Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 9, Nomor 2, Agustus 2024, pp 868-876
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

Time Series Implementation for Sales Forecasting of Furniture Products at PT XYZ

Pragnanta Yopie Pramastya[1], Evangs Mailoa[2]
[1,2]Universitas Kristen Satya Wacana, Salatiga, Indonesia
Email :672020133@student.uksw.edu[1], evangs.mailoa@uksw.edu[2]

Abstract

PT XYZ is an importer, distributor, and seller of furniture. The company operates in two business lines: B2B (Business to Business) and B2C (Business to Consumer). The B2B segment is divided into Agents/Local Distributors, Modern Trade, and Partner Retail Stores, while the B2C segment is divided into Direct Sales and Online Intermediaries. This study applies the Time Series method to forecast future furniture sales using the SARIMA (Seasonal Autoregressive Integrated Moving Average) model. The results of the study using the SARIMA model provide sales forecasts from 2024 to 2026.

***Keywords**: Prediction, Time Series, Sales*

1. INTRODUCTION

Business activities in the field of marketing have rapidly developed alongside Indonesia's economic growth. Furniture is one of Indonesia's craft industries that can compete with other industries. Furniture focuses on home and office equipment such as chairs, tables, doors, windows, and cabinets [1]. The current business competition demands that business actors be more consistent with the desires and needs of consumers regarding the products offered, necessitating accurate predictions in sales planning and inventory management [2]. Sales forecasting is an essential step for every company and can help predict sales for the coming years [3]. Forecasting or sales prediction is beneficial for obtaining information for sales planning and inventory management in a company [4].

PT XYZ is a company engaged in furniture sales. The company experienced a decline in sales in 2021 due to the COVID-19 pandemic. The company faced problems in controlling inventory that did not match consumer demand, resulting in overstock and losses. Accurate forecasting in sales planning for the future can help address these issues [5]. This research applies the Time Series method to predict future furniture sales at PT XYZ using the SARIMA (Seasonal Autoregressive Integrated Moving Average) model. The Time Series method can help identify different patterns in sales data over several years [6]. Similar research with the implementation of the Time Series method has been conducted for sales data at PT Gaikindo. The testing yielded the best SARIMA model prediction with parameters (p) = 1, (d) = 1, (q) = 1, (P) = 1, (D) = 0, (Q) = 0, and (s) = 12 with an AIC value of -187.929 [5]. The research question formulated is whether the SARIMA time series model can determine sales and optimize furniture inventory management in the future.

2. RESEARCH METODOLOGY

The stages of this research are illustrated in the diagram shown in Figure 1.

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 9, Nomor 2, Agustus 2024, pp 868-876
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

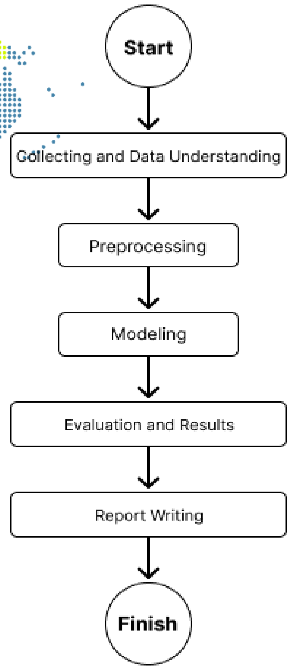

Figure 1. Research Methodology

The first stage is to collect sales data from PT XYZ. The dataset used includes previous sales data from the period of 2020 to March 2024. An understanding of the data obtained from PT XYZ is then developed. The second stage is preprocessing. This involves cleaning the data by removing unnecessary variables for forecasting, such as Document No, Sell to customer No, Quantity, Unit Price, Disc normal, Disc customer, and Disc product. The variables used are Date_order and Sales.

The third stage is modeling. The rolling statistics test is conducted to check if the dataset is stationary [6]. The rolling statistics test is presented in the form of a data plot. The results of the rolling statistics test are divided into two conditions: if the actual data and the rolling data have a significant difference, it means the data series is not stationary [7]. The Augmented Dickey-Fuller (ADF) test is then used to make the data stationary by differencing it once, taking into account the p-value and confidence levels of 1%, 5%, and 10% [8].

Next, the dataset is split into training and observation sets, with the 2021 sales data used for training. The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) values are then calculated. The SARIMA model is applied by determining the parameter values p, d, and q. These parameters are set as p (order of autoregression) to determine the number of lag variables needed, d (order of integration) to determine the number of differences, and q (order of moving average) [9].

The predicted values are visualized and compared with actual sales data. Finally, the forecast results are plotted for the next three years.

The forecasting results are evaluated using MAPE (Mean Absolute Percentage Error).

Table 1. MAPE Evaluation Criteria [5]

MAPE	Parameter
<10%	Highly accurate Forecasting
10% - 20%	Good forecasting accuracy
20% - 50%	Reasonable forecasting accuracy
>50%	Weak and inaccurate predictability

Based on the evaluation parameters in Table 1, if the MAPE value is less than 10%, the method is considered to have highly accurate forecasting. If the MAPE value is between 10% and 20%, the method is considered to have good forecasting accuracy. If the MAPE value is between 20% and 50%, the method is considered to have reasonable forecasting accuracy. If the MAPE value is greater than 50%, the method is considered to have weak and inaccurate predictability. The research process concludes with the writing of the report.

3. RESULT AND DISCUSSION

The study utilized several libraries for forecasting, implementing the SARIMA model, evaluating methods, and displaying data plots.

3.1. Collecting and data understanding

The dataset obtained consists of sales data from October 2020 to April 2024. The dataset comprises 9 variables and 53,520 records. Detailed data can be seen in Figure 2.

Row id	Document No.	Order Date	Sell-to Customer No.	Quantity	Unit Price	Disc. Normal	Disc. Product	Disc. Customer	Sales
1	PSI102-202010-0001	2020-10-02	102-0000063	2	2528000	30	9,77443	5	2757818
2	PSI102-202010-0002	2020-10-03	102-0000011	1	2315000	30	10,1641	5	1257274
3	PSI102-202010-0003	2020-10-05	102-0000216	30	837000	30	29,04	10	10204887
4	PSI102-202010-0003	2020-10-05	102-0000216	120	308000	30	29,04	10	15020813
5	PSI102-202010-0003	2020-10-05	102-0000216	25	837000	30	29,04	10	8504072
....									
53515	PSI102-202404-0215	2024-04-22	102-0000058	2	1750000	40	0	0	1891892
53516	PSI102-202404-0215	2024-04-22	102-0000058	8	385000	40	0	0	1664865
53517	PSI102-202404-0215	2024-04-22	102-0000058	3	1750000	40	0	0	2837838
53518	PSI102-202404-0215	2024-04-22	102-0000058	12	385000	40	0	0	2497297
53519	PSI102-202404-0216	2024-04-22	102-0000251	1	1550000	0	0	0	1396396
53520	PSI102-202404-0216	2024-04-22	102-0000251	1	1550000	0	0	0	1396396

Figure 2. Sales Data of PT XYZ

3.2. Preprocessing

The dataset consists of Document No, Sell to customer No, Quantity, Unit Price, Disc normal, Disc customer, Disc product, Order Date, and Sales. For this research, only the variables Order Date and Sales will be used, and the other variables will be removed. Selects the variables to be used, namely 'Order Date' and 'Sales', and displays the data in Table 2.

Table 2. Sales Data

	Order Date	Sales
0	2020-10-02	2757818
1	2020-10-03	1257274
2	2020-10-05	10204887
3	2020-10-05	15020813
4	2020-10-05	8504072
...	...	...
53515	2024-04-22	1664865
53516	2024-04-22	2837838
53517	2024-04-22	2497297
53518	2024-04-22	1396396
53519	2024-04-22	1396396

Retrieves the data by converting daily sales data into monthly sales data, taking the average sales for each month, and visualizing the monthly sales data as shown in Figure 3.

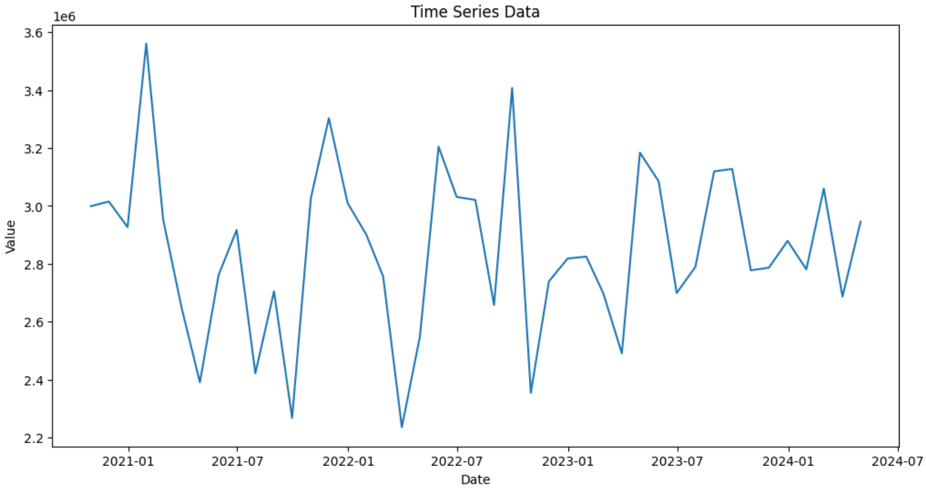

Figure 3. Plot of Sales Data from PT XYZ

The graph illustrates the monthly sales data at PT XYZ from October 2020 to April 2024. The graph indicates a decline in sales during several months in 2021 and 2022. This decrease in sales is attributed to the aftermath of the COVID-19 pandemic recovery period. Sales began to stabilize and rise from 2023 onward, continuing into 2024.

3.3. Modeling

Augmented Dickey-Fuller (ADF) test is a statistical test used to check whether a time series data has a unit root, indicating that the series is non-stationary. Stationarity of the data is checked using the Augmented Dickey-Fuller (ADF) test with the formula [10]:

$$\Delta Yt = \beta 1 + \beta 2 + \delta Yt\text{-}1 + \phi i \sum k\ i = 1\ \Delta Y t - 1 + \varepsilon t \qquad (1)$$

Where :

ΔYt : First Difference of Y, indicating the change in Y at time

β1 : Constant or intercept term

β2 : Coefficient for trend, representing the trend component over time t

δ : Coefficient for lagged Y

ϕ : Coefficient for lagged differences of Y

ε : Error term

k : Lag

t : Time

The ADF test resulted in an ADF statistic of -5.845714389779301, indicating that the data exhibits stationarity. A dataset is considered stationary if the p-value is less than 0.05. In this case, the obtained p-value is 3.6878457549448 41e-07, which is less than 0.05. Therefore, based on these results, the null hypothesis (H0) is rejected, and the alternative hypothesis (H1) is accepted, indicating that the data is stationary.

3.3.1. Parameter Model Identification

Identifying the SARIMA model parameters (p, d, q, P, D, Q, s) using the ACF and PACF plots. ACF measures the correlation between observations in a time series that are separated by various lags with the formula [11]:

$$\rho(k) = \frac{\sum_{t=k+1}^{T} (yt - \bar{y})(yt - k - \bar{y})}{\sum_{t=1}^{T} (yt - \bar{y})^2} \qquad (2)$$

Where :

$\rho(k)$: ACF value at lag k

T : Total number of observations in the time series

yt : Observation value at time t

$\bar{y}$: Mean of all observation values in the time series

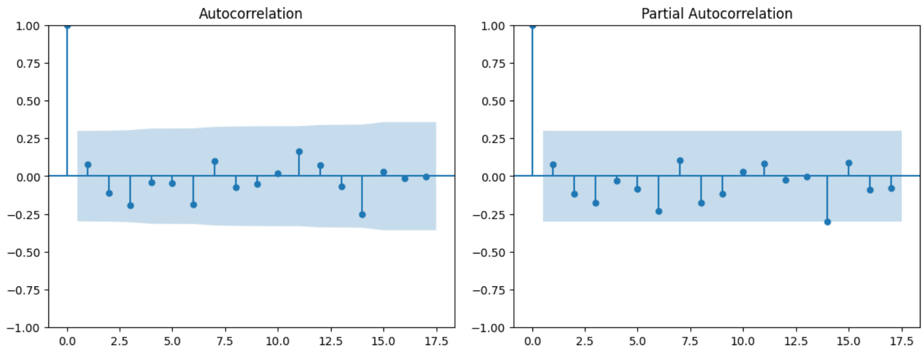

Figure 4. Plot ACF and Plot PACF

Both plots have a horizontal axis indicating lags and a vertical axis showing correlation coefficients ranging from -1 to 1, where -1 indicates perfect negative correlation, 0 indicates no correlation, and 1 indicates perfect positive correlation.

Volume 9, Nomor 2, Agustus 2024, pp 868-876
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

The shaded blue area represents the confidence interval for the correlation coefficients. Based on Figure 4, the PACF plot shows no lags exceeding the confidence interval lines, indicating that the data is stationary.

3.3.2. Data separation

Data splitting in SARIMA (Seasonal Autoregressive Integrated Moving Average) prediction serves to ensure that the constructed model can make accurate and reliable forecasts. The dataset is divided into two parts: training data and testing data. Splitting the monthly_sales dataset for 2021 into training data and using the monthly_sales dataset for 2022 as testing data.

3.3.3. Best Parameters Sarima

Finding the optimal parameters for the SARIMA (Seasonal Autoregressive Integrated Moving Average) model using the grid search method with the equation [11]:

$$\begin{aligned}&\phi_p(\mathrm{B})\Phi_p B^S(1-\mathrm{B})^d\,(1\,-B^S)^D\,\dot{\mathrm{Z}}_t\\&=\theta_q\,(\mathrm{B})\Theta_Q(B^S)\mathrm{a}_t\end{aligned} \tag{3}$$

Where :

$\phi_p(\mathrm{B})$: Non-seasonal autoregressive operator
$\Phi_p B^S$: Seasonal autoregressive operator
$(1-\mathrm{B})^d$: Non-seasonal differencing order d
$(1\,-B^S)^D$: Seasonal differencing order D
$\theta_q\,(\mathrm{B})$: Non-seasonal moving average operator
$\Theta_Q(B^S)$: Seasonal moving average operator
$\dot{\mathrm{Z}}_t$: $\dot{\mathrm{Z}}_t - \mu$
s : Number of seasonal periods
a_t : : Error term at time t

The best SARIMA model parameters ((0, 1, 1), (0, 1, 0, 12)) were identified with a minimum AIC of 57.44. The analysis reveals a pattern using the seasonal SARIMA model of order (0, 1, 1)x(0, 1, 0, 12), comprising non-seasonal components (p, d, q) with autoregressive (AR) order 0, moving average (MA) order 1, and differencing order 1. The seasonal components (P, D, Q, S) include seasonal autoregressive (SAR) order 0, seasonal moving average (SMA) order 0, seasonal differencing order 1, a seasonal period of 12, and an Akaike Information Criterion (AIC) value of 57.44.

3.3.4 Implementation of the SARIMA Model

Application of the SARIMA model and visualization of the training, testing, and forecasted data plots. The visual results can be seen in Figure 5.

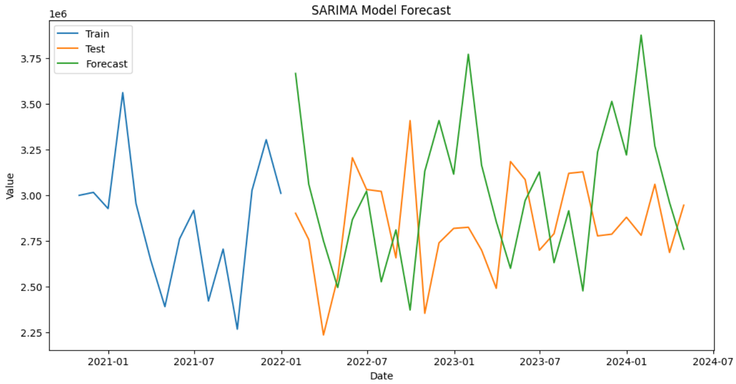

Figure 5. SARIMA Model Forecast

Plotting the training data with a blue line, the testing data with an orange line, and the forecasted data with a green line. The plot shows that the forecasted values for the test period closely match the actual testing data. The forecasted values appear higher compared to both the training and testing data, resulting in forecasts from January 2022 to March 2024 showing a graph that closely resembles and aligns with the testing data.

3.3 Evaluation and Result

Model evaluation using MAPE with the formula [5]:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{A_t - F_t}{A_t} \text{ x } 100\% \qquad (4)$$

The MAPE evaluation resulted in an accuracy of 16.02%. Referring to Table 1, this accuracy falls within the range of 10% - 20%, indicating that the model's forecasting performance is considered good forecasting accuracy, making it sufficiently reliable for use. Based on the analysis conducted on sales data at PT XYZ from 2020 to 2024, using the SARIMA method yielded a MAPE calculation of 16.02%, which falls within the criteria of being good, indicating that the model is suitable for forecasting. With a lower MAPE value, sales predictions for the next 3 years were made using Google Colaboratory with the SARIMA method. The forecast results for the next 3 years can be seen in Figure 6.

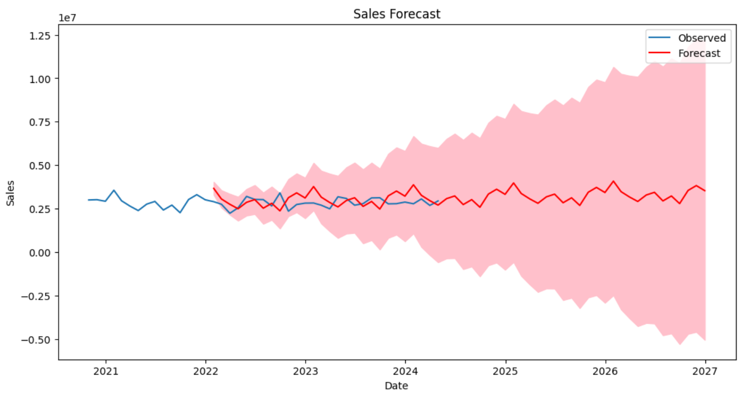

Figure 6. Plot Sales Forecast

The sales forecast results from May 2024 to December 2026. The blue line represents the actual data, while the red line depicts the forecasted results. From the graph, it is evident that the forecasted values closely track the actual data, with the forecast values being higher than the actual data. Consequently, the average sales forecast from May 2024 to December 2026 is obtained. Detailed monthly forecast results can be seen in Table 3.

Table 3. Forecast Result

Bulan	Prediksi
Mei 2024	3.075533
Juni 2024	3.231454
Juli 2024	2.736044
Agustus 2024	3.019537
September 2024	2.581949
Oktober 2024	3.340366
November 2024	3.617416
Desember 2024	3.324370
Januari 2025	3.979439
Februari 2025	3.372900
Maret 2025	3.064290
April 2025	2.809774
Mei 2025	3.180287
Juni 2025	3.336209
Juli 2025	2.840799
Agustus 2025	3.124291

4. SIMPULAN

Based on the research findings, it can be concluded that the SARIMA time series model successfully predicts the sales outcomes of PT XYZ from 2024 to 2026. The SARIMA method applied to sales data from PT XYZ produced monthly sales predictions for the years 2024 to 2026, with an evaluated MAPE accuracy of 16.02%. This indicates that the SARIMA model meets the criteria for good performance in prediction tasks and SARIMA time series model can determine sales and optimize furniture inventory management in the future.

DAFTAR PUSTAKA

[1] D. D. Prasetya *Et Al.*, “Analisis Strategi Pemasaran Guna Meningkatkan Penjualan Furniture Pada Finture . Id (Marketing Strategy Analysis To Increase Furniture Sales At Finture . Id) Pendahuluan,” Pp. 36–41, 2022.

[2] Salwa, N. Tatsara, R. Amalia, And A. F. Zohra, “Peramalan Harga Bitcoin Menggunakan Metode Arima Autoregressive Integrated Moving Average ,” J. Data Anal., Vol.1, No.1, Pp. 21–31, 2018, Doi: 10.24815/Jda.V1i1.11874

[3] W. Satria, “Jaringan Syaraf Tiruan Backpropagation Untuk Peramalan Penjualan Produk (Studi Kasus Di Metro Electronic Dan Furniture),” *Djtechno J. Teknol. Inf.*, Vol. 1, No. 1, Pp. 14–19, 2021, Doi: 10.46576/Djtechno.V1i1.966

[4] A. Kurniawati, M. S. Ahmad, M. Fhadli, And S. Lutfi, “Analisis Perbandingan Metode Time Series Forecasting Untuk Prediksi Penjualan Obat Di Apotek (Studi Kasus : Kimia Farma Apotek Takoma)\ Vol. 3, No. 1, Pp. 96–106, 2023, Doi: 00.0000/Jati.

[5] M. H. Widiyanto, R. Mayasari, F. I. Komputer, U. S. Karawang, And T. Timur, "Implementasi Time Series Pada Data Penjualan Di Gaikindo Menggunakan Algoritma Seasonal Sarima," Vol. 7, No. 3, Pp. 1501–1506, 2023.
[6] K. Silalahi, "Forecasting Of Poverty Data Using Seasonal Arima Modeling In West Java Province," Vol. 4, No. 1, Pp. 76–86, 2020.
[7] Utomo, P. (2020). Peramalan Jumlah Penumpang Kereta Api Di Indonesia Menggunakan Metode Seassonal Autoregressive Integrated Moving Average (Doctoral Dissertation, Uin Sunan Ampel Surabaya).
[8] S. Suseno And Suryo Wibowo, "Penerapan Metode Arima Dan Sarima Pada Peramalan Penjualan Telur Ayam Pada Pt Agromix Lestari Group," J. Teknol. Dan Manaj. Ind. Terap., Vol. 2, No. I, Pp. 33–40, 2023, Doi: 10.55826/Tmit.V2ii.85.
[9] G. Christie, D. Hatidja, And R. Tumilaar, "Penerapan Metode Sarima Dalam Model Intervensi Fungsi Step Untuk Memprediksi Jumlah Pegunjung Objek Wisata Londa (Application Of The Sarima Method In The Step Function Intervention To Predict The Number Of Visitors At Londa Tourism Object)," J. Ilm. Sains, Vol. 22, No. 2, P. 96, 2022, Doi: 10.35799/Jis.V22i2.40961.
[10] A. F. Muzakki, D. Aditama, And I. Gita Anugrah, "Penerapan Metode Autoregressive Integrated Moving Average Untuk Memprediksi Penggunaan Barang Medis Pada Logistik Medis Rumah Sakit Muhammadiyah Gresik," *Indexia*, Vol. 4, No. 1, P. 1, 2022, Doi: 10.30587/Indexia.V4i1.3595.
[11] Y. I. Katabba, "Metode Seasonal Autoregressive Integrated Moving Average (Sarima) Untuk Memprediksi Jumlah Penumpang Kereta Api Di Pulau Sumatera," P. 83, 2021.