JIKO (Journal of Informatics and Computers) Accredited KEMENDIKBUD RISTEK, No.105/E/KPT/2022
Vol. 7, No. 2, August 2024, pp. 71-78 p-ISSN: 2614-8897
DOI: 10.33387/jiko.v7i2.8266 e-ISSN: 2656-1948

FORECASTING SALES USING SARIMA MODELS AT THE SINAR PAGI BUILDING MATERIALS STORE

Ahmad Adiib Aminullah[1], Mohammad Idhom[2], Wahyu Syaifullah Jauharis Saputra[3]

[1,2,3]Department of Data Science, Universitas Pembangunan Nasional Veteran Jawa Timur, Surabaya, Indonesia
[*]Email: [1]20083010015@student.upnjatim.ac.id, *[2]idhom@upnjatim.ac.id,
[3]wahyu.s.j.saputra.if@upnjatim.ac.id

(Received: 24 June 2024, Revised: 11 July 2024, Accepted: 23 July 2024)

Abstract

Sinar Pagi Building Materials Store faces the challenge of maintaining optimal stock levels of goods to avoid excess and understock, which affects customer satisfaction and operational efficiency. This study applies the Seasonal Autoregressive Integrated Moving Average (SARIMA) method to forecast sales in the store. Leveraging its ability to model seasonal patterns on historical sales data, various SARIMA models were analyzed and compared using the Akaike Information Criterion (AIC) and Root Mean Square Error (RMSE). The dataset is divided by a 95:5 ratio into training and testing sets for robust evaluation. The results show that the SARIMA model with SARIMA notation $(p,d,q)(P,D,Q)^m$ has the best model value of $(1,0,0)\ (3,1,1)^7$. This model is the most suitable model based on the lowest AIC value of 1245 and the lowest RMSE of 7,95 compared to other SARIMA models after model identification using the model looping test. For other models such as model (1,0,1) $(3,1,1)^7$ and (0,0,1) $(3,1,1)^7$, the AIC and RMSE values are greater, namely model (1,0,1) $(3,1,1)^7$ with AIC 1246 and RMSE of 8,05, while model (0,0,1) $(3,1,1)^7$ gets an AIC of 1252 and an AIC of 8,15 .The lower the AIC value, the better the model and the lower the RMSE value, the better the model. This shows a superior balance between model complexity and prediction accuracy. The model manages to capture seasonal patterns in sales data, providing a pretty good prediction framework. This study shows that the SARIMA (1,0,0) $(3,1,1)^7$ model is effective in the accuracy of the sales forecasting process so that Sinar Pagi Building Materials Store can make more reliable sales predictions, which can help in inventory planning and marketing strategies.

Keywords: *SARIMA, Sales Forecasting, Predictive Modeling, RMSE*

**Corresponding Author: Mohammad Idhom*

1. INTRODUCTION

Sinar Pagi Building Materials Store, a major player in the building materials and construction tools market, faces the challenge of maintaining adequate stock levels of goods while minimizing overstocking and understocking. This balance is crucial to meet customer demands in a timely manner and maintain a high level of customer satisfaction [1].

Determining the right level of inventory of goods is very important because it relates to the flow of money and can affect the performance of the store [2]. Excessive inventory can lead to buildup in warehouses and reduce capital. Conversely, too little inventory can lead to running out of goods so that sales that should have happened become missed, and profits from the store are reduced [3]. Therefore, good inventory management is needed with a focus on two things: improving customer satisfaction and keeping inventory from emptying or piling up, while keeping inventory costs as low as possible [4].

One of the effective tools in inventory management is sales forecasting. Forecasting is the activity of estimating (measuring) the magnitude or quantity of something in the coming time [5]. Sales forecasting is a business activity that aims to predict the future sales of goods [6]. This can be done by taking historical data and projecting sales data into the future [7]. Sales forecasting not only helps in predicting future sales but also ensures optimal inventory management, efficient resource allocation, and increased customer satisfaction.

Various methods can be used for sales forecasting, in this study the appropriate method to formulate sales forecasting problems is the SARIMA (Seasonal Auto-Regressive Integrated Moving Average) method. The implementation of the SARIMA method requires the analysis of historical sales data to identify seasonal patterns and existing trends [8]. Once the SARIMA model is built and validated using the error percentage rate, the forecasting results can be used to make better inventory management decisions.

Various related studies that have been carried out previously in sales forecasting using the SARIMA method by Wawan Gunawan obtained an RMSE error percentage rate of 3.61 with the best SARIMA model, namely (0,0,0) The dataset used is tire and wheel sales data in 2021 for 5 months, namely June, July, August, September, and October. With 854 rows of data divided into training data and test data with a ratio of 80:20 $(0,1,1)^{12}$ [3]. SARIMA has an advantage in predicting because in the process there is a feature to find the best parameters to be implemented on the model.

Another study by Hakim used the SARIMA method in forecasting the price of purebred chicken eggs. An RMSE value of 1491.30 and a MAPE of 3.40% were obtained with the best SARIMA model of (0,1,0). The dataset in this study is data on the average daily price of chicken eggs from the traditional market in Manokwari district in January 2018 – February 2024 $(0,1,1)^{12}$[8]. SARIMA has the advantage of predicting historical data that is seasonally patterned [9].

The sales data of goods at the Sinar Pagi Building Materials Store is a type of time series data. Autoregressive Integrated Moving Average (ARIMA) is one of the methods that is usually used to forecast time series data based on historical data. This method is a development of the method carried out by George Box and Gwilym Jenkins in combining the Autoregressive (AR) and Moving Average (MA) models [10]. ARIMA is used about short-term prediction. However, the ARIMA method does not have a special component to analyze certain seasonal patterns in data. Therefore, the ARIMA method has a shortcoming in predicting data that has a seasonal pattern over a period of time [11]. Based on the above background, the sales data at the Sinar Morning Building Materials Store has a seasonal pattern that repeats over a period of time. So forecasting methods that pay attention to seasonal patterns are more appropriately used, such as the Seasonal Autoregressive Integrated Moving Average (SARIMA) method.

This study will use the Seasonal Autoregressive Integrated Moving Average (SARIMA) method in forecasting sales of goods. By utilizing technology capable of handling seasonal data, the SARIMA method has the potential to become a more effective tool in accelerating and improving the accuracy of the sales forecasting process. SARIMA model has advantages in its well-known statistical properties and an effective modeling process and can be used when the seasonal time series is stationary and has no missing data [12]. In this study, the best model of the SARIMA method will be analyzed and the percentage of error will be calculated with the Root Mean Square Error (RMSE). By obtaining the best SARIMA model value with a low percentage of error using a low Root Mean Square Error (RMSE) value, the forecast results will be better. This effort, is expected to make a significant contribution to efficient inventory management, improve store operational efficiency, and support economic growth in the trade sector.

2. RESEARCH METHOD

As a guide in carrying out re-searching to achieve the expected goals, the research methodology is prepared systematically. The stages of this research are explained using a research flow diagram. Figure 1 illustrates the stages of the research to be followed and serves as a complete guide during the research process.

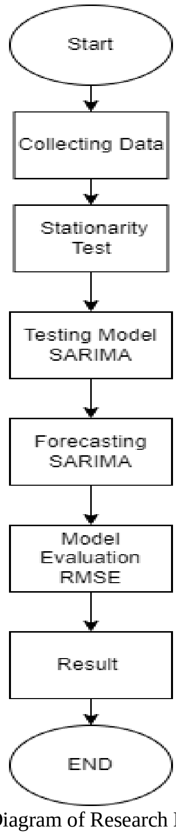


Figure 1. Diagram of Research Method

2.1 Collecting Data

The data used is cement sales data at the Sinar Morning Building Shop which is accumulated on a daily basis with a span of 7 months, namely from September 1, 2023 to March 31, 2024. The data is a series of data where the data totals 231 data consisting of 2 variables, namely the Date and Sales variables.

2.2 Stationarity Test

Stationary testing is a statistical procedure used to determine whether or not a time series data has a stationary nature. Time series data is said to be stationary if its statistical properties remain constant over time, such as averages, variances, and coarians between observations [13]. Stationary means that the data does not show significant fluctuations around the base value or zero. This means that there is no clear tendency to rise or fall in the medium or long term, which is often referred to as a trend. Trends and seasonal components in time series data can be separated, but when the data is said to be stationary, this usually applies only to the seasonal component. Data stationarity is one of the assumptions that must be fulfilled in forecasting [14].

2.3 Auto Regressive Integrated Moving Average

The Auto-Regressive Integrated Moving Average (ARIMA) is a method that completely ignores the independent variables in making forecasts. In making short-term forecasting, ARIMA uses past and present values according to dependent variables. The use of historical value of variables is used to select good statistical interactions, which is done by comparison of the predicted variables as the goal of ARIMA [15]. The ARIMA method has a high level of accuracy in conducting short-term forecasting, but for long-term forecasting, the forecasting accuracy is not good [16].

The three characteristics of the ARIMA model are parameters (p, d, q) where each is the 'p' model is Autoregressive (AR), 'q' Moving Average (MA), and 'd' Integrated (I). The time series (AR) model has mean, white noise variance, and p parameters. The general form of the Autoregressive (AR) process is defined as follows [17] :

$$X_t = \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_0 X_{t-p} + \varepsilon_t \quad (1)$$

X_tWhere the information in the formula above is a stationary series, β1, β2, β3 as constants and model coefficients, X_{t-p} as the past value of the series concerned, and εt as a prediction error (error). Meanwhile, the Moving Average (MA) model is obtained by adding and finding the average value of a certain number of periods, then removing past values and adding new values. The general form of the Moving Average (MA) process is defined as follows [17]:

$$X_t = \beta_0 + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \cdots - \beta_q \varepsilon_{t-q} \quad (2)$$

Where the information in the formula above is as a stationary series value, X_tεt−1, εt−2 as past forecasting errors, and β1, β2, β3 as constants. Meanwhile, the Autoregressive Model (AR) and the Moving Average (MA) model are a combination of the ARMA model. Generally, the ARMA model form [17] :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \alpha_t - \theta_1 \alpha_{t-1} - \theta_2 \alpha_{t-2} - \cdots - \theta_q \alpha_{t-q} \quad (3)$$

$X_t \phi_p$Where is the value of the response variable with time t, is the value of the Autoregressive (AR) parameter and is the value of the Moving Average (MA) parameter.θ_q

To build a prediction model, a stationary time series is required. To remove non-stationery from a row, differencing is usually done (I). Sometimes, if the time series is more complex, more than one differencing operation (I) may be required. Therefore, the difference in value "d" value is the minimum amount of difference required to convert a time series into a stationary time series [18]. The value of 'd' will be 0, if the series is already stationary. However, if it is not stationary, eating will be done in the differencing stage (I) or what can be called 'd'. The ARIMA model has proven to be very useful in time series analysis because it provides a basic methodology for modeling the dependency effects of a data series and allows for valid statistical testing. Forecasting using the ARIMA model can be done with the formula [17] :

$$X_t(1 - B)(1 - \Phi_1 B) = \mu' + (1 - \theta_1 B)e_t \quad (4)$$

Where $X_t(1 - \Phi_1 B)(1 - \theta_1 B)$ as the value of the variable of the response period of the ket, as the value of AR, as the value of the Moving Average (MA), and as a forecasting error (error).e_t

2.4 Seasonal Auto-Regressive Integrated Moving Average

The Seasonal Autoregressive Integrated Moving Average which can be referred to as SARIMA is an extension of the ARIMA method with time series forecasting with variable (fluctuating) models with data in the form of trend and seasonal patterns [19]. Therefore, SARIMA is a suitable model for seasonal situations. The notation form of SARIMA is (p,d,q)$(P,D,Q)s$ where 'p' is a non-seasonal component Autoregressive (AR), 'd' Non-seasonal differencing order, 'q' Non-seasonal component of the Moving Average (MA), 'P' Value of seasonality Autoregressive (AR), 'D' Seasonal differencing Order, 'Q' Value of the seasonal Moving Average (MA), 's' number of periods per season [20]. The following is the general formula of SARIMA [17] :

$$X_t(1 - B)^d(1 - B^S) = (1 - B\Phi_1)(1 - \theta_1 B^S)e_t \quad (5)$$

Where the information in the formula above is as a non-seasonal differentiator, as a seasonal differentiator, as a non-seasonal Moving Average (MA), and () as a seasonal Moving Average (MA), as a residual term.$(1 - B)^d(1 - B^S)\ \theta_1(B)\ B^S e_t$

2.5 Root Mean Square Error

Root Mean Square Error is one of the commonly used evaluation metrics in statistics and data science to measure how close a forecasting or prediction model is to actual data. The RSME measures how much the difference between the predicted value and the actual observed value in the same unit. The smaller the RSME value, the better the model's ability to perform accurate forecasting.

Root Mean Square Error (RMSE) is the square root of the mean of the square difference between the calculation results of the tested method and the actual value. RMSE is used to compare the predictions given by a hypothetical model with the observed values. In simple terms, RMSE measures the degree of agreement between the actual data and the model's predictions. In the context of regression models, RMSE is one of the most commonly used evaluation metrics [21]. The RMSE calculation can be done using the following equation:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(Y_t - Y_t')^2}{n}} \quad (6)$$

Where:

$RMSE$ = *Root Mean Square Error*
Y_t = Actual data in period t
Y'_t = Data on forecast results in the period t
n = Total number of data periods

3. RESULT AND DISCUSSION

3.1 Dataset

The dataset used was taken from cement sales records at the Sinar Pagi Building Materials Store located in the Sidoarjo area. The data used was 7 months, from September 1, 2023, to March 31, 2024, with a total of 213 data rows. Here's an example of the data used:

Table 1. Datasets

Date	Cement Sales
09/01/2023	30
09/02/2023	15
09/03/2023	6
...	...
03/29/2023	33
03/30/2023	26
03/31/2023	17

The procedure for collecting cement sales data from the Sinar Pagi Building Materials Store is carried out through direct interviews with shop owners. Interviews are conducted periodically to ensure that the data obtained is accurate and reflects the daily sales that occur at the store. During the interview, the information collected includes the date of sale and the number of cement sales each day. With this approach, the data obtained is expected to provide an accurate picture of the sales pattern of cement in stores during a certain period.

The data collected consists of two main variables, namely the date and number of cement sales. The date variable records the day, month, and year of each sales transaction, while the cement sales variable records the number of cement units sold on each of those dates. The use of these two variables allows for a more in-depth analysis of cement sales trends, including daily, weekly, and seasonal fluctuations. This data will then be analyzed to identify sales patterns that can help in stock planning and future sales strategies.

3.2 Plotting Dataset

The data that has been collected is then visualized in the form of a plot to be able to see fluctuations in cement sales at the Sinar Pagi Building Materials Store.

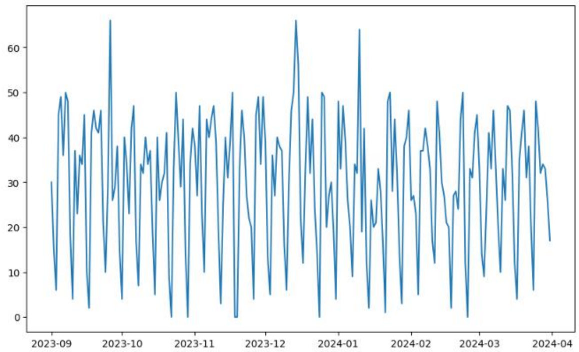


Figure 2. Daily Data on Cement Sales

It can be seen from Figure 1 that cement sales data fluctuates significantly every day.

3.3 Static Test on Dataset

To ensure that the cement sales data from the Sinar Pagi Building Materials Store is stationary, a stationery test is carried out using the Augmented Dickey-Fuller (ADF) method in Python. The ADF test is one of the commonly used methods to test the hypothesis of whether a time series of data has a root unit, which indicates that the data is not stationary. In this procedure, cement sales data that has gone through a seasonal differencing process is tested to see if there is a shift in the variance and mean data over time.

```
Results of Augmented Dickey-Fuller Test:
Test Statistic                 -4.619772
p-value                         0.000119
#Lags Used                     13.000000
Number of Observations Used   199.000000
Critical Value (1%)            -3.463645
Critical Value (5%)            -2.876176
Critical Value (10%)           -2.574572
dtype: float64
```

Figure 2. Stationary Test Results

The results of the Dickey-Fuller Augmented Test (ADF) in Figure 3 show that the data has met the stationery criteria. The p-value obtained is 0.000119, which is much smaller than the conventional significance level of 0.05. This further strengthens the conclusion that the data is stationary. Thus, based on the results of this ADF test, it can be concluded that the tested data does not contain the root unit and exhibits stationarity properties, which is an important prerequisite for time series analysis and further modeling.

To test the seasonal stationarity on cement sales data from Sinar Pagi Building Materials Store, an Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plot analysis was carried out using Python. This method helps in identifying correlation patterns in time series data. The ACF graph shows the degree of correlation between the values in the data and the previous lags, while the PACF graph helps identify significant lags after eliminating the effects of shorter lags.

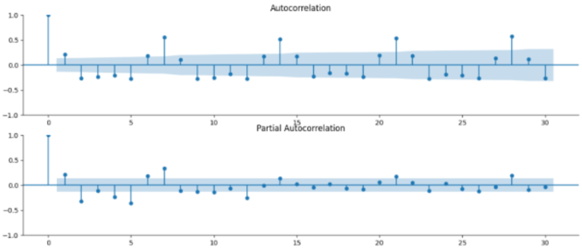


Figure 3. ACF and PACF plots

It can be seen that based on the ACF (Autocorrelation Function) plot produced, there is a clear pattern in certain lags. In particular, there was a significant increase in each 7-lag interval, indicating a weekly seasonal pattern in cement sales data. Meanwhile, in the PACF plot, there is a significant and irregular increase in lag, but the lag with the highest significance is in the 7th lag.

This pattern indicates that the sales value has a positive autocorrelation with the values that occurred in the previous seven-day time interval. This pattern of recurrence every 7 lags indicates the presence of a weekly seasonal component that needs to be considered in time series data modeling. Based on the ACF plot above, it shows that the dataset needs to be differentiated seasonally.

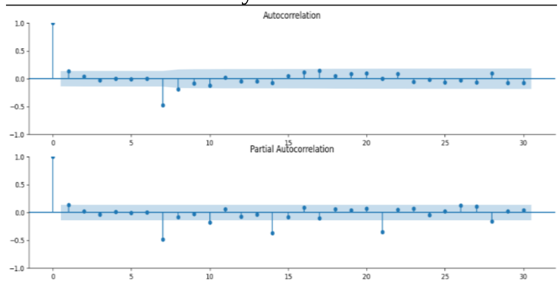
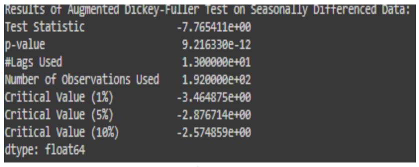


Figure 4. ACF and PACF Plots After Seasonal Differencing

In Figure 5, it can be seen that after the seasonal differencing stage with a seasonal value of 7, each lag has followed the significance limit. In the ACF plot, it can be seen that lag 7 has a significant increase, and after that, no more lag has a significant increase. In the PACF plot, it can be seen that 3 lags have increased to exceed the significance limit, namely at lags 7, 14, and 21 or it can be said that interval 7.



```
Results of Augmented Dickey-Fuller Test on Seasonally Differenced Data:
Test Statistic                -7.765411e+00
p-value                        9.216330e-12
#Lags Used                     1.300000e+01
Number of Observations Used    1.920000e+02
Critical Value (1%)           -3.464875e+00
Critical Value (5%)           -2.876714e+00
Critical Value (10%)          -2.574859e+00
dtype: float64
```

Figure 5 Stationary Test After going through the seasonal differencing stage

In Figure 6, it can be seen that the results of the Dickey-Fuller Augmented Test (ADF) on the data that have been carried out seasonal differencing show that the data has met the stationarity criteria. The p-value obtained was 9.6330e-12 or 0.000000000000921633, which is much smaller than the conventional significance level of 0.05 and much smaller than the p-value value of the initial dataset. This further strengthens the conclusion that the data is stationary. Thus, based on the results of the ADF test after going through this seasonal differencing stage, it can be concluded that the tested data does not contain the root unit and shows stationarity properties, which is an important prerequisite for time series analysis and further modeling using the SARIMA method.

3.4 Identification and Forecasting SARIMA Model

After the stationarity test and the previous differencing stage, the SARIMA modeling can be known by looking at the ACF and PACF plots. For the determination of the SARIMA model, it can be seen from the lags in ACF and PACF which have spikes beyond the limit of significance. In Figure 5 where the first most significant ACF lag is in the 7th lag and however the 7th lag is a seasonal lag, after that for the 1st lag has an increase that is between the significance limits, then it is found that for the order 'q' (MA) is 0 or 1. For 'd'(I) it is rated 0 because the initial data does not experience normal differencing and is already stationary according to Figure 3. and to determine the order 'p'(AR) is determined by observing the PACF plot in Figure 5 where the 7th lag is the most significant lag. However, the lag is a seasonal lag so it cannot be counted in the order 'p' (AR). As can be seen in Figure 5, the 1st lag in the PACF plot is a non-seasonal lag and has increased between the significance limits, from the observation above, it can be concluded that for the order 'p'(AR) is 1 or 0. Therefore, from the above observations, the ARIMA model that may be obtained is ARIMA with order (1,0,1), ARIMA with order (0,0,1), (1,0,0).

Furthermore, the values of the 'P' (SAR) and 'Q' (SMA) orders are determined at seasonal lags of 7, 14 and so on which are significant. In the ACF plot, it occurred significantly in the 7th lag. This means that the Q order (SMA) is valued at 1 because it uses a seasonal value of 7. Then for the P order (SAR), it is worth 3 because the 7th, 14th, and 21st seasonal lags are lags that exceed the significance limit. And for the order 'D' (Differencing) has a value of 1 because the dataset experiences one seasonal differencing. The value of 's' itself is 7 beuse the known seasonal pattern is a weekly pattern. From the above observations, it is possible for the SARIMA model, $(3,1,1)^7, (2,1,1)^7, (1,1,1)^7$.

After finding the possibility for several SARIMA models, after that a looping test was carried out to find the best model with the Akaike Information Criterion (AIC) parameter. AIC is a metric for assessing the quality of a model. The Akaike Information Criterion (AIC) is used to compare statistical models in terms of their quality and complexity. AIC helps select the best model from a set of different models by considering the trade-off between the model's suitability to the data and the complexity of the model. The lower the AIC value, the better the model [22]. The lower the AIC value, the better the model. The division of training data with test data is carried out with a ratio of 95:5 where cement sales data totaling 213 data will be divided into 202 training data and test data totaling 11. In this study, there will be 11 prediction data. Where the forecast data will be compared with the actual data or it can be called testing data.

Table 2. The best SARIMA models based on the lowest AIC

It	Type	AIC
1.	(1,0,0) $(3,1,1)^7$	1245
2.	(1,0,1) $(3,1,1)^7$	1246
3.	(0,0,1) $(3,1,1)^7$	1252

Based on Table 2, the 3 best models from the lowest AIC parameters are obtained, namely the SARIMA Model (1,0,0) $(3,1,1)^7$ (1,0,1)$(3,1,1)^7$, (0,0,1)$(3,1,1)^7$ .It can be seen that the SARIMA model (1,0,0)$(3,1,1)^7$ is the best model because it has the lowest AIC value of 1245. After obtaining the best model, the sales forecasting process was carried out with the best model obtained, namely the SARIMA (1,0,0) $(3,1,1)^7$model.

Table 3. Results of Cement Sales Data Forecasting

Date	Actual Cement Sales (Data Testing)	Forecast
03/21/2024	31	40
03/22/2024	38	36
03/23/2024	20	15
03/24/2024	6	3
03/25/2024	48	33
03/26/2024	42	34
03/27/2024	32	35
03/28/2024	34	41
03/29/2024	33	34
03/30/2024	26	15
03/31/2024	17	6
TOTAL	327	292

From Table 3, the forecast results using the SARIMA model (1,0,0)$(3,1,1)^7$ are obtained which is the best model based on the lowest AIC where there are 11 sales testing data from March 21, 2024 to March 31, 2024. The Date column explains the daily cement sales time, the Actual Cement Sales column is the testing data and the forecast column is the forecast result of the SARIMA model (1,0,0). It can be seen from Table 3 that the total testing data is worth 327 and the total forecast data is worth 292.$(3,1,1)^7$

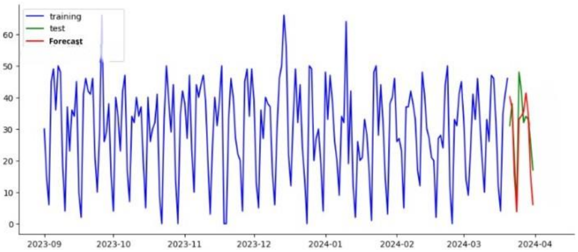


Figure 6 Graph of Forecast Results

It can be seen in Figure 6 where the training data graph is depicted in blue, the testing data graph is depicted in green, and the forecast result graph is depicted in red. In figure 6, the forecast results graph can be observed simply by following the testing data graph, this shows that the forecast of cement sales using the SARIMA model (1,0,0) $(3,1,1)^7$ shows quite good prediction results in following the pattern of cement sales data.

3.5 Model Evaluation

After the forecasting is carried out, it is continued with the evaluation of the model by determining the percentage of errors using the Root Mean Square Error (RMSE).

Table 4 Model Evaluation Using RMSE

It	Type	AIC	RMSE
1.	(1,0,0) $(3,1,1)^7$	1245,119671	7,95
2.	(1,0,1) $(3,1,1)^7$	1246,32888	8,05
3.	(0,0,1) $(3,1,1)^7$	1252,058219	8,15

From Table 4, with a 95:5 ratio data division, the SARIMA Model (1,0,0) $(3,1,1)^7$was successfully used to model cement sales data with an RMSE prediction error rate of 7.95 compared to the 2 models with the lowest AIC after that, namely the SARIMA model (1,0.1) $(3,1,1)^7$ with an RMSE value of 8.05 and the SARIMA model (0.0.1)$(3,1,1)^7$with an RMSE value of 8.15. This ensures that the model is trained with the majority of available data and tested with a small amount of data to evaluate its prediction performance.

The results of the evaluation show that the SARIMA model (1,0,0) $(3,1,1)^7$ is the best model for predicting sales. The model (1,0,0)$(3,1,1)^7$ has several parameters, namely (p,d,q) which is a non-seasonal component of the ARIMA model where p is the order of the autoregressive part (AR), d is the order of differencing to create stationary data, and q is the order of the moving average (MA) part. The seasonal components of the SARIMA model consist of $(P,D,Q)^m$where P is the order of the seasonal autoregressive part (SAR), D is the order of the seasonal differencing part, Q is the order of the seasonal moving average part, and m is the length of the seasonal period.

In this model, the non-seasonal components are p=1, d=0, and q=0, indicating that the model has one lag of the non-seasonal autoregressive part, no differencing applied to the non-seasonal component, and no non-seasonal moving average component. The seasonal component consists of P=3, D=1, Q=1, and m=7, which shows that the model has three lags of the seasonal autoregressive part, applies seasonal differencing once to create stationary seasonal data, has one lag of the seasonal moving average part, and the season repeats every 7 periods or can be called a weekly period. The SARIMA model was selected as the best model based on two evaluation metrics: Akaike Information Criterion (AIC) and Root Mean Square Error (RMSE).

The SARIMA model (1,0,0) $(3,1,1)^7$ shows the best balance between model complexity and prediction accuracy. A lower AIC value compared to other models shows that the model (1,0,0) $(3,1,1)^7$ is more efficient in capturing data patterns without becoming too complex, thus avoiding overfitting. In addition, a lower RMSE value indicates that this model has the least prediction error, which means that the predictions generated by this model are closer to the actual value of cement sales.

The evaluation results show that the model has a fairly good ability to predict cement sales, although there is room for further improvement to improve the accuracy of the predictions.

4. CONCLUSION

Based on the results of the research that has been carried out, the SARIMA model in forecasting the sales of Sinar Pagi building material stores with the notation SARIMA (p,d,q)(P,D,Q)^m has the best model value is (1,0,0) $(3,1,1)^7$ This model is the most suitable model for predicting cement sales at Sinar Pagi Building Material Stores. This is because the model (1,0,0) $(3,1,1)^7$ has the lowest AIC value and the smallest RMSE, with an AIC value of 1245 and an RMSE of 7.95 compared to other SARIMA models after model identification using model looping testing. These results show the best balance between model complexity and prediction accuracy. Lower AIC values indicate that the model is more efficient at capturing data patterns without becoming too complex, thus avoiding overfitting. In addition, a lower RMSE indicates that this model has the least prediction error, which means that the predictions generated by this model are closer to the actual value of cement sales. The model manages to capture seasonal patterns in sales data, providing a pretty good prediction framework. With this model, Sinar Pagi Building Materials Store can make more reliable sales predictions, which can help in inventory planning and marketing strategies.

5. REFERENCE

[1] 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.

[2] M. Romi, I. Apriliana, S. Wulandari, P. Studi, T. Industri, and U. M. Sidoarjo, "Masonry Concrete Sales Forecasting Using ARIMA and SARIMA Methods for Production Planning [Peramalan Penjualan Beton Masonry Menggunakan Metode ARIMA Dan SARIMA Untuk Perencanaan Produksi]," pp. 1–10, 1991.

[3] W. Gunawan and M. Ramadani, "Analisa Perbandingan Penerapan Metode SARIMA dan Prophet dalam Memprediksi Persediaan Barang PT XYZ," *Fakt. Exacta*, vol. 16, no. 2, pp. 88–97, 2023, doi: 10.30998/faktorexacta.v16i2.13803.

[4] W. Stephany, A. S. Albadry, A. Sofa, and Tarjo, "Analisis Pengendalian Persediaan Barang Dagang Dalam Menunjang Kelancaran Transaksi Jual Beli," *J. Ekopendia*, vol. 06, no. 1, pp. 171–193, 2021.

[5] M. Idhom and S. M. Huda, "879-2916-1-Pb," vol. XII, pp. 25–34, 2017.

[6] M. A. Firmansyah, A. Panji Sasmito, and H. Zulfia Zahro', "Aplikasi Forecasting Penjualan Bahan Bangunan Menggunakan Metode Trend Moment (Studi Kasus Di Ud. Hasil Bumi)," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 5, no. 2, pp. 526–533, 2021, doi: 10.36040/jati.v5i2.3759.

[7] E. S. Putri and M. Sadikin, "Prediksi Penjualan Produk Untuk Mengestimasi Kebutuhan Bahan Baku Menggunakan Perbandingan Algoritma LSTM dan ARIMA," *Format J. Ilm. Tek. Inform.*, vol. 10, no. 2, p. 162, 2021, doi: 10.22441/format.2021.v10.i2.007.

[8] I. L. Hakim, M. Sanglise, and C. D. Suhendra, "Analisis Peramalan Harga Telur Ayam Ras Dengan Menggunakan Metode SARIMA," vol. 8, no. April, pp. 966–977, 2024, doi: 10.30865/mib.v8i2.7610.

[9] N. E. Susanti, R. Saputra, and I. A. Situmorang, "Perbandingan Metode SARIMA, Double Exponential Smoothing dan Holt-Winter Additive dalam Peramalan Retail Sales Mobil Honda," *J. Sains Mat. dan Stat.*, vol. 10, no. 1, p. 58, 2024, doi: 10.24014/jsms.v10i1.26375.

[10] D. A. Fajari, M. F. Abyantara, and H. A. Lingga, "Peramalan Rata-Rata Harga Beras Pada Tingkat Perdagangan Besar Atau Grosir Indonesia Dengan Metode Sarima (Seasonal Arima)," *J. Agribisnis Terpadu*, vol. 14, no. 1, p. 88, 2021, doi: 10.33512/jat.v14i1.11460.

[11] A. S. Riyadi, I. P. Wardhani, and A. Perdana, "APLIKASI PERBANDINGAN PREDIKSI

HARGA BITCOIN MENGGUNAKAN DEEP LEARNING DENGAN METODE ARIMA , SARIMA , LTSM DAN GRADIENT," vol. 7, 2023.

[12] H. Chhabra, "A Comparative Study of ARIMA and SARIMA Models to Forecast Lockdowns due to SARS-CoV-2," *Adv. Tech. Biol. Med.*, vol. 11, no. 1, pp. 1–12, 2023, doi: 10.35248/2379-1764.23.11.399.

[13] S. Aktivani, "Uji Stasioneritas Data Inflasi Kota Padang," *Statistika*, vol. 20, no. 2, pp. 83–90, 2020.

[14] C. A. Melyani, A. Nurtsabita, G. Z. Shafa, and E. Widodo, "Peramalan Inflasi Di Indonesia Menggunakan Metode Autoregressive Moving Average (Arma)," *J. Math. Educ. Sci.*, vol. 4, no. 2, pp. 67–74, 2021, doi: 10.32665/james.v4i2.231.

[15] A. Juwanda *et al.*, "Analisa Prediksi Penjualan Mobil dengan Metode Autoregressive Integrated Moving Average (ARIMA)," *Semin. Nas. Mhs. Ilmu Komput. dan Apl. Jakarta-Indonesia*, no. September, pp. 1–7, 2021, [Online]. Available: https://conference.upnvj.ac.id/index.php/senamika/article/view/1787

[16] I. B. B. Mahayana, I. Mulyadi, and S. Soraya, "Peramalan Penjualan Helm dengan Metode ARIMA (Studi Kasus Bagus Store)," *Inferensi*, vol. 5, no. 1, p. 45, 2022, doi: 10.12962/j27213862.v5i1.12469.

[17] S. Putri and A. Sofro, "Peramalan Jumlah Keberangkatan Penumpang Pelayaran Dalam Negeri di Pelabuhan Tanjung Perak Menggunakan Metode ARIMA dan SARIMA," *MATHunesa J. Ilm. Mat.*, vol. 10, no. 1, pp. 61–67, 2022, doi: 10.26740/mathunesa.v10n1.p61-67.

[18] U. M. Sirisha, M. C. Belavagi, and G. Attigeri, "Profit Prediction Using ARIMA, SARIMA and LSTM Models in Time Series Forecasting: A Comparison," *IEEE Access*, vol. 10, no. December, pp. 124715–124727, 2022, doi: 10.1109/ACCESS.2022.3224938.

[19] N. S. Khoiriyah, M. Silfiani, R. Novelinda, and S. M. Rezki, "Peramalan Jumlah Penumpang Kapal di Pelabuhan Balikpapan dengan SARIMA," *J. Stat. dan Komputasi*, vol. 2, no. 2, pp. 76–82, 2023, doi: 10.32665/statkom.v2i2.2303.

[20] A. Agung Dwi Ramadhan and A. Fauzan, "Prediksi Nilai Ekspor Non Migas Di Jawa Barat Menggunakan Metode Seasonal Auto Regresif Integrated Moving Average (SARIMA)," *Emerg. Stat. Data Sci. J.*, vol. 1, no. 1, pp. 10–19, 2023, doi: 10.20885/esds.vol1.iss.1.art2.

[21] V. R. Prasetyo, H. Lazuardi, A. A. Mulyono, and C. Lauw, "Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression," *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 1, pp. 8–17, 2021, doi: 10.25077/teknosi.v7i1.2021.8-17.

[22] L. I. Harlyan, E. S. Yulianto, Y. Fitriani, and Sunardi, "Aplikasi Akaike Information Criterion (AIC) pada Perhitungan Efisiensi Teknis Perikanan Pukat Cincin di Tuban, Jawa Timur," *Mar. Fish. J. Mar. Fish. Technol. Manag.*, vol. 11, no. 2, pp. 181–188, 2021, doi: 10.29244/jmf.v11i2.38550.