Jurnal CoreIT, Vol.11, No.1, Juni 2025
ISSN 2460-738X (Print)
ISSN 2599-3321 (Online)

Red Curly Chili Forecast in Southeast Sulawesi Using Auto Regressive Integrated Moving Average (ARIMA)

Muhammad Vannes Al Qadri[1], Rizal Adi Saputra[2*], Bambang Pramono[3]
[123] Program Studi Teknik Informatika, Universitas Halu Oleo, Indonesia
[1]muhammadvannesalqadri@gmail.com, [2]rizaladisaputra@uho.ac.id (**corresponding author*), [3]bambangpramono@uho.ac.id

Abstract. Price is a crucial aspect in the world of trade. Red curly chili peppers have become one of the plants favored by many consumers. This research aims to develop a forecasting model that can provide a more accurate insight into the future prices of red chili peppers, particularly in Southeast Sulawesi. Because price forecasting plays a crucial role in predicting future price trends, the Auto Regressive Integrated Moving Average (ARIMA) method becomes one of the models that can be used for time series analysis. The data for this research is sourced from the National Food Body Price Panel Website. The data period starts from August 8, 2022, to December 15, 2023, with the last 500 days' prices used as both test and training data. In this study, the ARIMA (1,1,1) model emerged as the best among the three ARIMA models analyzed. The ARIMA (1,1,1) model yielded a MAPE percentage of 17.97%, indicating that this model is suitable or reliable for time series forecasting. Furthermore, the results of this experiment show that the forecasted prices for the next 10 days do not experience significant decreases or increases, referring to several recent data points used as training data samples.

Keywords: ARIMA, Data, Forecast, Price, Red Curly Chili

Received March 2025 / **Revised** May 2025 / **Accepted** June 2025

INTRODUCTION

The data released by the Food and Agriculture Organization (FAO) of the United Nations (UN), published by *The Agriculture News*, states that Indonesia ranked fourth as the largest chili producer in the world in 2019. Total production reached 2.5 million tons, with three main chili variations: bird's eye chili, curly red chili, and large red chili [1]. In Indonesia, approaching Eid and New Year celebrations, the prices of red chili and bird's eye chili tend to rise due to high demand. Although only used as a kitchen spice, the increase in chili prices has the potential to cause inflation and reduce people's purchasing power [2].

Currently, curly red chili is often used as an additional ingredient in various dishes worldwide. This is because curly red chili offers diverse benefits, both in culinary applications and health [3]. In recent years, the price fluctuation of curly red chili has become a major concern for agricultural industry players. This fluctuation is usually influenced by factors such as climate variations, agricultural policies, and changes in market demand. Therefore, price forecasting is crucial to assist farmers and traders in production planning and business decision-making [4].

Previous studies on curly red chili price forecasting using causal forecasting methods have been conducted by Ainayyah Fatihah [5], Nurul Hidayati et al. [6], who forecasted chili prices to maintain inflation stability in Banda Aceh City, and Wan Habibi, who analyzed the price cointegration of curly red chili in Pekanbaru [7]. Additionally, research by Waskito, H. et al. investigated the response of growth and yield of curly red chili (Capsicum annuum L.) Ck5 due to the application of NPK fertilizer and biofertilizer [8].

Curly red chili price forecasting can be conducted through statistical methods and artificial intelligence [9]. Based on linear regression analysis in a case study of curly red chili price forecasting, accuracy can be calculated, which is influenced by dependent and independent variables. Meanwhile, with an artificial intelligence approach, a dataset can be trained to develop a system capable of predicting or forecasting curly red chili prices in Southeast Sulawesi. This study will use the Auto Regressive Integrated Moving Average (ARIMA) method as an artificial intelligence approach. The ARIMA method aims to develop a forecasting model that provides a more accurate outlook on future curly red chili prices, particularly in Southeast Sulawesi.

DOI: 10.24014/coreit.v11i1.21075

This study presents several important distinctions compared to previous research on red curly chili price forecasting. First, the method used in this study is the Auto Regressive Integrated Moving Average (ARIMA), a statistical time series approach that has not been widely applied specifically in the context of red curly chili price forecasting in Southeast Sulawesi. Prior studies, such as those conducted by Ainayyah Fatihah and Nurul Hidayati et al., mainly employed causal forecasting approaches or linear regression analysis, while this research adopts the ARIMA method by conducting data stationarity testing using the Augmented Dickey-Fuller (ADF) test, determining parameters through ACF and PACF plots, and evaluating the model using the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Mean Absolute Percentage Error (MAPE).

Second, this study utilizes 500 consecutive days of daily red curly chili price data (from August 8, 2022, to December 15, 2023), sourced from the official Price Panel of the National Food Agency. The dataset is systematically divided into training and testing subsets to ensure a more accurate and valid prediction model. Third, the geographical focus of this study is Southeast Sulawesi Province, which has rarely been the subject of similar research using the ARIMA approach. Finally, this study not only develops a forecasting model but also provides projected price estimates for several upcoming days, along with predicted minimum and maximum ranges. These results can be directly utilized by farmers, traders, and policymakers in planning production and distribution strategies.

METHODS

A. Data Collection

Data is a collection of initial information or facts in the form of symbols, numbers, words, or images obtained through the process of observation or searching from certain sources [10]. Data collection is carried out to obtain the information needed in order to achieve research objectives. The stated objectives are temporary answers to research questions. These answers need to be tested empirically and this is why data collection is needed. The data collected consists of a set of analysis units as research targets

B. Identifying Data Stationarity

In identifying the stationarity of data, this can be done using the Augmented Dickey-Fuller (ADF) test with the null hypothesis being the presence of a unit root, namely ÿ1 = 1. The test statistic formula is [11]:

$$t_{\hat{\delta}} = \frac{\hat{\delta}}{SE(\hat{\delta})} \qquad (1)$$

Information:

$t_{\hat{\delta}}$ = Calculated t-statistic value.

$\hat{\delta}$ = Estimator of population parameters

SE = Standard *error* of the estimator.

If the ADF result gives a p-value greater than the significance level of 0.05, then the null hypothesis (H0) cannot be rejected and the data has a unit root, and is non-stationary. Meanwhile, if the ADF result gives a p-value smaller than the significance level of 0.05, then the null hypothesis (H0) can be rejected and the data does not have a unit root, and is stationary. It should be noted that if the data is non-stationary, then before proceeding to ARIMA modeling, a differentiation test must be carried out first so that the data changes into stationary data.

C. ARIMA Modeling

In adjusting ARIMA modeling with time series based on the order (p,d,q), you must first create an Autocorrelation Function (ACF) plot and a Partial Autocorrelation Function (PACF) plot. The purpose of creating an ACF plot is to visualize the correlation between time series values at different lags and determine the autoregressive order (p) in the ARIMA model. While creating a plot PACF aims to visualize the partial correlation between time series values at different lags and helps determine the order of the moving average (q) in the ARIMA model. While d in the ARIMA model indicates the level of differentiation applied to the time series to make it stationary. The differencing process (d) involves reducing the value of previous observations. If after the first differencing the time series is still not stationary, then the differencing process can be repeated until the desired stationarity is achieved. Furthermore, after having the order (p, d, q), ARIMA modeling can be built.

The Autoregressive (AR) equation is mathematically:

$$Y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \cdots + \beta_k y_{t-k} \qquad (2)$$

Information:

Y_t	= Values in the time series at time parameters t.
$\beta_1, \beta_2, \dots, \beta_k$	= To be estimated.
$y_{t-1}, y_{t-2}, \cdots, y_{t-k}$	= The values of the time series before time *t* are used as input to the model.

The mathematical equation for Moving Average (MA) [12]:

$$Y_t = \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \alpha_3 \varepsilon_{t-3} + \cdots + \alpha_k \varepsilon_{t-k} \qquad (3)$$

Y_t	= Values in the time series at time parameters t.
$\alpha_1, \alpha_2, \dots, \alpha_k$	= To be estimated.
$\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, \varepsilon_{t-k}$	= Error terms as time before *t*.

D. Comparison AIC and BIC Values

Comparing the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values of ARIMA models that have been created helps in determining the model that best fits the data. Both of these criteria provide a measure of the statistical fit of the model to the data, taking into account the complexity of the model. AIC tends to give preference to models that provide a good fit, but with model complexity. The lower the AIC, the better the model is considered. While BIC imposes a greater penalty on model complexity than AIC. In other words, BIC tends to favor a simpler model if the additional complexity is not offset by a significant increase in the model's fit to the data. Comparing the AIC and BIC values for several ARIMA models can help determine the optimal balance between model fit to the data and model complexity. The choice that minimizes both of these criteria can provide better and more accurate prediction results general.

E. Division of Test Data and Training Data

Training data is pre-existing information, based on events that have already occurred. While the test data is information that has been labeled and will be used for calculations according to the method equation and evaluation of classification accuracy. The relationship between training data and test data is very close, where training data is a reference in calculating the probability of test data [13]. All data that is sampled in the study will be divided into two, namely test data and training data [14]. Data division is carried out with training data of 80% and test data of 20%.

F. Diagnostics and Mean Absolute Percentage Error (MAPE)

Diagnosis in the ARIMA model aims to evaluate how well the model meets the assumptions and how reliable the forecast results are. While MAPE is a calculation used to calculate the average percentage of absolute error between the actual value and the predicted value so that the lower the MAPE value, the better the quality of the model's prediction. Here is the formula for finding MAPE [15]:

$$MAPE = \sum(|Actual - Forecast| \,/\, Actual) * 100/n \qquad (4)$$

Information:

MAPE = Percentage absolute error between actual value and predicted value.
Actual = Actual value.
Forecast = Prediction value.
n = Number of periods used for the calculation.

RESULT AND DISCUSSION

A. Data Description

The data used in this study comes from the National Food Agency Price Panel Website . This data is the commodity price data for curly red chilies per day in Southeast Sulawesi for the last 500 days starting from August 8, 2022 to December 15, 2023 with the type of panel data, namely retail traders, as shown in Table 1

Table 1. Red Curly Chili Price Data in Southeast Sulawesi

Date	**Price**
03-08-22	56700
04-08-22	58430
05-08-22	57200
06-08-22	57380
07-08-22	56480
08-08-22	58810
09-08-22	59380
…	…
09-12-23	71200
10-12-23	74100
11-12-23	70180
12-12-23	78290
13-12-23	75870
14-12-23	76100
15-12-23	75410

B. Building an ARIMA Model of Order (p, d, q)

The ADF test of 500 data on "Price of Red Curly Chili in Southeast Sulawesi" shows a value of -1.463706 as the ADF Statistic and 0.551383 as the p-value. This indicates that the data is non-stationary. Therefore, differentiation is needed to change the data to be stationary.

Figure 1. Differentiation Graph, PACF, and ACF (First)

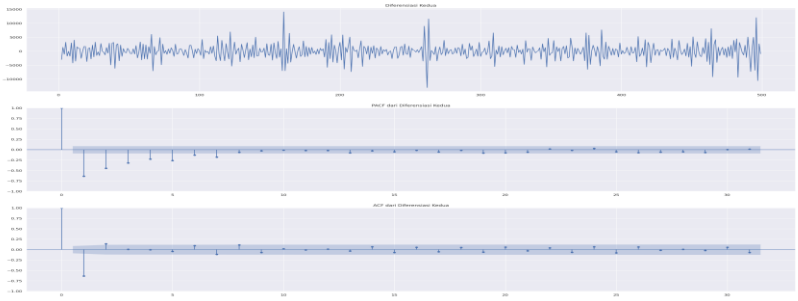

Figure 2. Differentiation Graph, PACF, and ACF (Second)

From Figure 1 and Figure 2, both show that the results of the first and second differentiation have changed the data to stationary because the p-value of both differentiations is below the significance level of 0.05. In this study, differentiation was carried out twice with the aim of finding which ARIMA model is suitable for forecasting "Price of Curly Red Chili in Southeast Sulawesi", whether in the first or second differentiation. The p value is determined from the ACF graph, the d value is determined from the differentiation graph used, and the q value is determined from the PACF graph.

C. Determining the ARIMA Model with AIC and BIC Values

This study selects the ARIMA (1,1,1), ARIMA (1,1,2), and ARIMA (3,2,1) models, then the AIC and BIC values will be used as the basis for determining the ARIMA model in the good category. The following are the results given from the three ARIMA models:

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                   85
Model:                 ARIMA(1, 1, 1)   Log Likelihood                -733.368
Date:                Sat, 06 Jan 2024   AIC                           1472.737
Time:                        06:58:33   BIC                           1480.029
Sample:                             0   HQIC                          1475.668
                                 - 85
Covariance Type:                  opg
==============================================================================
```

Figure 3. Results ARIMA Model (1,1,1)

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                   85
Model:                 ARIMA(1, 1, 2)   Log Likelihood                -733.051
Date:                Sat, 06 Jan 2024   AIC                           1474.103
Time:                        06:58:34   BIC                           1483.826
Sample:                             0   HQIC                          1478.011
                                 - 85
Covariance Type:                  opg
==============================================================================
```

Figure 4. Results ARIMA Model (1,1,2)

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                   85
Model:                 ARIMA(3, 2, 1)   Log Likelihood                -759.995
Date:                Sat, 06 Jan 2024   AIC                           1529.990
Time:                        06:58:33   BIC                           1542.084
Sample:                             0   HQIC                          1534.849
                                 - 85
Covariance Type:                  opg
==============================================================================
```

Figure 5. Results ARIMA Model (3,2,1)

The results of Figure 3, Figure 4, and Figure 5 show that the ARIMA (1,1,1) model is the best model of the three models. The ARIMA (1,1,1) model provides an AIC value of 1472.736794 and a BIC of 1480.029244. The ARIMA (1,1,1) model was chosen because the lower the AIC value, the better the model is considered and the lower the BIC value, the more complex the fit between the model and the data.

D. Dividing Test Data and Training Data

Dividing the data on "Price of Red Curly Chilies in Southeast Sulawesi" into two parts, namely training data and test data. Training data is taken from the first 400 data and test data is taken from the last 100 data.

Table 2. Training Data

Date	**Price**
03-08-22	56700
04-08-22	58430
05-08-22	57200
…	…
04-09-23	41760
05-09-23	40090
06-09-23	41380

Table 3. Test Data

Date	**Price**
07-09-23	41640
08-09-23	42160
09-09-23	41840
…	…
13-12-23	75870
14-12-23	76100
15-12-23	75410

The division of training data and test data in Table 2 and Table 3 can be visualized using ARIMA (1,1,1) model as in Figure 6.

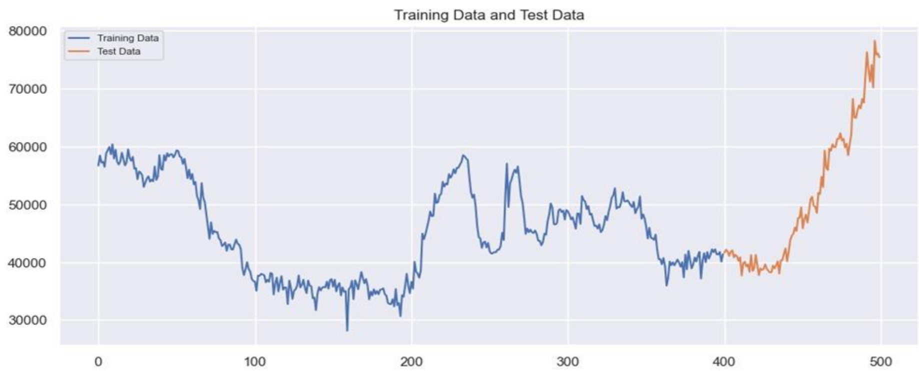

Figure 6. Training Data and Test Data Graph

E. Diagnostic Graphs and MAPE Values

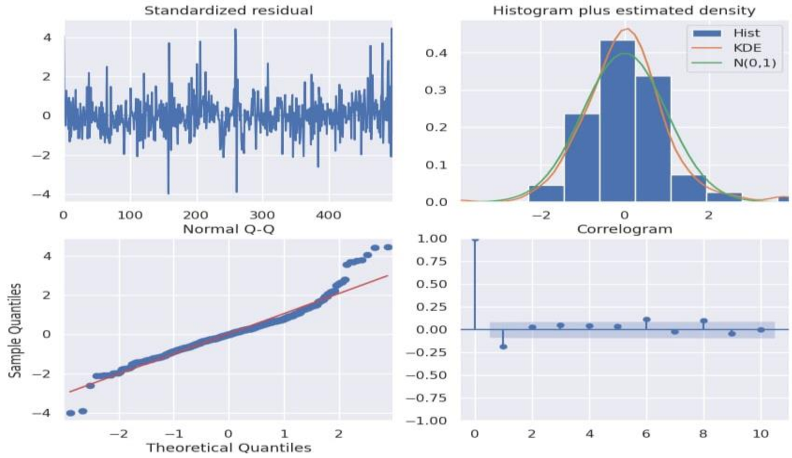

Figure 7. Diagnostic Graph

The MAPE value of the ARIMA (1,1,1) model is 17.97%. This means that the ARIMA (1,1,1) model is suitable or reliable in time series analysis because this model is in good category for forecasting the price of curly red chilies in Southeast Sulawesi.

F. Forecasting Results Using the ARIMA Model (1,1,1)

The following is a graph and table of forecast prices for curly red chilies in Southeast Sulawesi for the next 10 days:

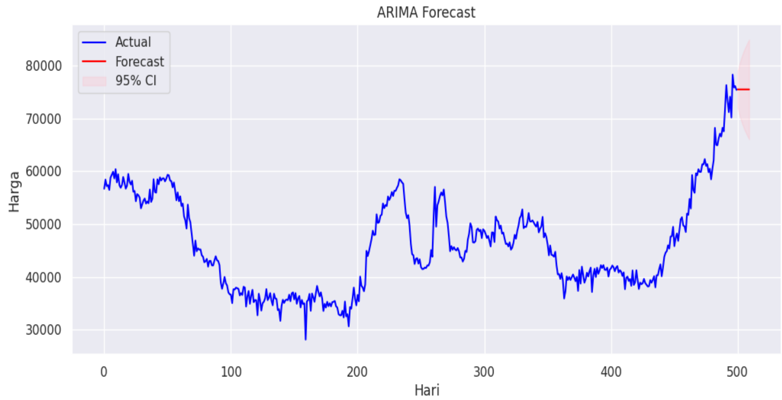

Figure 8. Graph of Forecasting Prices of Red Curly Chilies in Southeast Sulawesi

Table 4. Forecast of Curly Red Chili Prices in Southeast Sulawesi

Date	Price Original	Price Prediction	Minimum Prediction	Maximum Prediction
16-12-23	76260	75496	72144	78848
17-12-23	69930	75495	71043	79946
18-12-23	68430	75495	70163	80827
19-12-23	74800	75495	69408	81581
20-12-23	70730	75495	68737	82252
21-12-23	74910	75495	68127	82862
22-12-23	73100	75495	67564	83425
23-12-23	69740	75495	67038	83951
24-12-23	61900	75495	66543	84446
25-12-23	67210	75495	66074	84915

CONCLUSION

Based on the results of research and discussion on the forecasting of curly red chili prices in Southeast Sulawesi, it can be concluded that the time series analysis process using the ARIMA method with the ARIMA (1,1,1) model provides a MAPE value of 17.97%, which means that the ARIMA (1,1,1) model is good enough to be used in time series forecasting. The results of the ARIMA (1,1,1) model state that the forecast value for the next 10 days does not experience a significant increase or decrease when referring to the last few data that are the training data samples.

REFERENCES

[1] Y. T. Santosa, "10 Negara Produsen Cabai Terbesar Di Dunia," *The Agriculture News.* [Online]. Available: https://theagrinews.com. [Accessed January 4, 2024].

[2] D. Erianto, "Komoditas Cabai: Sejarah, Manfaat, Produksi, Distribusi dan Perkembangan Harga," *Kompas,* April 22, 2022. [Online]. Available: https://www.kompas.id. [Accessed January 4, 2024].

[3] I G. B. T. Ananta and D. G. A. Anjasmara, "Poteksi Ekstrak Buah Cabai Merah Keriting (Capsicum annum var. Longum) Sebagai Antioksidan dan Antibakteri," *Jurnal Ilmiah Medicamento,* vol. 8, pp. 48-55, 2022.

[4] H. Khatimah, W. G. Abdullah, and Abdi, "Analisis Peramalan Produksi dan Harga Cabai Merah (Capsicum annum L.) di Provinsi Sulawesi Tenggara," *Journal of Food System and Agribusinnes (JoFSA),* vol. 7, pp. 113-122, 2023.

[5] A. Fatihah, "Peramalan Produksi Cabai Merah Keriting Menggunakan Metode Causal Forecasting (Studi Kasus pada Pasar Lelang Cabai Sleman)," B.S thesis, Dept. of Industrial Engineering, Islamic University of Indonesia, Yogyakarta, 2022.

[6] N. Hidayati, S. Anwar, and R. Rahmah, "Peramalan Harga Cabai Merah Sebagai Upaya Menjaga Stabilitas Inflasi Kota Banda Aceh," *Agriekonomika,* vol.11, pp. 31-42, 2022.

[7] W. Habibi, "Analisis Kointegrasi Harga Cabai Merah Keriting (Capsicum annuum L.) di Kota Pekanbaru," B.S thesis, Dept. of Agribusiness, Islamic University of Riau, Pekanbaru, 2019.

[8] Waskito H., A. Nuraini, and N. Rostini, "Respon Pertumbuhan dan Hasil Cabai Keriting (Capsicum annuum L.) Ck5 Akibat Perlakuan Pupuk NPK dan Pupuk Hayati," *Jurnal Kultivasi,* vol. 17, pp. 676-681, 2018.

[9] A. C. Fauzan and S. D. Purwanto, "Analisis Perbandingan Algoritma Naïve Bayes Classifier dan Learning Vector Quantization dalam Sistem Identifikasi Boraks pada Bakso Daging Sapi," *Jurnal CoreIT,* vol. 7, pp. 43-50, 2021.

[10] A. M. Sari, "Pengertian Data, Fungsi, Jenis-jenis, dan Manfaatnya," *Fakultas Ilmu Komputer dan Teknologi Informasi,* August 2, 2023. [Online]. Available: https://fikti.ac.id. [Accessed January 7, 2024].

[11] R. Santra, "Tests for Stationarity in Time Series – Dickey Fuller Test & Augmented Dickey Fuller (ADF) Test," *Medium,* May 12, 2023. [Online]. Available: https://medium.com. [Accessed January 5, 2024].

[12] C. Shetty, "Time Series Models," *Medium,* September 22, 2020. [Online]. Available: https://medium.com. [Accessed January 5, 2024].

[13] D. Anggreani, Herman, and W. Astuti, "Kinerja Metode Naïve Bayes dalam Prediksi Lama Studi Mahasiswa Fakultas Ilmu Komputer," *Prosiding Seminar Nasional Ilmu Komputer dan Teknologi Informasi,* vol. 3, pp. 107-111, 2018.

[14] P. H. Zainal, Y. Angraini, and A. Rizki, "Penerapan Metode Generalized Auto-Regressive Conditional Heteroskedasticity untuk Peramalan Harga Minyak Mentah Dunia," *Journal of Statistics,* vol. 12, pp. 12-21, 2023.

[15] M. A. Maricar, "Analisa Perbandingan Nilai Akurasi Moving Average dan Exponential Smoothing untuk Sistem Peramalan Pendapatan pada Perusahaan XYZ," *Jurnal Sistem dan Informatika,* vol. 13, pp. 36-45, 2019.