Vol. 3 No. (1) (June 2024) 19-31

Journal of Data Insights

http://journalnew.unimus.ac.id/index.php/jodi

Forecasting Honda Car Retail Sales Using the Seasonal Autoregressive Integrated Moving Average Method

Lea Angelina[1], Alia Rahma Permata[2], Jesicha Roselina Arsusma[3] , Firochul Masichah[4], Ihsan Fathoni Amri[5*], M. Al Haris[6]
[1,2,3,4,5,6]Universitas Muhammadiyah Semarang, Indonesia

DOI: https://doi.org/10.26714/jodi.v3i1.416

Article Information

Artiicle History:
Submitted 22th July 2024
Revised 23th June 2025
Accepted 27th June 2025

Keywords:
SARIMA; MAPE; Retail sales; Honda Car.

Abstract

This article discusses the forecasting of Honda car retail sales using the Seasonal Autoregressive Integrated Moving Average (SARIMA) method. The study aims to forecast Honda car retail sales for the upcoming year. Various SARIMA models have been tested to determine the best model, and the results show that the SARIMA (1,1,0)(1,1,1)^{12} model provides the lowest Mean Absolute Percentage Error (MAPE) among all tested models, which is 17,74%. Therefore, this model was chosen for forecasting sales over the next 12 months. The forecast results are expected to assist management in making optimal decisions regarding stock and marketing, as well as significantly enhancing operational efficiency and customer satisfaction in the future.

*Corresponding Author:
E-mail: ihsanfathoni@unimus.ac.id

e-ISSN: 2988 - 2109

I. INTRODUCTION

Car retail sales involve a process in which resellers or dealers sell directly to end consumers and are heavily influenced by customer desires. This study utilizes data on the retail selling price value of Honda brand cars from January 2013 to September 2023 obtained from the Association of Indonesian Automotive Industries (GAIKINDO) [1].

The automotive sector plays a crucial role in Indonesia's economy. As a country with the fourth largest population in the world, Indonesia offers a potentially large market for the automotive industry. Honda, one of the major players in the industry, is known as the leading vehicle brand in Indonesia. Honda's car sales reflect the dynamics of the national automotive market and are an important indicator for businesses and the government in making strategic decisions [2].

Analysis of car sales data is necessary to predict future trends and plan effective business strategies. In this case, data from GAIKINDO provides a comprehensive and detailed report on car sales of various brands, including Honda [1]. In 2022, Honda managed to get the third position as a highly desirable car brand in Indonesia [3]. Honda is internationally renowned and dominates the market in real terms. Due to Honda's rapid growth and the loyalty of its users, analyzing its changing trends and user satisfaction would be meaningful.

Effective planning requires accurate estimation to determine desired goals, including for car dealers. The automotive industry requires precise sales estimates to set sales targets [4]. In Indonesia, car sales are increasing after experiencing a decline in 2020 [5]. According to data from GAIKINDO, retail car sales rose 17.4% in 2022, reaching more than one million units compared to 800 thousand units in the previous year. In 2022, retail sales of Honda cars also increased by 34 thousand from the previous year which was only 91,393 units. The automotive industry is a major contributor to the national economy and creates many jobs.

Data from GAIKINDO shows that retail sales of Honda cars experience seasonal fluctuations, with an increase in March and December [6]. Due to the increase in a period of eight months, forecasting methods that consider seasonal patterns, such as SARIMA, are more appropriate for forecasting methods [7]. SARIMA is the development of the Autoregressive Integrated Moving Average (ARIMA) model, with a seasonal component to capture periodic fluctuations in the data, the result of the development of George Box and Gwilym Jenkins [8]. SARIMA can analyze seasonal data patterns such as quarterly, semi-annual, and annual [9]. ARIMA is used with the assumption that the time series data used must be stationary [10]. The SARIMA model is symbolized as ARIMA $(p,d,q)(P,D,Q)^s$, and is selected by comparing the Akaike Information Criterion (AIC) values. The smallest AIC value indicates the best model, with AIC being an estimator of the relative quality of a statistical model for a given set of data [9].

This method calculates the seasonal constituents in the data and is capable of estimating values with advanced accuracy [11]. SARIMA is popularly used for time series analysis due to its ability to handle data with seasonal patterns [9]. SARIMA has the advantage of accommodating trends and seasonal patterns in sales data, as well as allowing the integration of Autoregressive (AR), Differencing (I), and Moving Average (MA) components. With SARIMA, forecasting results are expected to be more accurate and reliable as a basis for decision making [12]. The results of the sales value estimation help manufacturers optimize their potential and strengthen their position even more in the Indonesian motor vehicle industry. Time series data is applicable for the future, considering that data from the past may occur again [13].

II. METHODS

2.1 Data Source

This study utilizes historical data in the form of the number of retail car sales units from PT Honda Prospect Motor, taken from the beginning of 2013 to September 2023, and obtained from the GAIKINDO website. The data used has a seasonal pattern. In this study, the data will be processed using RStudio software.

2.2 SARIMA Method

The SARIMA method is a time series model that includes Seasonal, Autoregressive, Differencing, and Moving Average components. SARIMA uses time series data with a fixed period that shows seasonal patterns including quarterly, monthly, semi-annual, and annual variations [14]. SARIMA is a development of the ARIMA forecasting method. The general notation of SARIMA is (p, d, q)(P, D, Q)^s, where p, d, q are non-seasonal components, while P, D, Q are seasonal components, and s is many seasonal periods. The basic elements of SARIMA include three main components, namely:

a) Autoregressive (AR or p)
   AR is the number of lags used in the model. The Autoregressive (AR) method states that current events are influenced by data from the previous period [15].
b) Integrated (I or d)
   Orde I is the degree of differentiation required to make the time series stationary. It represents the number of differentiation steps required.
c) Moving Average (MA or q)
   MA is the number of lags applied in the model.
   The basic formula of SARIMA is:

$$\phi p(BS)\phi p(B)(1-B)d(1-BS)DZt = \theta p(B)\Theta Q(BS)\alpha t$$

Description:

Zt	=	Observation at time -t
$\phi p(BS)$	=	Seasonal Autoregressive
$\phi p(B)$	=	Non-Seasonal Autoregressive
αt	=	Residual at time-t
$(1-B)d$	=	Non-Seasonal Differencing
$(1-BS)D$	=	Seasonal Differencing
$\theta p(B)$	=	Non-Seasonal Moving Average
$\Theta Q(BS)$	=	Seasonal Moving Average

2.3 Best Model Retrieval

In the assessment of time series data, there may be several models available with significant parameters that meet the parameter significance test, so it is necessary to select the best model. Taking the most optimal model in this study is done through the AIC value. One of the best ways is based on the smallest AIC value. The AIC value can show the suitability of the model to the existing data and the resulting value in the future. The following is the formula for calculating the AIC value:

$$AIC(M) = nlog\sigma^2 + 2M$$

Description:

σ^2 = variance of the model residuals.
M = number of parameters in the model.
n = the number of deviations that can be calculated from a series

2.4 MAPE (Mean Absolute Percentage Error)

MAPE is the average percentage error of forecasting against actual data in a given period. A good forecasting method does not have a large error rate. The smaller the error rate, the closer the forecast results will be to the actual data. The MAPE formula can be written as follows:

$$MAPE = \frac{1}{n}\left(\sum_{t-1}^{n}\left|\frac{Z_t - \hat{Z}_t}{Z_t}\right|\right) \times 100\%$$

with,

n = number of observations

Z_t = observed value at time t

$\hat{Z}_t$ = forecasting value at time t t

The accuracy criteria of the error calculation method using MAPE are as follows.

Table 1. MAPE Accuracy Criteria

Forecasting Crteria	MAPE Percentage Limit
Very Good	MAPE ≤ 10%
Good	10% ≤ MAPE ≤ 20%
Simply	20% ≤ MAPE ≤ 50%

2.5 Stages of Data Analysis

The research stage for forecasting Honda car retail sales data using the SARIMA method is :

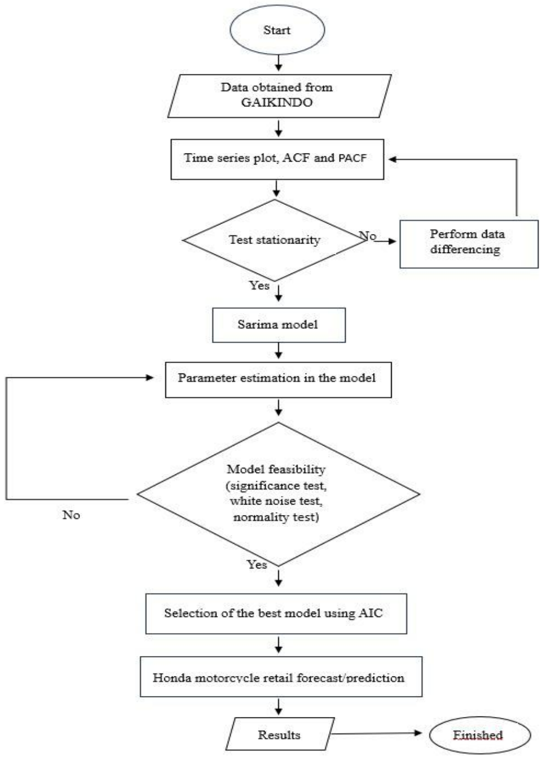


Figure 1. SARIMA *Flowchart*

III. RESULTS AND DISCUSSION

3.1. Descriptive Analysis

Descriptive analysis was applied to extract information from the data. The descriptive analysis techniques applied included the calculation of mean, median, standard deviation, highest score, and lowest score.

Table 2. Deskriptif Analysis

Analysis	Value
Average	11503
Median	12062
Standar Deviation	3824.506
Maxsimum	20523
Minimum	1291

Based on Table 2, the average value of retail sales of Honda cars is 11,503 units, which can be classified as a seasonal increase in sales. In addition, the median value was recorded at 12,062 units, with a standard deviation of 3824.506. The highest sales reached 20,523 units, while the lowest sales were 1,291 units.

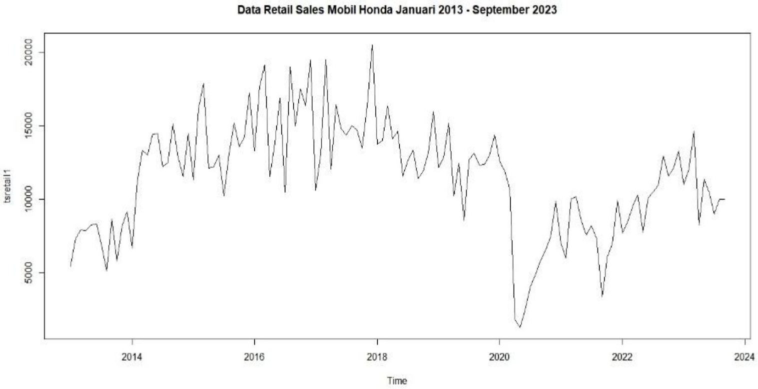


Figure 2. Plot of Honda Car Retail Data

Referring to Figure 2, it can be observed that the pattern of honda car retail marketing data is non-stationary. The data is seasonal after being identified from the pattern of repeated up and down fluctuations at consistent time intervals. In order to facilitate the recognition of this seasonal pattern, a data decomposition process is carried out using RStudio.

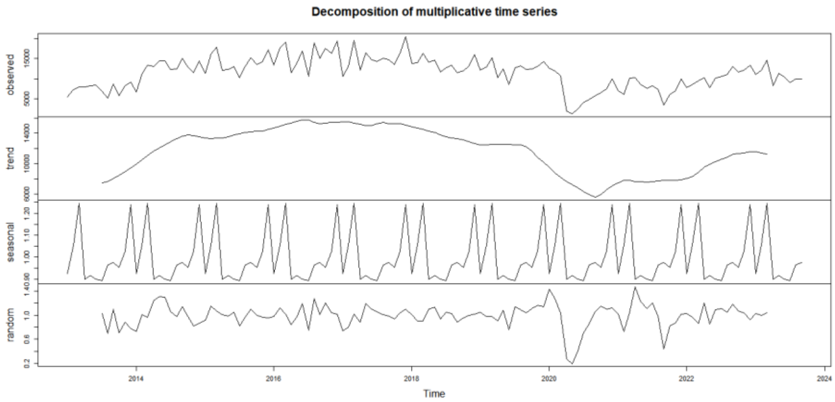


Figure 3. Decomposition Test

Referring to Figure 3, it is identified that the decomposition results of Honda car retail sales data show repeated fluctuations every year.

3.2. Data Stationarity Test

According to the plot in Figure 2, it can be visually seen that the data tends to fluctuate and be inconsistent around the mean value over time, indicating that the data is not yet stationary in the mean. To check stationarity in means, the Augmented Dickey-Fuller (ADF) test can be used.

Hypotesis
H_0 : Not stationarity data

H_1: Stationarity data

Table 3. Testing Results

Test	P-value	Alpha	Decision
ADF	0.325	0.05	Failure to Reject H_0

Referring to Table 3, the result of the ADF test shows a P-value of 0.325, which is smaller than the significance level of 0.05. With a confidence level of 95%, it is concluded that with these data H_0 fails to be rejected. Thus, the Honda car retail selling price data is not stationary.

3.3. Non-Seasonal and Seasonal Differencing

The stationarity test results show that the data has not reached stationarity. If the data is unstable to the mean, it can be converted into stationary data by differencing, which is replacing the original data with the difference between values.

To produce stationary data, differencing is done once non-seasonal (d=1) and to eliminate the strong influence of seasonality, differencing is done once seasonal (D=1). From the differencing results then back tested ADF.

Table 4. Non-seasonal and Seasonal ADF Test Results

Component	Test Statistics	P-value
Non-seasonal	ADF	0.01
Seasonal	ADF	0.01

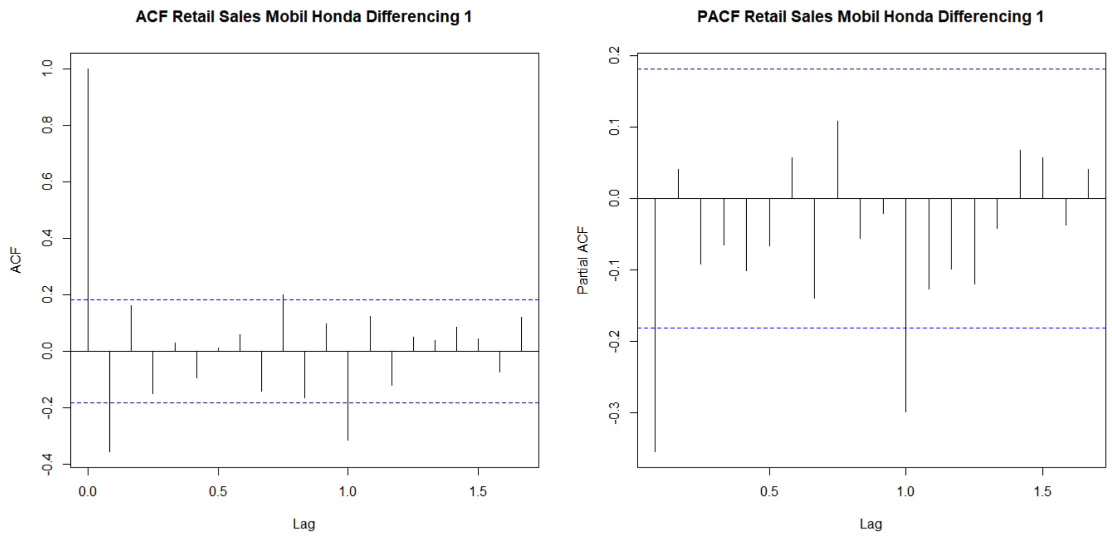
Based on Table 4, after performing non-seasonal differencing, with a confidence level of 95%, the final result obtained is that H_0 fails to be rejected. This means that the Honda car retail sales data is stationary in the average, which then needs to be checked for seasonal data stationarity.

After seasonal differencing, the final result obtained based on the hypothesis, with a confidence level of 95%, is that H_0 fails to be rejected. This means that after seasonal differencing, the Honda car retail selling price data is stationary (does not contain a unit root).

Based on Table 3, it is concluded that the Honda car retail sales data for both components can be said to be stationary when it has undergone 1 time differencing with a P-value of 0.01 each. The next stage is to identify using the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots.

3.4. Model Identification

If it is found that the data is stationary, then the SARIMA model must be identified with the (p, d, q)(P, D, Q)^S arrangement, then plot the honda car retail selling price data after differencing to the ACF and PACF plots.



(a) (b)

Figure 4. (a). ACF plot for non-seasonal and (b). PACF plot for seasonal after differencing

Referring to Figure 4a, there are 2 lags that exceed the limit, so the value of q is obtained as 2. Meanwhile, for order p , the PACF plot shows 1 lag that exceeds the limit, so P is 1. The value of d is 1 because non-seasonal differencing has been performed once.

In order to determine the seasonal model or (P ,D, Q), the number of lags that cross the boundary at multiples of 12, 24, and 36 is examined. Based on Figure 4b on the PACF plot, the value of P is 1 because there are no lags that cross the boundary, so P = 1. Meanwhile, from the ACF plot, there is 1 lag that exceeds the boundary, so the value of Q is 1. The value of D is 1 because seasonal differencing has been performed once.

The results of the model identification obtained the main SARIMA $(1,1,2)(1,1,1)^{12}$ model. The main model that has been obtained is combined into a smaller model order to estimate the model. In this study, a total of 9 models have been combined, with model 10 (the next model) will be applied for model estimation.

3.5. SARIMA Model Forecast

SARIMA model estimation is effective by testing parameters based on the combination of models formed. Parameters are considered significant if their P-value is less than the significance level of 0.05. Of the 10 SARIMA models, four models are significant.

Table 5. SARIMA Model Estimation

No	Model	Parameters	Estimation	P- Value	Decision
1	SARIMA $(1,1,0)(1,1,1)^{12}$	AR1	-0.378424	9.716e-06	Significant
		SAR1	0.279830	0.004989	Significant
		SMA1	-0.999986	5.625e-13	Significant

2	SARIMA $(1,1,0)(1,1,0)^{12}$	AR1	-0.352080	5.012e-05	Significant
		SAR1	-0.326384	0.0002057	Significant
3	SARIMA $(1,1,0)(0,1,1)^{12}$	AR1	0.389593	4.635e-06	Significant
		SMA1	-0.745275	9.894e-09	Significant
4	SARIMA $(0,1,1)(0,1,1)^{12}$	MA1	-0.430639	7.678e-07	Significant
		SMA1	-0.778484	4.807e-09	Significant

Based on Table 5, a model is obtained where all model parameters are significant, namely the SARIMA $(1,1,0)(1,1,1)^{12}$, SARIMA $(1,1,0)(1,1,0)^{12}$, SARIMA $(1,1,0)(0,1,1)^{12}$, and SARIMA $(0,1,1)(0,1,1)^{12}$ models. Furthermore, the four models are carried out diagnostic tests, namely the white noise assumption test and the normality test.

A model whose residuals have no autocorrelation is called white noise. To validate the white noise assumption in the model, the Ljung-Box test is used. The null hypothesis (H_0) is accepted and the model is considered to meet the white noise criteria if the Ljung-Box test results show a P-value greater than the 0.05 significance level.

Table 6. Ljung-Box Test

Model	X-squared	P-value
SARIMA $(1,1,0)(1,1,1)^{12}$,	0.10907	0.7412
SARIMA $(1,1,0)(1,1,0)^{12}$	0.02724	0.8689
SARIMA $(1,1,0)(0,1,1)^{12}$	0.31746	0.5731
SARIMA $(0,1,1)(0,1,1)^{12}$	0.070736	0.7903

Based on Table 6, it is concluded that the four models successfully accept the null hypothesis (H0) and the four models are considered to fulfill the white noise assumption because the P-value is above the significance level of 0.05. Then, a normality test is performed on the model using the Lilliefors test statistic (Kolmogorov-Smirnov).

Tabel 7. Lilliefors Test (Kolmogorov-smirnov)

Model	*P-value*
SARIMA $(1,1,0)(1,1,1)^{12}$,	0.0767
SARIMA $(1,1,0)(1,1,0)^{12}$	0.06452
SARIMA $(1,1,0)(0,1,1)^{12}$	0.07729
SARIMA $(0,1,1)(0,1,1)^{12}$	0.752

Based on Table 7, since after the normality test, the P-value of the four SARIMA models>α (0.05) is obtained, H_0 fails to be rejected. Thus, the residuals are normally distributed.

3.6. Determining The Best Model

The next step is to determine the best model among the models that have met the assumptions of white noise and normal distribution, based on the lowest AIC value among the models.

Table 8. AIC Value

Model	AIC
SARIMA $(1,1,0)(1,1,1)^{12}$	2133.57
SARIMA $(1,1,0)(1,1,0)^{12}$	2152.24
SARIMA $(1,1,0)(0,1,1)^{12}$	2137.99
SARIMA $(0,1,1)(0,1,1)^{12}$	2136.7

Based on Table 8, the smallest AIC value obtained in the SARIMA (1,1,0)(1,1,1)(12) model is 2133.57.

Based on the previous parameter estimation and significance test (Table 5), the SARIMA model is obtained, namely:

$$(1 - 0{,}378424B)(1 - 0{,}279830B^{12})(1 - B)^1(1 - B^{12})^1 Y_t = (1 - 0{,}999986B^{12})\varepsilon t$$

Then the SARIMA $(1,1,0)(1,1,1)^{12}$ model is suitable for forecasting Honda car retail sales data for the next 12 months.

3.7. Forecasting

Furthermore, forecasting Honda car sales in the retail market for the next year, from October 2023 to September 2024, will be carried out using the best SARIMA model, namely SARIMA $(1,1,0)(1,1,1)^{12}$.

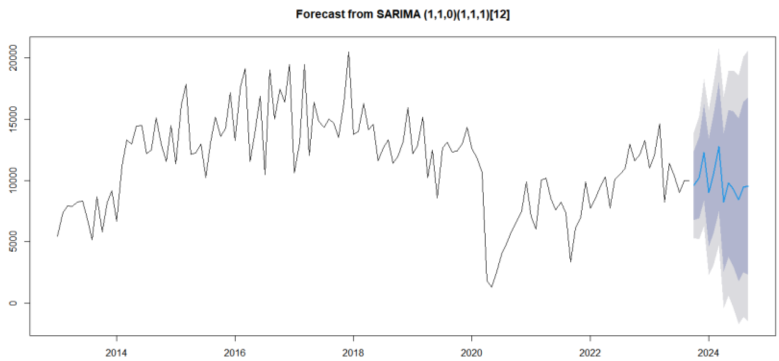


Figure 5. Chart of Forecasting Result Data

Based on Figure 5, shows that the results of Honda car retail sales forecasting tend to fluctuate following seasonal patterns.

Table 9. Forecasting Results

Period	Forecasting Results
October 2023	9596.553
November 2023	10240.877
December 2023	12304.030
January 2024	9019.018
February 2024	10547.180
March 2024	12761.674
April 2024	8228.086
May 2024	9779.552
June 2024	9319.434
July 2024	8422.163
August 2024	9480.293
September 2024	9577.796

Based on Table 9, the results of forecasting for the next 12 months starting from October 2023 to September 2024, sales are expected to peak in March 2024, with a total of 12,762 units.

3.8. MAPE Value

MAPE is the proportion of error from the mean in absolute terms. This method produces information on how much the prediction error is when measured against the actual value.

Table 10. MAPE Value

Model	MAPE (%)
SARIMA $(1,1,0)(1,1,1)^{12}$	17.7491
SARIMA $(1,1,0)(1,1,0)^{12}$	20.22867
SARIMA $(1,1,0)(0,1,1)^{12}$	19.00666
SARIMA $(0,1,1)(0,1,1)^{12}$	18.97876

Based on Table 10, the lowest MAPE value obtained in the SARIMA (1,1,0)(1,1,1)(12) model is 17.7491%. This means that the model can be claimed to be efficient for forecasting Honda car retail sales data even though there is an intervention in the actual data in 2020 due to the COVID-19 case.

IV. CONCLUTION

Referring to the results of the analysis and discussion, it can be concluded that the most effective model for forecasting retail sales of Honda cars using the SARIMA method is SARIMA $(1,1,0)(1,1,1)^{12}$ with a MAPE accuracy of 17.74%, which means that the model's predictive ability is quite good. The SARIMA $(1,1,0)(1,1,1)^{12}$ model has the following model equation.

$$(\mathbf{1}-\mathbf{0},\mathbf{378424}\boldsymbol{B})(\mathbf{1}-\mathbf{0},\mathbf{279830}\boldsymbol{B}^{\mathbf{12}})(\mathbf{1}-\boldsymbol{B})^{\mathbf{1}}(\mathbf{1}-\boldsymbol{B}^{\mathbf{12}})^{\mathbf{1}}\boldsymbol{Y}_{\boldsymbol{t}}=(\mathbf{1}-\mathbf{0},\mathbf{999986}\boldsymbol{B}^{\mathbf{12}})\boldsymbol{\varepsilon t}$$

After obtaining the best model, then forecasting is carried out for the next year, starting from October 2023 to September 2024. The forecast results show that retail sales of Honda cars experience variations. According to the prediction, the highest sales occurred in March at 12,762 units and the lowest sales in April at 8,229 units. To ensure increased sales, it is important to pay attention to the

trends generated by the model and consider factors that can affect the retail sales of Honda cars such as economic factors, government policies on taxes, product innovations, and promotions through automotive exhibitions. Regular monitoring and adjustment of the model is also necessary to maintain the accuracy of the predictions.

REFERENCES

[1] Gabungan Industri Kendaraan Bermotor Indonesia (GAIKINDO). (2023). Data Penjualan Kendaraan Bermotor. Sumber: https://www.gaikindo.or.id/statistik-penjualan. Diakses 28 Juni 2024.

[2] Honda Indonesia. (2023). Sejarah dan Profil Perusahaan. Sumber: https://www.honda-indonesia.com/profil. Diakses 28 Juni 2024.

[3] Gabungan Industri Kendaraan Bermotor Indonesia (GAIKINDO). (2023). Mobil Terlaris di Indonesia 2022. Sumber: https://www.gaikindo.or.id/mobil-terlaris-2022. Diakses 28 Juni 2024.

[4] I. H. Susilowati, "Peramalan Penjualan Mobil pada PT Toyota Astra Motor Indonesia dengan Metode Trend Semi Average dan Metode Least Square," JIMF (Jurnal Ilmiah Manajemen Forkamma), vol. 6, no. 1, p. 28, Nov. 2022

[5] Kementerian Perindustrian Republik Indonesia. (2023). Laporan Industri Otomotif 2020-2023. Sumber: https://www.kemenperin.go.id/laporan-industri-otomotif. Diakses 28 Juni 2024.

[6] Gabungan Industri Kendaraan Bermotor Indonesia (GAIKINDO). (2023). Tren Musiman Penjualan Mobil. Sumber: https://www.gaikindo.or.id/tren-musiman-penjualan. Diakses 28 Juni 2024

[7] I. H. Susilowati, "Peramalan Penjualan Mobil pada PT Toyota Astra Motor Indonesia dengan Metode Trend Semi Average dan Metode Least Square," JIMF (Jurnal Ilmiah Manajemen Forkamma), vol. 6, no. 1, p. 28, Nov. 2022

[8] Box, G.E.P., Jenkins, G.M., & Reinsel, G.C. (2015). *Time series* Analysis: Forecasting and Control. Wiley

[9] Muhammad Irfan Rizki, Teguh Ammar Taqiyyuddin, "Penerapan Model SARIMA untuk Memprediksi Tingkat Inflasi di Indonesia," JSMS (Jurnal Sains Matematika dan Statistika), vol. 7, no. 2, p. 67-72, Juli. 2021

[10] Laila Qadrini, Asrirawan Asrirawan, Nur Mahmudah, Muhammad Fahmuddin, Ihsan Fathoni Amri, "Forecasting Bank Indonesia Currency Inflow and Outflow Using ARIMA, *Time series* Regression (TSR), ARIMAX, and NN Approaches in Lampung," JMSK (Jurnal Matematika, Statistika & Komputasi), vol. 17, no. 2, p. 166-177, Jan. 2021

[11] M. I. Rizki and T. A. Taqiyyuddin, "Penerapan Model SARIMA untuk Memprediksi Tingkat Inflasi di Indonesia," Jurnal Sains Matematika dan Statistika, vol. 7, no. 2, Aug. 2021

[12] McKinsey & Company. (2023). Competitive Strategies in Automotive Market. Sumber: https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/competitive-strategies-in-automotive-market. Diakses 28 Juni 2024

[13] Anis Mahfud Al'afi, Widiarti Widiarti, Dian Kurniasari dan Mustofa Usman, "Peramalan Data *Time series* Seasonal Menggunakan Metode Analisis Spektral", Jurnal Siger Matematika, Vol 1, No 1, 10-15, 2020

[14] D. S. Fahik and M. T. Jatipaningrum, "Peramalan Jumlah Penumpang Penerbangan Internasional di Bandar Udara Internasional Soekarno Hatta dengan Metode Holt Winters Exponential Smoothing dan Seasonal Arima," Jurnal Statistika Industri dan Komputasi, vol. 6, no. 01, pp. 77–87, 2021

[15] Z. Li, J. Bi, and Z. Li, "Passenger Flow Forecasting Research for Airport Terminal Based on SARIMA *Time series* Model," 2017