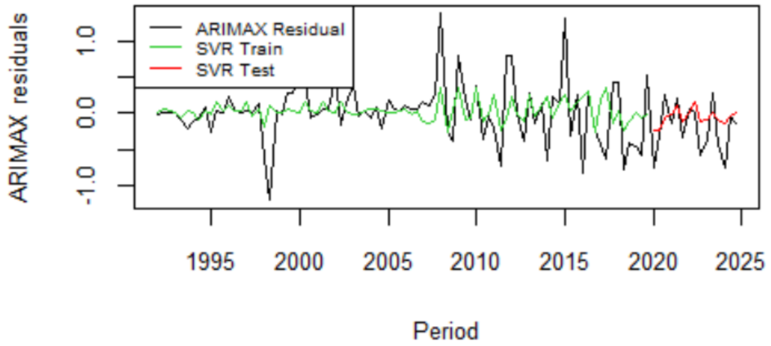


Figure 7. Comparison of ARIMAX Residuals and SVR Forecast

3.4 Performance Evaluation

The forecast accuracy for evaluating the performance of the seasonal ARIMA, ARIMAX, ARIMA-SVR, and ARIMAX-SVR hybrid models is shown in Table 12. Numerically, SVR modeling of seasonal ARIMA residuals did not improve accuracy, as the RMSE and MAPE values for the seasonal ARIMA model testing data were smaller than those for the seasonal ARIMA-SVR hybrid model. Therefore, in this case, SVR modeling of seasonal ARIMA residuals was not considered particularly necessary.

Conversely, SVR modeling of seasonal ARIMAX residuals successfully improved model performance, as evidenced by the RMSE and MAPE values in the seasonal ARIMAX-SVR hybrid model being smaller than those in the seasonal ARIMAX model. Thus, in this case, SVR modeling of seasonal ARIMAX residuals was considered quite necessary. Among them, the seasonal ARIMAX (1,1,1)(0,1,1)[3]-SVR hybrid model is identified as the most effective, achieving the lowest RMSE and MAPE values.

Table 12. Accuracy of The Model

Model	RMSE		MAPE (%)	
	Training	Testing	Training	Testing
Seasonal ARIMA (1,1,1)(0,1,1)[3]	1.830	4.536	6.991	23.739
Seasonal ARIMA (1,1,1)(0,1,1)[3]-SVR Hybrid	1.825	4.699	6,896	24.799

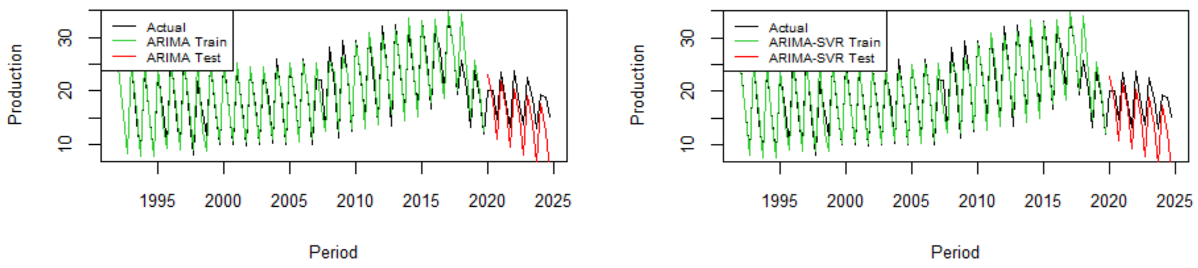
Model	RMSE		MAPE (%)	
	Training	Testing	Training	Testing
Seasonal ARIMAX (1,1,1)(0,1,1)[3]	0.413	0.385	1.436	1.832
Seasonal ARIMAX (1,1,1)(0,1,1)[3]-SVR Hybrid	**0.389**	**0.304**	**1.309**	**1.473**

To further ensure that the difference in model performance was statistically significant, the Diebold-Mariano (DM) test was conducted. This test compared the errors of two forecasting models against the actual data. Table 13 shows that the SVR model significantly improves the performance of seasonal ARIMAX but has minimal impact on seasonal ARIMA.

Table 13. DM Test

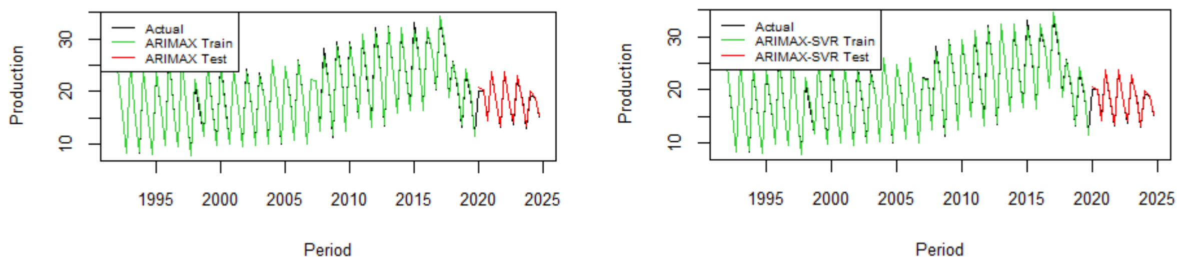
Model	p-value	Accuracy Differences Between Models
Seasonal ARIMA (1,1,1)(0,1,1)[3] Vs Seasonal ARIMA(1,1,1)(0,1,1)[3]-SVR	0.9998	Not Significant
Seasonal ARIMA (1,1,1)(0,1,1)[3] Vs Seasonal ARIMAX(1,1,1)(0,1,1)[3]	0.0039	Significant
Seasonal ARIMA (1,1,1)(0,1,1)[3] Vs Seasonal ARIMAX(1,1,1)(0,1,1)[3]-SVR	0.0038	Significant
Seasonal ARIMAX (1,1,1)(0,1,1)[3] Vs Seasonal ARIMAX(1,1,1)(0,1,1)[3]-SVR	0.0123	Significant
Seasonal ARIMA (1,1,1)(0,1,1)[3]-SVR Vs Seasonal ARIMAX(1,1,1)(0,1,1)[3]-SVR	0.0032	Significant

A comparison between the actual data and the predicted results for both the training and testing datasets is presented in Figs. 8 – 9. The results show that the seasonal ARIMA-SVR hybrid model does not significantly improve predictions compared to the regular seasonal ARIMA model. Likewise, the forecasts from the seasonal ARIMAX-SVR hybrid model are only slightly better than those from the seasonal ARIMAX model alone. However, both seasonal ARIMAX and seasonal ARIMAX-SVR hybrid outperform seasonal ARIMA and seasonal ARIMA-SVR hybrid.



(a) (b)

Figure 8. The Comparison of Forecast Seasonal ARIMA and Actual Data (a), The Comparison of Forecast Seasonal ARIMA-SVR Hybrid and Actual Data



(a) (b)

Figure 9. The Comparison of Forecast Seasonal ARIMAX and Actual Data (a), The Comparison of Forecast Seasonal ARIMAX-SVR Hybrid and Actual Data

3.5 Forecasting with Best Model

The seasonal ARIMAX (1,1,1)(0,1,1)[3]-SVR hybrid model was used to forecast paddy production values. Table 14 shows the results of paddy production forecasting from sub-round 1 (January-April) 2025, to sub-round 3 (September-December) 2026. The forecast results show a decrease in paddy production from 23.413 million tons of dry unhusked paddy in sub-round 1 in 2025 to 22.161 million tons of dry unhusked paddy in sub-round 1 in 2026. Meanwhile, sub-rounds 2 and 3 in 2025 experienced an increase compared to the same period in 2026.

Table 14. Forecasting of Paddy Production

Period	Forecast Value (million tons)
Sub-round 1 2025	23.413
Sub-round 2 2025	18.094
Sub-round 3 2025	13.867
Sub-round 1 2026	22.161
Sub-round 2 2026	18.601
Sub-round 3 2026	14.537

The time series plot of the forecast value for the next 6 periods is shown in Fig. 10. The green line represents the seasonal ARIMAX-SVR hybrid forecast, while the blue dotted line indicates the historical average production, and the red dotted lines mark the historical minimum and maximum ranges as reference points for reasonable values.

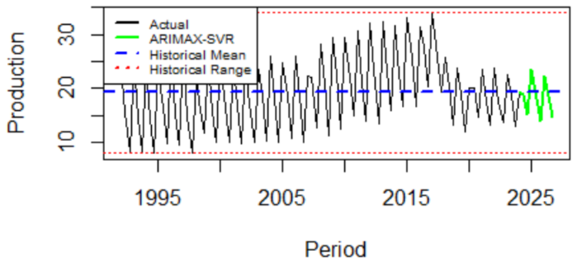


Figure 10. Forecast Plot for Paddy Production in Sub-round 1 2025 to Sub-round 3 2026

The forecast figures from the seasonal ARIMAX-SVR hybrid model are within the historical range, i.e., between the minimum and maximum values of actual past data. In general, it can be concluded that the predictions are within reasonable limits, indicating that the model has learned the pattern stably.

4. CONCLUSION

Significantly, the seasonal ARIMAX-SVR hybrid produces more accurate forecasts than other models. This improvement can be attributed to the inclusion of harvested area covariates and ASF dummy variables, which significantly enhance predictive performance within the seasonal ARIMAX model to forecast paddy production. Furthermore, applying SVR to the residuals of the seasonal ARIMAX model effectively captures remaining linear patterns that were previously undetected. As a result, the seasonal ARIMAX-SVR hybrid model was identified as the most appropriate forecasting approach to forecast paddy production. Future research could incorporate additional covariates such as climatic conditions (rainfall variability, extreme temperatures, droughts) that likely influence paddy production trends. Integrating these factors into models like ARIMAX or hybrid approaches may improve forecast accuracy and provide deeper insights into paddy production dynamics, supporting evidence-based policies for sustainable food security.

Author Contributions

I'lmisukma Risnawati: Conceptualization, Data Curation, Writing-Original Draft, Software. Farit Mochamad Afendi: Methodology, Formal Analysis, Supervision. I Made Sumertajaya: Writing-Review AND Editing, validation. All authors discussed the results and contributed to the final manuscript.

Funding Statement

This research was supported by funding from the Agricultural Human Resources Development Agency of the Ministry of Agriculture.

Acknowledgment

The authors thank the Agricultural Human Resources Development Agency of the Ministry of Agriculture, which supported this research.

Declarations

The authors confirm that there are no conflicts of interest to disclose regarding this study. No affiliations, financial relationships, or personal connections exist that could affect the objectivity or validity of the research fundings.

REFERENCES

[1] G. Chitikela *et al.*, "ARTIFICIAL-INTELLIGENCE-BASED TIME-SERIES INTERVENTION MODELS TO ASSESS THE IMPACT OF THE COVID-19 PANDEMIC ON TOMATO SUPPLY AND PRICES IN HYDERABAD, INDIA," *Agronomy*, vol. 11, no. 9, Sep. 2021, doi: https://doi.org/10.3390/agronomy11091878.

[2] A. Kurnia and I. Made Sumertajaya, "ANALISIS INFLASI MENGGUNAKAN DATA GOOGLE TRENDS DENGAN MODEL ARIMAX DI DKI JAKARTA *," 2020.

[3] K. Li *et al.*, "INTERNET SEARCH DATA COULD BE USED AS NOVEL INDICATOR FOR ASSESSING COVID-19 EPIDEMIC," *Infect Dis. Model*, vol. 5, 2020, doi: https://doi.org/10.1016/j.idm.2020.10.001.

[4] M. R. Susila, "PENGARUH HARI RAYA IDUL FITRI TERHADAP INFLASI DI INDONESIA DENGAN PENDEKATAN ARIMAX (VARIASI KALENDER)," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 14, no. 3, 2020, doi: https://doi.org/10.30598/barekengvol14iss3pp369-378.

[5] M. Aprilianto, M. N. S. Hasibuan, and S. Z. Harahap, "FORECASTING HEALTH SECTOR STOCK PRICES USING ARIMAX METHOD," *Sinkron*, vol. 7, no. 2, 2022, doi: https://doi.org/10.33395/sinkron.v7i2.11418.

[6] D. I. Purnama, "PERAMALAN HARGA EMAS SAAT PANDEMI COVID-19 MENGGUNAKAN MODEL HYBRID AUTOREGRESSIVE INTEGRATED MOVING AVERAGE - SUPPORT VECTOR REGRESSION," *Jambura Journal of Mathematics*, vol. 3, no. 1, pp. 52–65, Jan. 2021, doi: https://doi.org/10.34312/jjom.v3i1.8430.

[7] Y. Pan, X. Jin, Y. Li, D. Chen, and J. Zhou, "A STUDY ON THE PREDICTION OF BOOK BORROWING BASED ON ARIMA-SVR MODEL," in *Procedia CIRP*, Elsevier B.V., 2021, pp. 93–102. doi: https://doi.org/10.1016/j.procs.2021.05.057.

[8] L. Rubio and K. Alba, "FORECASTING SELECTED COLOMBIAN SHARES USING A HYBRID ARIMA-SVR MODEL," *Mathematics*, vol. 10, no. 13, Jul. 2022, doi: https://doi.org/10.3390/math10132181.

[9] H. F. Chow, "SHORT-TERM ELECTRICITY GRID MAXIMUM DEMAND FORECASTING WITH THE ARIMAX-SVR MACHINE LEARNING HYBRID MODEL," *HKIE Transactions Hong Kong Institution of Engineers*, vol. 28, no. 1, 2021, doi: https://doi.org/10.33430/V28N1THIE-2020-0005.

[10] Badan Pusat Statistik, "LUAS PANEN DAN PRODUKSI PADI DI INDONESIA 2021 (ANGKA TETAP)," *BPS*, vol. 2021, no. 21, 2022.

[11] BPS-Statistics Indonesia, "PADDY HARVESTED AREA AND PRODUCTION IN INDONESIA 2024 (FINAL FIGURES)". *Official Statistics News*, No. 15/02/Th. XXVIII, 2025.

[12] M. A. P. Munasingha and N. A. D. N. Napagoda, "TREND ANALYSIS AND FORECASTING FOR PADDY PRODUCTION IN SRI LANKA," *Applied Economics & Business*, vol. 5, no. 2, 2021, doi: https://doi.org/10.4038/aeb.v5i2.33.

[13] N. Nurviana, A. Amelia, R. P. S. Riezky Purnama Sari, U. N. Ulya Nabilla, and T. Talib, "FORECASTING RICE PADDY PRODUCTION IN ACEH USING ARIMA AND EXPONENTIAL SMOOTHING MODELS," *CAUCHY: Jurnal Matematika Murni dan Aplikasi*, vol. 7, no. 2, 2022, doi: https://doi.org/10.18860/ca.v7i2.13701.

[14] A. Andita and W. Sulistijanti, "PERBAIKAN PERAMALAN PRODUKSI PADI DI KABUPATEN KENDAL DENGAN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)," *Proceeding of The URECOL*, 2018.

[15] A. Asrirawan, S. U. Permata, and M. I. Fauzan, "PENDEKATAN UNIVARIATE TIME SERIES MODELLING UNTUK PREDIKSI KUARTALAN PERTUMBUHAN EKONOMI INDONESIA PASCA VAKSINASI COVID-19," *Jambura Journal of Mathematics*, vol. 4, no. 1, pp. 86–103, Jan. 2022, doi: https://doi.org/10.34312/jjom.v4i1.11717.

[16] D. C. Montgomery, C. L. Jennings, and M. Kulahci, "INTRODUCTION TO TIME SERIES ANALYSIS AND FORECASTING," *Wiley Series in Probability and Statistics*, 2015.

[17] M. L. S. Putera, "IMPROVISASI MODEL ARIMAX-ANFIS DENGAN VARIASI KALENDER UNTUK PREDIKSI TOTAL TRANSAKSI NON-TUNAI," *Indonesian Journal of Statistics and Its Applications*, vol. 4, no. 2, 2020, doi: https://doi.org/10.29244/ijsa.v4i2.603.

[18] Y. P. Chen *et al.*, "MODELING AND PREDICTING PULMONARY TUBERCULOSIS INCIDENCE AND ITS ASSOCIATION WITH AIR POLLUTION AND METEOROLOGICAL FACTORS USING AN ARIMAX MODEL: AN

ECOLOGICAL STUDY IN NINGBO OF CHINA," *Int J Environ Res Public Health*, vol. 19, no. 9, 2022, doi: https://doi.org/10.3390/ijerph19095385.

[19] N. Puspita, F. M. Afendi, and B. Sartono, "COMPARISON OF SARIMA, SVR, AND GA-SVR METHODS FOR FORECASTING THE NUMBER OF RAINY DAYS IN BENGKULU CITY," *BAREKENG: J. Il. Mat. & Ter*, vol. 16, no. 1, pp. 355–362, 2022, doi: https://doi.org/10.30598/barekengvol16iss1pp353-360.

[20] J. J. Zou *et al.*, "APPLICATION OF A COMBINED MODEL WITH SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE AND SUPPORT VECTOR REGRESSION IN FORECASTING HAND-FOOT-MOUTH DISEASE INCIDENCE IN WUHAN, CHINA," *Medicine (United States)*, vol. 98, no. 6, Feb. 2019, doi: https://doi.org/10.1097/MD.0000000000014195.

[21] D. I. Purnama and S. Setianingsih, "SUPPORT VECTOR REGRESSION (SVR) MODEL FOR FORECASTING NUMBER OF PASSENGERS ON DOMESTIC FLIGHTS AT SULTAN HASANUDIN AIRPORT MAKASSAR," *Jurnal Matematika, Statistika dan Komputasi*, vol. 16, no. 3, 2020, doi: https://doi.org/10.20956/jmsk.v16i3.9176.

[22] P. Pakrooh and E. Pishbahar, "FORECASTING AIR POLLUTION CONCENTRATIONS IN IRAN, USING A HYBRID MODEL," *Pollution*, vol. 5, no. 4, pp. 739–747, 2019, doi: https://doi.org/10.22059/poll.2019.274827.572.

[23] F. Islam and M. Alam, "USE OF TELECONNECTIONS TO PREDICT WESTERN AUSTRALIAN SEASONAL RAINFALL USING ARIMAX MODEL," *Hydrology*, vol. 7, no. 3, 2020, doi: https://doi.org/10.3390/hydrology7030052.