JURNAL STATISTIKA DAN APLIKASINYA

e-ISSN 2620 – 8369
VOLUME 9 (ISSUE 2) DECEMBER 2025: 82-92
https://doi.org/10.21009/JSA.09208

FORECASTING THE PRICE OF CURLY RED CHILI PEPPERS IN EAST JAVA PROVINCE USING ARIMA MODEL WITH ITERATIVE OUTLIER DETECTION PROCEDURE

Fareka Erdien[1], Widyanti Rahayu[2], Bagus Sumargo[3*], Ika Yuni Wulansari[4], Didiq Rosadi Ali[5]

[1,2,3]Department of Statistics, Faculty of Mathematics and Natural Sciences, Jakarta State University
Jl. Rawamangun Muka., East Jakarta 13320, Indonesia
[4]School of Mathematical and Physical Sciences, University of Technology Sydney
15 Broadway, Ultimo, NSW 2007, Australia.
[5]Global Development Department, University of Copenhagen
Øster Farimagsgade 5, Copenhagen, 1353, Denmark.

Corresponding author's e-mail: ***bagussumargo@unj.ac.id***

ABSTRACT

Article History:
Received: 27, November 2025
Revised: 15, December 2025
Accepted: 29, December 2025
Published: 31, December 2025
Available online.

Keywords:
Time Series; Forecasting; Detection Outlier; Chili Price

Curly red chili is one of the vegetables with high economic value because it plays a role in supporting the food industry and meeting domestic needs. Fluctuations in the price of curly red chili peppers can change at any time, requiring forecasting to prevent losses for economic actors. This research aims to get the best model for forecasting and determine the accuracy of forecasting the price of curly red chili. The Autoregressive Integrated Moving Average (ARIMA) model is one method that can be used for forecasting with limitations requiring data that must be stationary. Outliers in the ARIMA model affect the autocorrelation structure of a time series so that the estimated values of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) become biased so that forecasting with the ARIMA model is less accurate and requires handling outliers in the form of outlier detection, one of which is an iterative procedure. From this study, it was found that the ARIMA(0,2,3) model with outlier detection was the best model for forecasting. Forecasting tends to show a downward trend with an accuracy level of MAPE value of 4.612, which means that the model is very good for forecasting.

How to cite this article:
F. Erdien, W. Rahayu, B. Sumargo, I. Y. Wulansari, D. R Ali, "FORECASTING THE PRICE OF CURLY RED CHILI PEPPERS IN EAST JAVA PROVINCE USING ARIMA MODEL WITH ITERATIVE OUTLIER DETECTION PROCEDURE", Jurnal Statistika dan Aplikasinya, vol. 9, iss. 2, pp. 82 – 92, December 2025

Journal e-mail: *jsa@unj.ac.id*
Research Article · Open Access

1. INTRODUCTION

The agricultural production sector has an important role in meeting economic needs in Indonesia. Vegetable agricultural products that have a high economic value because of their significant role in meeting domestic needs and supporting the food industry, one of which is chili. According to Pickersgill (1997), among the five species that have economic potential are Red Chili (Capsicum Annum) and Cayenne Pepper (Capsicum Frutescens) [1].

According to the Central Statistics Agency (BPS) (2023), chili production in Indonesia in 2022 touched 3,020,262 tons [2]. This figure is an increase from 2021 which only touched 2,747,018 tons. Indonesia has a good market potential in world trade by looking at the large chili production [3]. One of the provinces that carries out the largest red chili production activities is East Java Province [4]. Chili production in East Java still greatly affects the supply of chili prices, especially cayenne pepper and curly red chili. The development of the retail price of curly red chili shows a fluctuating pattern from day to day. Fluctuations in the price of curly red chili peppers that can change at any time require forecasting to prevent losses for economic actors.

Forecasting is the prediction of an event or several events in the future. Most forecasting problems involve the utilization of time series data, which is a series of observations on a variable that is sequential in time or chronologically [5]. One method used for forecasting is the Box-Jenkins method with the Autoregressive Integrated Moving Average (ARIMA) model, which was developed by George E. P. Box and Gwilym M. Jenkins in 1970 [6]. The use of ARIMA models in the presence of outliers can affect the autocorrelation structure of a time series so that the estimated values of the Autocorrelation Function (ACF) and partial Autocorrelation Function (PACF) become biased [7]. The impact is that forecasting with the ARIMA model produces less accurate forecasts where the model is unable to provide forecasts that are close to the true value, so it is important to use procedures that will detect and correct outliers.

Outlier detection was first studied by Fox (1972) by introducing two types of outliers: additive outliers (AO) and innovational outliers (IO). Additive outliers are outliers that have an effect on time series data for one period only, while innovational outliers are outliers whose effect follows an ARMA process. One method for outlier detection is to use an iterative procedure, which was first developed by Chang, Tiao, and Chen (1983). The iterative procedure is used to perform detection and handling when the time of outliers is unknown. One of the studies using iterative procedures for outlier detection is Putri and Suhartono's (2015) research which concluded that the outlier detection method is able to overcome outliers and can be applied to the ARIMA model of short-term electricity data [8]. In addition, outlier detection has been widely used in previous studies, including those conducted by Ahmar (2018) concluded that the SARIMA model with outlier detection on palm oil production data is the best model for forecasting palm oil production at PTPN XIII for the period January 2010 to December 2017 [9].

Based on the background description above, curly red chili is one of the important agricultural products in fulfilling Indonesia's economic needs. Until now, there has been no research that examines how to predict the price of curly red chili in the future, especially in East Java Province by considering outlier detection in the data. In this study, forecasting and outlier detection are carried out on curly red chili price data to be able to provide insight to economic actors so that they can design appropriate strategies to overcome changes in the price of curly red chili in the future.

2. METHODS

Material and Data

This study uses daily data on the price of curly red chili peppers in East Java Province with the period October 1, 2023 to April 13, 2024 2024 obtained from the site https://panelharga.badanpangan.go.id/. The amount of data used is 196 as a sample. The variables used in this study are daily data on the price of curly red chili peppers for the period October 1, 2023 to April 13, 2024 (Z_t). The time series data for that period was chosen because the price of curly red chili in East Java Province slowly increases and tends to fluctuate.

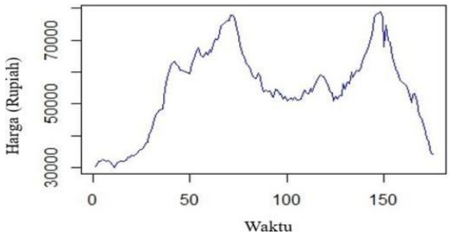

Figure 1. Daily Curly Red Chili Price During October 1, 2023 - April 13, 2024.

Research Method

The analysis procedure of curly red chili price data in this study was carried out with the following steps: (1) Exploring data using time series plots on curly red chili price data. (2) Categorize the data into two parts, namely training data and test data. (3) Modeling data with ARIMA models using training data. (4) Perform outlier detection on the selected ARIMA model to determine the type of outliers contained in the data. (5) Comparing the AIC, BIC, and MSE values between the ARIMA model with outlier detection and the ARIMA model without outlier detection to determine the best model. (6) Forecasting the price of curly red chili using the best model. See the accuracy of the best model forecasting results by calculating the MAPE value.

Time Series Analysis

Time series analysis aims to obtain a model of the variable under study and perform forecasting for future values using past data [6]. In time series analysis, the stationarity assumption is divided into two, including stationarity of variance and stationarity of average [5]. Data that are stationary with respect to variance show that the values of the variables have a fixed and constant movement over time. Meanwhile, data that is stationary with respect to the mean shows that the values of the variables move steadily around a constant mean [10]. One way to check stationarity is to conduct a unit root test through the Augmented Dickey Fuller (ADF) test. One way to check stationarity against variance is by looking at the Box-Cox plot. Data that are not stationary in terms of variance can be overcome by Box-Cox transformation.

Autoregressive Integrated Moving Average (ARIMA)

The Autoregressive Integrated Moving Average (ARIMA) model is a model formed from the Autoregressive (AR) model and the Moving Average (MA) model with Integrated (I) which shows that the data used has been differenced so that the data is stationary [10]. The ARIMA model is a model developed from the AR, MA, and ARMA models but is used for time series data that is not stationary so it is necessary to do differencing in order to achieve stationarity. The ARIMA model can be denoted by ARIMA(p,d,q) with d representing differencing, in general the ARIMA model can be written as follows [11]:

$$\phi_p(B)(1-B)^d Z_t = \theta_0 + \theta_q(B)a_t \quad (1)$$

where ϕ_p is the autoregressive coefficient, Z_t is the value of the t-th time series, θ_q is the moving average coefficient, $(1 - B)^d$ represents differencing process at d-th order, p is the autoregressive order, q is the moving average order, and a_t is the residuals at time t.

Model Identification

Data that have not been stationary on average need to be differenced several times so that the data is stationary. Meanwhile, data that has not been stationary to the variance can be overcome by Box-Cox transformation. Data that have been stationary in both average and variance can be made ACF and PACF graph plots to identify the ARIMA model. Table 1 shows model identification with ACF and PACF graph plots as follows [11]:

Table 1. Identification of ARIMA models with ACF and PACF plots

	ACF	PACF
AR(p)	Descends exponentially or tails off	Cut off after p-th lag
MA(q)	Cut off after q-th lag	Descends exponentially or tails off
AR(p) or MA(q)	Cut off after q-th lag	Cut off after p-th lag
ARMA(p, q)	Descends exponentially or tails off	Descends exponentially or tails off

Parameter Estimation with Maximum Likelihood

The MLE method uses a probability density function where the joint probability density function of an n observation $(x_1, x_2, \dots, x_n)$ is a random sample from an identical population that exhibits a probability density function that depends on the parameters θ can be denoted as $f(x|\theta)$ and can be written as follows [11]:

$$f(x_1, x_2, \dots, x_n|\theta) = \prod_{t=1}^{n} f(x_i|\theta) = L(\theta|x_1, x_2, \dots, x_n) \tag{2}$$

where $(x_1, x_2, \dots, x_n|\theta)$ is the probability density function $f(x|\theta)$, L is the likelihood function, and θ is the unknown parameter. With the maximum likelihood estimator for the parameter θ can be written as follows:

$$lnL = l(\theta) = ln \prod_{t=1}^{n} f(x_i|\theta) = \sum_{i-1}^{n} ln\ f(x_i|\theta) \tag{3}$$

Joint probability density function of residuals $a = (a_1, a_2, \dots, a_n)'$ and a_t is white noise with $N(0, \sigma^2)$ can be written as follows:

$$\begin{aligned} f(a_t) &= \left(2\pi\sigma_a^2\right)^{-1\backslash 2} \exp\left[-\frac{1}{2}\left(\frac{a_t - 0}{\sigma_a}\right)^2\right] \\ &= \left(2\pi\sigma_a^2\right)^{-1\backslash 2} \exp\left[-\frac{1}{2}\left(\frac{a_t}{\sigma_a}\right)^2\right] \end{aligned} \tag{4}$$

with Equation (2), Equation (3), and Equation (4), the joint probability distribution for $a_1, a_2, \dots, a_n$ can be written as follows [6]:

$$\begin{aligned} f(a_1, a_2, \dots, a_n) &= \prod_{t=1}^{n} \left(2\pi\sigma_a^2\right)^{-1\backslash 2} \exp\left[-\frac{1}{2}\left(\frac{a_t}{\sigma_a}\right)^2\right] \\ &= \left(2\pi\sigma_a^2\right)^{-1\backslash 2} \exp\left[-\frac{1}{2}\left(\frac{a_t}{\sigma_a}\right)^2\right] \end{aligned} \tag{5}$$

Model Diagnosis, Best Model Selection, and Forecasting Accuracy Test

Diagnosis of the ARIMA model is used to assess whether the ARIMA model formed is suitable for use or not by checking whether the residual assumptions in the model are met [11]. Test assumptions that must be met, namely the assumption of white noise residuals and the assumption of normally distributed residuals. The model is said to be feasible if the residuals $\{a_t\}$ is able to fulfill the white noise residual assumption test with mutually independent residuals where the residuals of the formed model are not correlated with each other [5].

The white noise assumption can be verified through residual ACF and PACF plots, while the Kolmogorov-Smirnov test is used to assess normality in residuals. These tests ensure the validity of the model's residuals. The best model is determined by the lowest Mean Square Error (MSE), which indicates better forecasting accuracy. Alternatively, models are compared using Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC), with preference given to the model with the smallest values.

One method that can be used to measure the level of forecasting accuracy is to look at the Mean Absolute Percentage Error (MAPE) value. The MAPE value is displayed in percentage form with the criteria that the smaller the MAPE percentage value, the better the forecasting results of the model. A good model has a MAPE value that meets the criteria in Table 2.

Table 2. MAPE Criteria

MAPE (%)	Accuracy Level
$MAPE < 10$	Very good
$10 \leq MAPE \leq 19$	Good
$20 \leq MAPE \leq 49$	Worth
$MAPE \geq 50$	Inaccurate

Iterative Outlier Detection Procedure

Outliers are observations that are unusual and significantly different from the rest of the data. Outliers can be caused by errors in measurement or data copying, or they can be due to sudden and temporary changes in the process being observed [6]. There are four types of outliers, namely innovational outliers (IO), additive outliers (AO), level shift (LS), and temporary change (TC). The four types of outliers can be seen in Figure 2 as follows [16]:

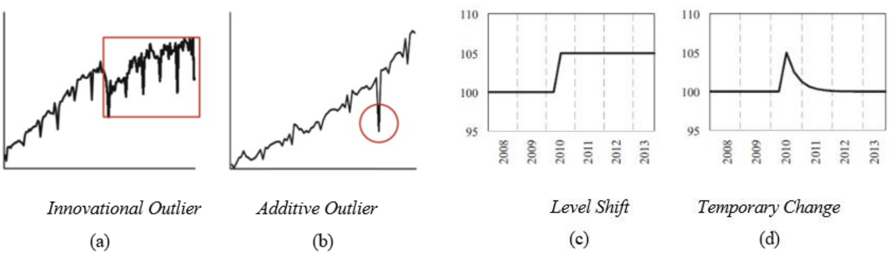

Figure 2. Outlier Model Graph Pattern Type

1. Additive Outlier (AO)
   Additive outliers (AO) are outliers that have an effect on time series data for one period only. AO only affects the observation t, Z_t. The AO model can be written as follows [11]:

$$Z_t = \begin{cases} X_t, & t \neq T \end{cases}$$

$$= \frac{\theta(B)}{\phi(B)} a_t + \omega I_t^{(T)} \quad (6)$$

where

$$I_t^{(T)} = \begin{cases} 1, & t \neq T, \\ 0, & t = T, \end{cases} \quad (7)$$

2. Innovational Outlier (IO)
Innovational outliers (IO) are outliers whose effects follow an ARMA process. Innovational outliers affect all observations $Z_T, Z_{T+1}, \ldots,$ beyond time T through the memory system described by $\frac{\theta(B)}{\phi(B)}$. The innovational outlier (IO) model can be written as follows [11]:

$$Z_t = X_t + \frac{\theta(B)}{\phi(B)} \omega I_t^{(T)}$$
$$= \frac{\theta(B)}{\phi(B)} \left(a_t + \omega I_t^{(T)} \right) \quad (8)$$

where Z_t is the value of the t-th time series, X_t is the value of a normal time series followed by data in the absence of outliers, $I_t^{(T)}$ is the indicator variable that represents the presence or absence of an outlier at time T, ω is the magnitude of the impact of the outlier that occurs at time T, $\theta(B)$ is the moving average (MA) operator of ARMA model, $\phi(B)$ is the autoregressive (AR) operator of the ARMA model, and a_t represents white noise at time t.

Outlier detection was first studied by Fox (1972) [11]. Suppose Z_t is the observed series and X_t is an outlier-free series. In general, time series data can contain several outliers of different types. The time series model to account for outliers can generally be written as follows [11]:

$$Z_t = \sum_{j=1}^{k} \omega_j v_j(B) I_t^{(T_j)} + X_t \quad (9)$$

where ω_j is the impact parameter of the outlier j, $v(B)$ is the polynomial lag operator describing the effect of outliers, X_t is the value of a normal time series followed by data without outliers, and $I_t^{(T)}$ represents the indicator function that takes value 1 if $t = T_j$ and 0 if $t \neq T_j$.

An iterative procedure was first proposed by Chang, Tiao, and Chen (1983) to detect and handle when the timing of outliers is unknown. If T is unknown, but the time series parameters are known, it can be calculated λ_1, for each $t = 1, 2, \ldots, n$, and a decision can be made based on the results. In reality, it is often found that the time series parameters ϕ_j, θ_j, π_j, and σ_2 are usually unknown and must be estimated. Parameter estimation will be biased by the presence of outliers, so outlier detection can be done with the following iterative procedure [5]. Then, the test statistic for AO can be written as follows:

$$\hat{\lambda}_{A,T} = \frac{\tau \hat{\omega}_{AT}}{\hat{\sigma}_a} \quad (10)$$

with

$$\hat{\omega}_{AT} = \frac{e_T - \sum_{j=1}^{n-T} \pi_j e_{T+j}}{\sum_{j=1}^{n-T} \pi_j^2} \quad (11)$$

$$\tau^2 = \sum_{j=1}^{n-T} \pi_j^2 \quad (12)$$

If $\hat{\lambda}_T = |\hat{\lambda}_{A,T}| > c$, then there is an AO detected at the Tth time. Where c is a predetermined constant value with a typical value of 3.0, 3.5, or 4.0 [5]. The effect of AO can be eliminated through residuals which can be written as follows:

$$\tilde{e}_t = \hat{e}_t - \hat{\omega}_{AT} \hat{\pi}(B) I_t^{(T)} = \hat{e}_t - \hat{\omega}_{AT} \hat{\pi}_{t-T}, t \geq T \quad (13)$$

Meanwhile, the test statistic for IO can be written as follows:

$$\hat{\boldsymbol{\lambda}}_{I,T} = \frac{\hat{\boldsymbol{\omega}}_{IT}}{\hat{\boldsymbol{\sigma}}_{a}} \tag{14}$$

with

$$\hat{\boldsymbol{\omega}}_{IT} = \boldsymbol{e}_{T} \tag{15}$$

If $\hat{\lambda}_T = |\hat{\lambda}_{I,T}| > c$, then there is an IO detected at the T-th time. The effect of IO can be eliminated through residuals which can be written as follows:

$$\tilde{\boldsymbol{e}}_t = \hat{\boldsymbol{e}}_t - \hat{\boldsymbol{\omega}}_{IT} = \boldsymbol{0} \tag{16}$$

If outliers are identified, then the residuals and modified variance estimates can be calculated using the same parameters $\hat{\pi}(B) = \frac{\hat{\phi}(B)}{\hat{\theta}(B)}$ to calculate the statistics $\hat{\lambda}_{A,T}$ and $\hat{\lambda}_{I,T}$. Each section is repeated until all outliers are identified. Outlier model for AO when the outlier identification procedure occurs at multiple points in time, e.g. $T_1, T_2, \dots, T_k$ which can be used for estimation on the order of Z_t can be written as follows:

$$\boldsymbol{Z}_t = \sum_{j=1}^{k} \boldsymbol{\omega}_j \boldsymbol{I}_t^{(T_j)} + \frac{\boldsymbol{\theta}(\boldsymbol{B})}{\boldsymbol{\phi}(\boldsymbol{B})} \boldsymbol{a}_t \tag{17}$$

Meanwhile, the outlier model for IO when the outlier identification procedure occurs at multiple points in time, for example $T_1, T_2, \dots, T_k$ can be written as follows:

$$\begin{aligned} \boldsymbol{Z}_t &= \sum_{j=1}^{k} \boldsymbol{\omega}_j \frac{\boldsymbol{\theta}(\boldsymbol{B})}{\boldsymbol{\phi}(\boldsymbol{B})} \boldsymbol{I}_t^{(T_j)} + \frac{\boldsymbol{\theta}(\boldsymbol{B})}{\boldsymbol{\phi}(\boldsymbol{B})} \boldsymbol{a}_t \\ &= \frac{\boldsymbol{\theta}(\boldsymbol{B})}{\boldsymbol{\phi}(\boldsymbol{B})} \left(\sum_{j=1}^{k} \boldsymbol{\omega}_j \boldsymbol{I}_t^{(T_j)} + \boldsymbol{a}_t \right) \end{aligned} \tag{18}$$

A modified time series observation containing outlier effects where the outlier effects are removed can be written with the following equation:

$$\hat{\boldsymbol{Z}}_t = \boldsymbol{Z}_t + \sum_{j=1}^{k} \hat{\boldsymbol{\omega}}_j \boldsymbol{I}_t^{(T_j)} \tag{19}$$

3. RESULTS

The training dataset used consists of $n = 155$ observations. The Augmented Dickey–Fuller test produces a p-value of 0.541 > 0.05, indicating that the data are not stationary in the mean. The Box–Cox plot shows a lambda value of 1.591 > 1, indicating that the data are stationary in variance. After second-order differencing (d = 2), the ADF test yields a p-value of 0.040 < 0.05, demonstrating stationarity. Temporary ARIMA model estimation shows that only the ARIMA(0,2,1), ARIMA(0,2,2), ARIMA(0,2,3), ARIMA(1,2,0), ARIMA(3,2,0), and ARIMA(1,2,1) models have significant parameters ($p < 0.05$). Based on minimum MSE, AIC, and BIC values, the ARIMA(0,2,3) model is selected as the best model.

Table 3. Residual Assumption Test of Selected ARIMA Model

Residual Assumption Test	*p-value*
Ljung-Box	0,942
Kolmogorov Smirnov	0,150

Residual diagnostic tests are presented in Table 3. The Ljung–Box test yields a p-value of 0.942 > 0.05, and the Kolmogorov–Smirnov test yields a p-value of 0.150 > 0.05, indicating that the residuals meet the assumptions of white noise and normality. The corresponding ARIMA(0,2,3) model equation is:

$$Z_t = 2Z_{t-1} - Z_{t-2} + a_t - 0{,}939a_{t-1} + 0{,}326a_{t-2} - 0{,}252a_{t-3} \quad (20)$$

Outlier detection is used to determine the presence of outliers in the data and can be seen in Figure 3.

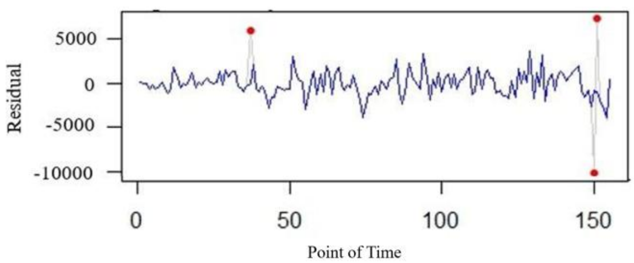

Figure 3. Plot of Outlier Detection Results

There are three outliers in the curly red chili price data in East Java Province from October 1 to March 3, 2024, namely in the 37th, 150th, and 151st data. To find out the type of outliers, it can be seen in Table 4.

Table 4. Outlier Detection of Selected ARIMA Models

Outlier **Type**	**Point**	**Residual Value**	**T-stat value**
AO	37	5982	4,701
IO	150	-9401	-6,892
AO	151	8475	6,261

The ARIMA(0,2,3) model equation with Additive Outlier (AO) on the 37th data can be written as follows:

$$\boldsymbol{Z_t = 2Z_{t-1} - Z_{t-2} + a_t - 0,930a_{t-1} + 0,316a_{t-2} - 0,263a_{t-3} + 0,230I^{(T_{37})}} \quad (21)$$

Meanwhile, the equation of the ARIMA(0,2,3) model with Innovational Outlier (AO) on the 150th data can be written as follows:

$$\boldsymbol{Z_t = 2Z_{t-1} - Z_{t-2} + a_t - 0,740a_{t-1} + 0,280a_{t-2} - 0,338a_{t-3} + 0,848I^{(T_{150})}} \quad (22)$$

The equation of the ARIMA(0,2,3) model with Additive Outlier (AO) on the 151st data:

$$\boldsymbol{Z_t = 2Z_{t-1} - Z_{t-2} + a_t - 0,765a_{t-1} + 0,048a_{t-2} - 0,191a_{t-3} + 0,634I^{(T_{151})}} \quad (23)$$

Equations (21), (22), and (23) represent the initial estimation of the model for each outlier type identified during the iterative detection process. Then, the addition of outlier types one by one into the ARIMA(0,2,3) model so that the results can be obtained in Table 5 [12].

Table 5. AIC Comparison of Models with Outliers

ARIMA(0,2,3) with Outlier Model Addition	AIC
151	2708,37
151, 37	2709,44
151, 37, 150	2659,42

Table 5 indicates that the ARIMA model with the best outlier detection is IMA (2,3) with the addition of three types of outliers.

Table 6. Comparison of AIC of ARIMA Model with Outlier Detection and Without Outlier Detection

Model	MSE	AIC	BIC
ARIMA(0,2,3)	2941,74	2724,45	2736,58
ARIMA(0,2,3) with *outlier* detection	1854,50	2659,42	2680,63

Table 6 indicates that the MSE, AIC, and BIC values of the ARIMA model with outlier detection are smaller than the MSE, AIC, and BIC values of the ARIMA model without outlier detection. Thus, the best model chosen is the ARIMA model with outlier detection so that the model can be used for forecasting. After the iterative process is completed, the parameters are re-estimated to produce the final combined model as shown in Equation (24). The best model equation or ARIMA(0,2,3) model with outlier detection can be written as follows:

$$\boldsymbol{Z_t = 2Z_{t-1} - Z_{t-2} + a_t - 0,716a_{t-1} + 0,243a_{t-2} - 0,312a_{t-3} + 0,169I^{(T_{37})} - 0,134I^{(T_{150})} + 0,790I^{(T_{151})}} \quad (24)$$

The forecasting plot in Figure 4 shows a decreasing pattern with price values ranging from 15,413.29 rupiah to 61,452.82 rupiah with an average of 38,474.76 rupiah.

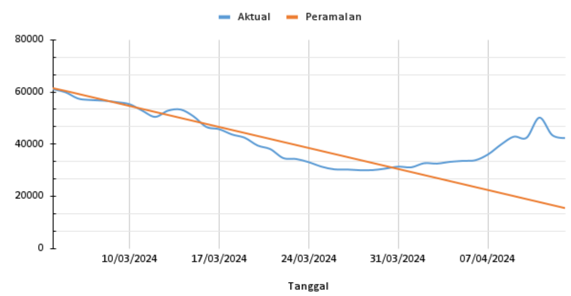

Figure 4. Forecasting with the Best Model

The smaller the MAPE value, the better the model for forecasting. Table 7 indicates that the ARIMA(0,2,3) model with outlier detection has a MAPE value of 4.612 < 10, so the model is very good for forecasting according to Lewis (1985) [13].

Table 7. **Forecasting Accuracy Level of the Best Model**

Model	MAPE
ARIMA(0,2,3) with outlier detection	4,612

4. DISCUSSIONS

The ADF test results indicate that the original training data contain a unit root, requiring differencing to achieve stationarity. The Box–Cox lambda value greater than 1 suggests that variance transformation is unnecessary. After second-order differencing, the stationarity requirement is satisfied, allowing ARIMA modeling. The evaluation of temporary ARIMA models shows that only six models have significant parameters, and among them, the IMA (2,3) model performs best based on MSE and information criteria. This highlights that higher-order integrated moving average structures better capture the behavior of red chili price fluctuations. The residual diagnostic tests confirm that the selected IMA (2,3) model satisfies both white-noise and normality assumptions, ensuring that the model is statistically reliable and adequately fits the data.

Outlier detection reveals three extreme observations. These outliers significantly influence the model, as evidenced by the improvement in AIC, MSE, and BIC after including their effects. This suggests that ignoring outliers may reduce model accuracy, and their incorporation leads to a more robust ARIMA specification. The final model incorporating all outliers demonstrates substantially better performance, indicating that unexpected shocks in chili prices—possibly due to supply disruptions or seasonal events—must be accounted for to improve forecasting precision. The forecasting results show a decreasing trend in predicted prices with reasonable variability. The MAPE value of 4.612 categorizes the model as “very accurate” according to Lewis (1985) [13], confirming that the ARIMA (0,2,3) model with outlier adjustments is suitable for predicting red chili prices in East Java Province.

The practical interpretation of the identified outliers reveals a strong correlation with market conditions in Indonesia. The Additive Outlier (AO) at point 37 (November 6, 2023) corresponds to the peak of the El Niño-induced drought, which caused a significant supply shock in major chili-producing regions. This led to a temporary but sharp price spike as harvests failed.

Furthermore, the Innovational Outlier (IO) at point 150 (February 27, 2024) followed by an AO at point 151 (February 28, 2024) highlights extreme volatility during the transition to the Ramadan season. The negative IO suggests systemic shock, possibly due to large-scale market interventions or sudden harvest inflows, while the immediate positive AO correction reflects the high market sensitivity and speculative pressure typical of the days leading up to the holy month.

5. CONCLUSION

Using R Studio and ARIMA(p,d,q) models with outlier detection, the analysis identified the optimal model as ARIMA(0,2,3). The model demonstrates a downward trend in forecasts, with predicted values ranging from 15,413.39 IDR to 61,452.82 IDR. The model's accuracy, reflected by a MAPE of 4.612%, indicates excellent forecasting performance. This study examines the forecasting of curly red chili prices in East Java Province for the period from October 1, 2023, to April 13, 2024, revealing a general downward trend. Factors such as rainfall, population size, and inflation may influence price fluctuations and warrant further consideration. Additionally, exploring alternative outlier detection methods could improve the identification of outlier values.

6. ACKNOWLEDGMENTS

The authors would like to express their gratitude to the Badan Pangan Nasional for their invaluable insights. We also thank all individuals and institutions for their technical and financial support, which were instrumental in the successful completion of this research

7. REFERENCES

[1] B. Pickersgill, “Genetic resources and breeding of Capsicum spp.,” *Euphytica*, vol. 96, pp. 129–133, 1997.

[2] Badan Pusat Statistik, “Produksi beberapa tanaman sayuran di Indonesia (ton),” 2023. [Online]. Available: https://www.bps.go.id/id/statistics-table/2/NjEjMg==/produksi-tanaman-sayuran.html [Accessed: 28 Feb. 2024].

[3] Pusat Data dan Sistem Informasi Pertanian, Kementerian Pertanian, *Analisis Kinerja Perdagangan Cabai Merah*, vol. 13, no. 1E, 2023.

[4] H. Fu, “Lima kabupaten yang menjadi sentra produksi cabai merah terbesar di Jawa Timur,” *Berita Daerah*, 2021. [Online]. Available: https://www.beritadaerah.co.id [Accessed: 28 Feb. 2024].

[5] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. New Jersey: John Wiley & Sons, 2015.

[6] K. S. Chan and J. D. Cryer, *Time Series Analysis with Applications in R. New York: Springer*, 2008.

[7] D. R. Arifanti and A. Asrirawan, “Peramalan volume debit air Kota Palopo menggunakan model ARIMA deteksi pencilan,” *Infin. J. Mat. dan Apl.*, vol. 1, no. 2, pp. 31–39, 2021.

[8] I. K. Putri and S. Suhartono, “Deteksi outlier pada model ARIMA musiman ganda untuk peramalan beban listrik jangka pendek di Jawa Timur,” *Jurnal Sains dan Seni ITS*, vol. 4, no. 1, pp. D31–D36, 2015.

[9] A. S. Ahmar, S. Guritno, A. Rahman, I. Minggi, M. A. Tiro, M. K. Aidid, … A. N. M. Arifin, “Modeling data containing outliers using ARIMA additive outlier (ARIMA-AO),” *Journal of Physics: Conference Series*, vol. 954, no. 1, pp. 012010, 2018.

[10] S. Makridakis, S. C. Wheelwright, and R. J. Hyndman, *Forecasting Methods and Applications*, 3rd ed. New York: John Wiley & Sons, 2008.

[11] W. W. S. Wei, *Time Series Analysis: Univariate and Multivariate Methods*, 2nd ed. Boston: Pearson Education, Inc., 2006.

[12] N. S. R. Ahmad, S. Martha, and N. Imro’ah, “Prediksi produksi kelapa sawit di PTPN XIII dengan additive outlier pada model seasonal autoregressive integrated moving average (SARIMA),” *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya*, vol. 8, no. 4, 2019.

[13] C. D. Lewis, *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*. London: Butterworths, 1982.