BAREKENG: Journal of Mathematics and Its Application
December 2022 Volume 16 Issue 4 Page 1487–1496
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol16iss4pp1487-1496

COMPARISON IN PREDICTING THE SHORT-TERM USING THE SARIMA, DSARIMA, AND TSARIMA METHODS

Muhammad Giovani [1*], Indira Anggriani[2], Syalam Ali Wira Dinata Simatupang[3]

[1,2]Mathematics Study Program, Department of Mathematics and Information Technology, Kalimantan Institute of Technology
[3]Statistics Study Program, Department of Mathematics and Information Technology, Kalimantan Institute of Technology
Jl. Soekarno Hatta No.KM 15, Karang Joang, Balikpapan City, East Kalimantan 76127, Indonesia

Corresponding author's e-mail: [1] giovani.smat@gmail.com*

Abstract. *The flow of data and information grows quickly and rapidly in various sizes and means, called big data. In dealing with future changes, a mature data processing analysis and design are needed so that the prediction framework produces good results. One of the big data processing efforts is realized in the prediction or forecasting method, which is used to predict future values or trends as a reference for past conditions. One example of Big Data in Balikpapan City is the temperature within 2 meters obtained from the NASA satellite published on the power.larc.nasa.gov website. One of the methods used in this study is the ARIMA method, and development is carried out according to the data used. Based on the data to be used, namely temperature data within 2 meters in Balikpapan City, data processing is developed to pay attention to three seasonal patterns, or the Triple Seasonal ARIMA model. In this research, we can be seen how to build the Triple Seasonal ARIMA model, the Seasonal ARIMA model, and Double Seasonal ARIMA, and it can be seen the comparison of the prediction accuracy results of the three models. The results obtained in this study obtained a comparison of methods in making predictions with a specified period; the results obtained from the Seasonal ARIMA model showed that it was very good at predicting a period of 2 weeks, Double Seasonal ARIMA for a period of 1 month, Double Seasonal ARIMA for a period of 3 months, and Triple Seasonal ARIMA for a period of 6 months.*

Keywords: *ARIMA, big data, forecasting, seasonal.*

rticle info:

Submitted: 24th August 2022 *Accepted: 20th November 2022*

How to cite this article:
M. Giovani, I. Anggriani and S. A. W. D. Simatupang, "COMPARISON IN PREDICTING THE SHORT-TERM USING THE SARIMA, DSARIMA AND TSARIMA METHODS", *BAREKENG: J. Math. & App.*, vol. 16, iss. 4, pp. 1487-1496, Dec., 2022.

1. INTRODUCTION

The flow of data and information is growing rapidly in various sizes and means, called Big Data [11]. Big data has become a phenomenon among scientists and the business world. Currently, using Big Data in business is the most important thing. It supports solutions to business challenges in all aspects, including the manufacturing industry, technology, education, and libraries. Data can be considered big data if it has a factor of 3V (volume, velocity, variety) [2].

This big data phenomenon is increasingly interesting in a discussion regarding this data, not only covering how this data is located and found but also how this data can be utilized and useful for the benefit of human progress. One way that can help big data become useful information is data mining [1]. As a science that continues to develop, data mining can be utilized by all fields in aspects of life [12]. Data mining is a technique that can help predict a phenomenon of an object under study by being able to know in advance a certain class or object through data [17]. In facing a change for the better in the future, careful analysis and design of a data processing system are needed, in which a prediction or prediction framework can formulate the right policy to be one of the efforts to make a good decision.

One of the appropriate Big Data processing efforts, which can be realized through one method, prediction or forecasting, is to predict future values or trends concerning the analysis of a condition in future data. Ago [6]. One good method is ARIMA, which is a time series analysis method where ARIMA has very good accuracy for short-term predictions. In contrast, for long-term forecasting, the accuracy of predictions could be better [16]. According to Makridakis, seasonal data patterns are formed due to seasonal factors, such as weather and holidays. Seasonal data patterns are fluctuations in data that repeat every certain period, such as days, weeks, months, and even years [14]. So when looking at a seasonal pattern, you can use ARIMA development to process seasonal data using the Seasonal ARIMA (SARIMA) method, as well as double seasonal processing or related to recurring seasonal patterns such as days and weeks or weeks and years using the Double Seasonal ARIMA (DSARIMA) method. It was developed for modeling data that has seasonal and multiple seasonal patterns. Where in this study, the ARIMA method will be developed to process and predict the short-term time horizon of one of Big Data which has a seasonal pattern and the development of Seasonal ARIMA processing that takes into account the time pattern of up to three patterns, which will later be referred to as Triple Seasonal ARIMA, with the aim of to see the accuracy in processing and predicting Big Data when compared to other ARIMA alternative methods in processing seasonal patterns, namely the Seasonal ARIMA method and Double Seasonal ARIMA in predicting short-term data.

2. RESEARCH METHODS

2.1 Stationarity

A process which is a stochastic process is a stationary process if there is a constant mean and variance [5]. If the value of n observations from the time series is seen to have fluctuations in the mean and variance values are constant and independent of time, then the observation data can be said to be stationary. On the other hand, if the value of n observations in the time series does not fluctuate with respect to the mean and constant variance, it can be said that the observational data is not stationary [8].

2.2 Autoregressive Integrated Moving Average (ARIMA)

ARIMA is an acronym for Autoregressive Integrated Moving Average. ARIMA is an analytical method for making predictions based on observations of variable data behavior by ignoring an independent variable because the analysis pays attention to the past values of the dependent variable for predictions with good accuracy [9]. ARIMA model can be formed from the classification of three groups, autoregressive is a purposeful stochastic model in which the representation of a process as a finite number, a linear collection of past data from a process on an unpredictable event. The moving average, a representation of an observation at time t, is a linear combination of the number of random errors [3].

$$\phi_p(B)(1-B)^d \dot{Y}_t = \theta(B)e_t \tag{1}$$

p is a orde autoregressive, q is a orde moving average, d is a orde differencing.

2.3 Seasonal Autoregressive Integrated Moving Average (SARIMA)

The Seasonal Autoregressive Integrated Moving Average or also known as the SARIMA model can be applied to data processing that has a seasonal pattern. A seasonal pattern can be interpreted as an event that repeats itself over a certain period, where a data separated in a full season can show the same nature in the following seasons. The seasonal ARIMA model, better known as SARIMA, has a general form [15]

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D\dot{Y}_t = \theta_q(B)\Theta_Q(B^s)e_t \tag{2}$$

P is a orde seasonal autoregressive, Q is a orde seasonal moving average, D is a orde seasonal differencing.

2.4 Double Seasonal Autoregressive Integrated Moving Average (DSARIMA)

Double Seasonal Autoregressive Integrated Moving Average or commonly referred to as DSARIMA which can be said if it has two seasonal patterns, the DSARIMA model can be written as follows [4].

$$\phi_p(B)\Phi_{P_1}(B^{s_1})\varphi_{P_2}(B^{s_2})(1-B)^d(1-B^{s_1})^{D_1}(1-B^{s_2})^{D_2}\dot{Y}_t = \theta_q(B)\Theta_{Q_1}(B^{s_1})\tau_{Q_2}(B^{s_2})e_t \tag{3}$$

P_2 is a orde double seasonal autoregressive, Q_2 is a orde double seasonal moving average, D_2 is a orde double seasonal differencing.

2.5 Triple Seasonal Autoregressive Integrated Moving Average (TSARIMA)

Often in data processing, a time series data is encountered that has more than one seasonal pattern, such as short-term electricity consumption data. Triple Seasonal Autoregressive Integrated Moving Average or commonly referred to as TSARIMA which can be said if it has three seasonal patterns, in general from the development of the ARIMA, SARIMA, and DSARIMA models mathematically into the TSARIMA model can be written as follows [13]:

$$\begin{aligned}\phi_p(B)\Phi_{P_1}(B^{s_1})\varphi_{P_2}(B^{s_2})\Psi_{P_3}(B^{s_3})(1-B)^d(1-B^{s_1})^{D_1}(1-B^{s_2})^{D_2}(1-B^{s_3})^{D_3}Y_t \\ = \theta_q(B)\Theta_{Q_1}(B^{s_1})\tau_{Q_2}(B^{s_2})\Xi_{Q_3}(B^{s_3})e_t\end{aligned} \tag{4}$$

P_3 is a orde triple seasonal autoregressive, Q_3 is a orde triple seasonal moving average, D_3 is a orde triple seasonal differencing.

2.6 Best Model Selection

The accuracy of measuring methods in predicting is one of the important things. This is because an accuracy method is useful in providing an evaluation of the results of predictions that have been made. Therefore, a prediction method must have an error. The smaller the error rate, the more accurate the prediction will be. In choosing the best prediction model, it is chosen by looking at the level of error in predicting. However, if there is more than one suitable model, the smallest error rate value can be compared [10].

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{\left|Y_t - \hat{Y}_t\right|}{Y_t} \times 100\% \tag{5}$$

The use of MAPE in outcome evaluation can avoid measurement accuracy of the magnitude of the actual value and the predicted value. MAPE score criteria shown in Table 1 [7].

Table 1. MAPE score criteria

MAPE (x)	Criteria
$x < 10\%$	Very good forecasting ability
$10\% \geq x > 20\%$	Good forecasting ability
$20 \geq x > 50\%$	Forecasting ability is sufficient
$x \geq 50\%$	Poor forecasting ability

3. RESULTS AND DISCUSSION

3.1. Data Set and Results

The process with the Box-Jenkins procedure for modeling a Triple Seasonal ARIMA model and other alternative models, namely Seasonal ARIMA and Double Seasonal ARIMA, begins with identifying the stationarity of the data in mean and variance. This identification used sample temperature data within 2 meters in Balikpapan City from 01 January 2001 to 31 August 2021. The stationarity of the data in variance can be identified through the time series plot that will be visualized. Suppose the time series plot that has been visualized shows that the data distribution is not constant around the mean or average. In that case, it can indicate that it is not stationary concerning the variance, so a transformation is needed.

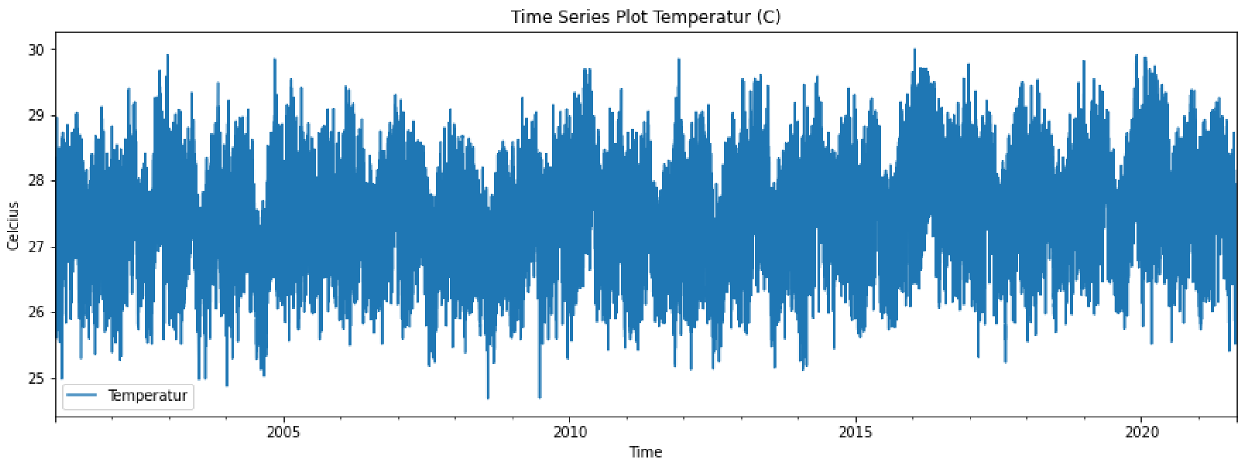

Figure 1. Plot time series Temperature Balikpapan City.

Figure 1, shows that the distribution of data is not constant with respect to the mean or average which can be said to be not stationary with respect to variance and a transformation process is required. The transformation process used, namely the Box-Cox transformation, is carried out by an appropriate in-sample data transformation process. Furthermore, the process of determining the value of is carried out in order to find out the form of transformation used. The Maximum Likelihood process is carried out to determine the estimated value of λ, the λ value is -0.05.

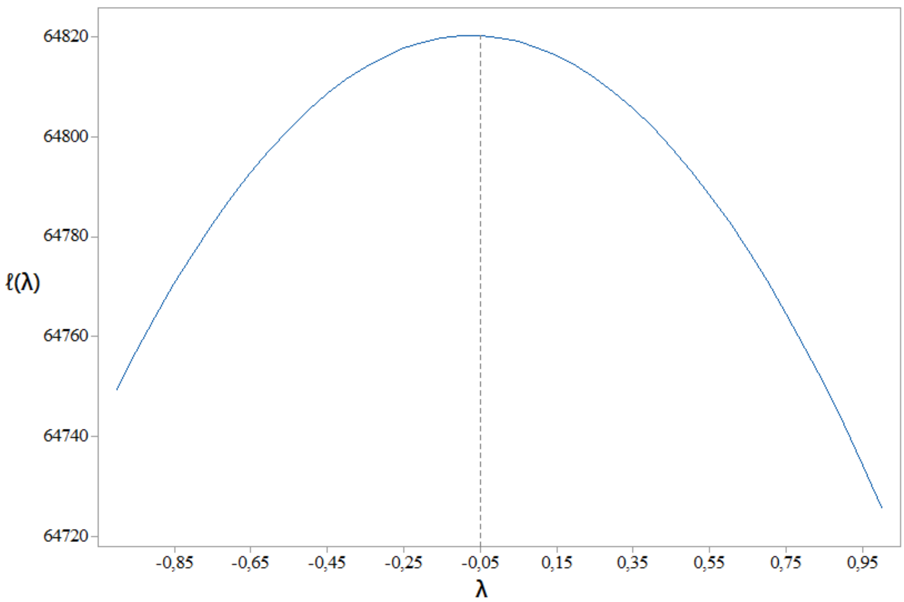

Figure 2. Estimate Value λ.

After obtaining the value of $\lambda = -0.05$. Therefore, a transformation process is carried out in which the process is $Y_t^{-0.05}$, for Y_t it is known that it is a data in sample. In the next stage, it will be seen that the stationary data in the mean or average can be identified through the ACF plot. The ACF plot which has a very slow descending pattern indicates that the data is not stationary in the mean, so differencing is necessary.

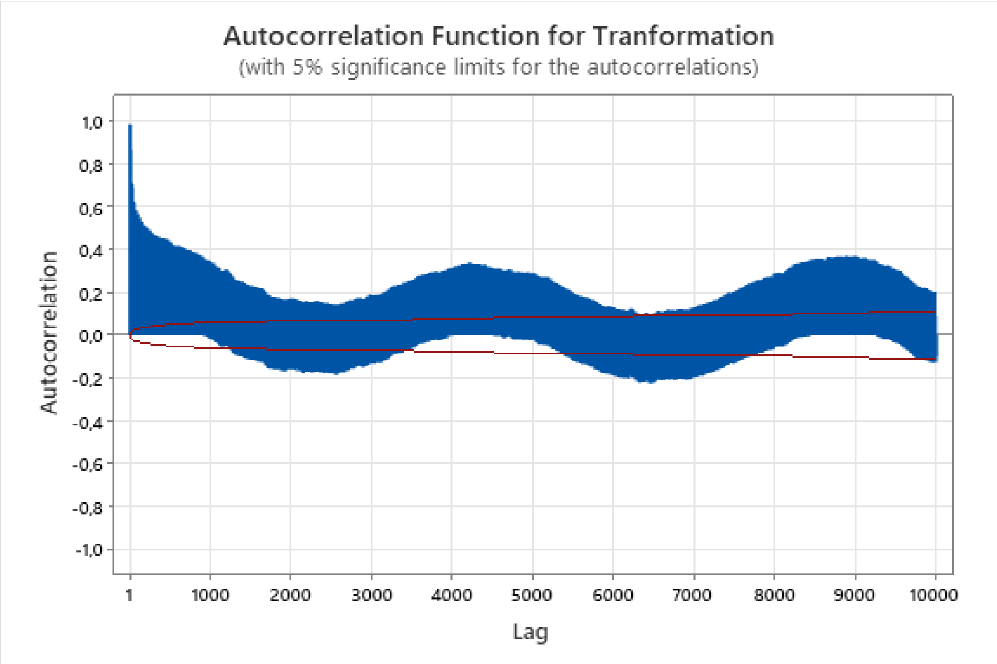

Figure 3. ACF Plot Temperature in 2 meters per hour.

Based on Figure 3, it shows that the data pattern is not stationary because the ACF plot still has a dying down pattern which indicates that the data is still not stationary with respect to the mean or average, so a differencing process is needed for temperature data within 2 meters in Balikpapan City. In the next stage where the data has been differencing, the ACF and PACF plots are re-formed to see that they are stationary with respect to the mean for the ACF and PACF formation processes.

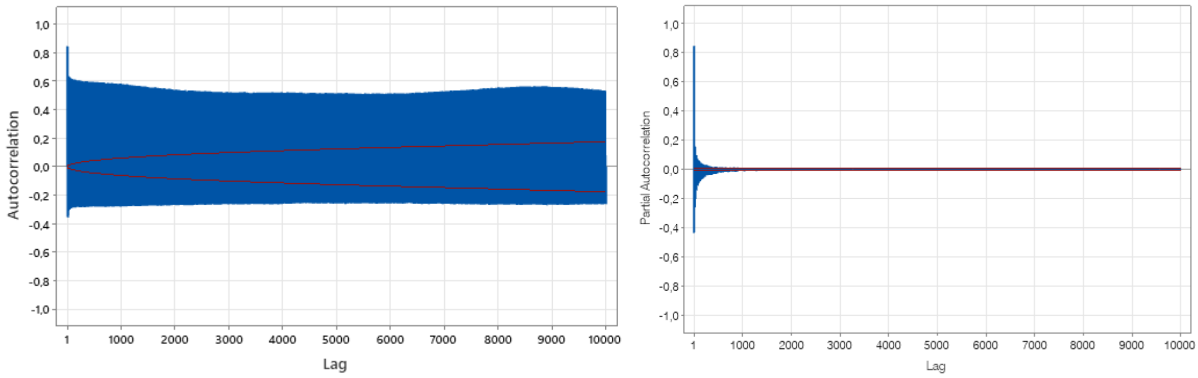

Figure 4. ACF and PACF plot Temperature in Balikpapan City.

Based on Figure 4 the ACF value has significance at lag 1 which can be said to have the model assumption of $MA(q) = 1$ and the PACF value has significance at lag 1 which can be said to have the model assumption of AR(p)=1. Furthermore, it can be assumed that regular ARIMA models are $(0,1,1), (1,1,0)$, and $(1,1,1)$. In the identification of models that have seasonal patterns in the PACF plot, it looks significant at lag 24, lag 168, and lag 8736. This means that the data that has been differencing in lag 1 with the data used, namely Temperature within 2 meters in Balikpapan City per hour has seasonal periods daily or 24 hours, weekly or $24 \times 7 = 168$ hours, and annual 52 weeks or $168 \times 52 = 8736$ hours. Based on Figure 5 and Figure 6, it can be said that the temperature data within 2 meters in Balikpapan City has a daily, weekly and annual seasonal pattern, which will then make some assumptions which are then used in the Box-Jenkins process in the model development stage.

First based on Single SARIMA. Based on Figure 4, the ACF plot is significant at lag 24, 168 and 8736, which has the assumption that the model $MA(Q_1) = 1$ for lag 24, lag 168 and lag 8736 so that we obtain an alleged model, namely:

Table 2. Assumption of SARIMA Model Development on Diff Lag 1.

SARIMA Model Assumption	
$(1,1,1)(0,0,1)^{24}$	$(1,1,1)(0,0,1)^{8736}$
$(1,1,0)(0,0,1)^{24}$	$(1,1,0)(0,0,1)^{8736}$
$(0,1,1)(0,0,1)^{24}$	$(0,1,1)(0,0,1)^{8736}$
$(1,1,1)(0,0,1)^{168}$	–
$(1,1,0)(0,0,1)^{168}$	–
$(0,1,1)(0,0,1)^{168}$	–

Next based on Figure 4, the ACF plot is significant at lag 24, 168 and 8736, which has the assumption that the model $MA(Q_2)$=1 for lag 24, lag 168 and lag 8736 so that we obtain an alleged model, namely:

Table 3. Assumption of DSARIMA Model Development on Diff Lag 1.

DSARIMA Model Assumption	
$(1,1,1)(0,0,1)^{24}(0,0,1)^{168}$	$(1,1,1)(0,0,1)^{168}(0,0,1)^{8736}$
$(1,1,0)(0,0,1)^{24}(0,0,1)^{168}$	$(1,1,0)(0,0,1)^{168}(0,0,1)^{8736}$
$(0,1,1)(0,0,1)^{24}(0,0,1)^{168}$	$(0,1,1)(0,0,1)^{168}(0,0,1)^{8736}$
$(1,1,1)(0,0,1)^{24}(0,0,1)^{8736}$	–
$(1,1,0)(0,0,1)^{24}(0,0,1)^{8736}$	–
$(0,1,1)(0,0,1)^{24}(0,0,1)^{8736}$	–

Based on Figure 4, the ACF plot is significant at lag 24, 168 and 8736, which has the assumption that the model $MA(Q_3) = 1$ for lag 8736 so that we obtain an alleged model, namely:

Table 4. Assumption of TSARIMA Model Development on Diff Lag g 1

TSARIMA Model Assumption
$(1,1,1)(0,0,1)^{24}(0,0,1)^{168}(0,0,1)^{8736}$
$(1,1,0)(0,0,1)^{24}(0,0,1)^{168}(0,0,1)^{8736}$
$(0,1,1)(0,0,1)^{24}(0,0,1)^{168}(0,0,1)^{8736}$

By using each model built from SARIMA, DSARIMA, and TSARIMA, then selecting the best model used for prediction by determining the smallest MAPE of each, next are the results of the calculation stages of the RMSE, sMAPE, and MAPE values with a time span of 2 weeks which are shown in Table 5.

Table 5. MAPE Values from 2 Weeks Range Prediction

Prediction	MAPE
$(0,1,1)(0,0,1)^{24}$	1,780%
$(1,1,0)(0,0,1)^{24}$	5,707%
$(1,1,1)(0,0,1)^{24}$	2,390%
$(0,1,1)(0,0,1)^{168}$	1,919%
$(1,1,0)(0,0,1)^{168}$	4,794%
$(1,1,1)(0,0,1)^{168}$	2,347%
$(0,1,1)(0,0,1)^{8736}$	1,823%
$(1,1,0)(0,0,1)^{8736}$	4,199%
$(1,1,1)(0,0,1)^{8736}$	2,394%
$(0,1,1)(0,0,1)^{24}(0,0,1)^{168}$	1,944%

Prediction	MAPE
$(1,1,0)(0,0,1)^{24}(0,0,1)^{168}$	4,953%
$(1,1,1)(0,0,1)^{24}(0,0,1)^{168}$	3,400%
$(0,1,1)(0,0,1)^{24}(0,0,1)^{8736}$	1,851%
$(1,1,0)(0,0,1)^{24}(0,0,1)^{8736}$	4,518%
$(1,1,1)(0,0,1)^{24}(0,0,1)^{8736}$	3,331%
$(0,1,1)(0,0,1)^{168}(0,0,1)^{8736}$	2,055%
$(1,1,0)(0,0,1)^{168}(0,0,1)^{8736}$	3,729%
$(1,1,1)(0,0,1)^{168}(0,0,1)^{8736}$	3,131%
$(0,1,1)(0,0,1)^{24}(0,0,1)^{168}(0,0,1)^{8736}$	2,083%
$(1,1,0)(0,0,1)^{24}(0,0,1)^{168}(0,0,1)^{8736}$	3,536%
$(1,1,1)(0,0,1)^{24}(0,0,1)^{168}(0,0,1)^{8736}$	2,083%

Based on Table 5, it can be seen that the best value for predicting temperature data in Balikpapan City with a span of 2 weeks, the best model can be chosen, namely the Single SARIMA model $(0,1,1)(0,0,1)^{24}$ with the equation $\dot{Y}_t = \dot{Y}_{t-1} + e_t + 0,76133_1 e_{t-1} + 0,37292_1 e_{t-24} + 0,2839151 e_{t-25}$ MAPE values 1,780% because the MAPE values are better than the model values compared to the DSARIMA model and the TSARIMA model. Next are the results of the calculation stages of MAPE values with a span of 1 months which are shown in Table 6.

Table 6. MAPE Values from 1 Month Range Prediction

Prediction	MAPE
$(0,1,1)(0,0,1)^{24}$	2,249%
$(1,1,0)(0,0,1)^{24}$	6,538%
$(1,1,1)(0,0,1)^{24}$	2,397%
$(0,1,1)(0,0,1)^{168}$	2,237%
$(1,1,0)(0,0,1)^{168}$	5,851%
$(1,1,1)(0,0,1)^{168}$	2,385%
$(0,1,1)(0,0,1)^{8736}$	2,253%
$(1,1,0)(0,0,1)^{8736}$	5,022%
$(1,1,1)(0,0,1)^{8736}$	2,396%
$(0,1,1)(0,0,1)^{24}(0,0,1)^{168}$	2,231%
$(1,1,0)(0,0,1)^{24}(0,0,1)^{168}$	6,002%
$(1,1,1)(0,0,1)^{24}(0,0,1)^{168}$	2,999%
$(0,1,1)(0,0,1)^{24}(0,0,1)^{8736}$	2,240%
$(1,1,0)(0,0,1)^{24}(0,0,1)^{8736}$	2,917%
$(1,1,1)(0,0,1)^{24}(0,0,1)^{8736}$	1,851%
$(0,1,1)(0,0,1)^{168}(0,0,1)^{8736}$	2,269%
$(1,1,0)(0,0,1)^{168}(0,0,1)^{8736}$	4,691%
$(1,1,1)(0,0,1)^{168}(0,0,1)^{8736}$	2,792%
$(0,1,1)(0,0,1)^{24}(0,0,1)^{168}(0,0,1)^{8736}$	2,273%
$(1,1,0)(0,0,1)^{24}(0,0,1)^{168}(0,0,1)^{8736}$	6,105%
$(1,1,1)(0,0,1)^{24}(0,0,1)^{168}(0,0,1)^{8736}$	2,827%

Based on Table 6, it can be seen that the best value for predicting temperature data in Balikpapan City with a span of 2 weeks, the best model can be chosen, namely the DSARIMA model $(1,1,1)(0,0,1)^{24}(0,0,1)^{8736}$ with the equation $\dot{Y}_t = 0,71611\dot{Y}_{t-1} + e_t + 0,38164 e_{t-1} \mp 0,24442 e_{t-24} + 0,093280448 e_{t-25} + 0,21311 e_{t-8736} + 0,081331 e_{t-8737} + 0,05208 e_{t-8760} + 0,01987 e_{t-8761}$ MAPE values 1,851% because the MAPE values are better than the model values compared to the Single SARIMA model and the TSARIMA model. Next are the results of the calculation stages of MAPE values with a span of 3 months which are shown in Table 7.

Table 7. MAPE Values from 3 Months Range Prediction

Prediction	MAPE
$(0,1,1)(0,0,1)^{24}$	2,754%
$(1,1,0)(0,0,1)^{24}$	8,242%
$(1,1,1)(0,0,1)^{24}$	2,244%
$(0,1,1)(0,0,1)^{168}$	2,518%
$(1,1,0)(0,0,1)^{168}$	7,678%

Prediction	MAPE
$(1,1,1)(0,0,1)^{168}$	2,302%
$(0,1,1)(0,0,1)^{8736}$	2,495%
$(1,1,0)(0,0,1)^{8736}$	6,569%
$(1,1,1)(0,0,1)^{8736}$	2,312%
$(0,1,1)(0,0,1)^{24}(0,0,1)^{168}$	2,437%
$(1,1,0)(0,0,1)^{24}(0,0,1)^{168}$	7,828%
$(1,1,1)(0,0,1)^{24}(0,0,1)^{168}$	2,092%
$(0,1,1)(0,0,1)^{24}(0,0,1)^{8736}$	2,437%
$(1,1,0)(0,0,1)^{24}(0,0,1)^{8736}$	6,928%
$(1,1,1)(0,0,1)^{24}(0,0,1)^{8736}$	2,125%
$(0,1,1)(0,0,1)^{168}(0,0,1)^{8736}$	2,166%
$(1,1,0)(0,0,1)^{168}(0,0,1)^{8736}$	6,313%
$(1,1,1)(0,0,1)^{168}(0,0,1)^{8736}$	2,111%
$(0,1,1)(0,0,1)^{24}(0,0,1)^{168}(0,0,1)^{8736}$	2,159%
$(1,1,0)(0,0,1)^{24}(0,0,1)^{168}(0,0,1)^{8736}$	6,105%
$(1,1,1)(0,0,1)^{24}(0,0,1)^{168}(0,0,1)^{8736}$	2,102%

Based on Table 7, it can be seen that the best value for predicting temperature data in Balikpapan City with a span of 2 weeks, the best model can be chosen, namely the DSARIMA model $(1,1,1)(0,0,1)^{24}(0,0,1)^{168}$ with the equation $\dot{Y}_t = e_t + 0{,}71790\dot{Y}_{t-1} + 0{,}38529e_{t-1} + 0{,}22331e_{t-168} + 0{,}0860391e_{t-169} + 0{,}21649e_{t-8736} + 0{,}083411e_{t-8737} + 0{,}04834e_{t-8904} + 0{,}34706e_{t-8905}$ get MAPE values 2,092% because the MAPE values are better than the model values compared to the Single SARIMA model and the TSARIMA model. Next are the results of the calculation stages of MAPE values with a span of 6 months which are shown in Table 8.

Table 8. MAPE Values from 6 Months Range Prediction

Prediction	MAPE
$(0,1,1)(0,0,1)^{24}$	2,754%
$(1,1,0)(0,0,1)^{24}$	8,242%
$(1,1,1)(0,0,1)^{24}$	2,244%
$(0,1,1)(0,0,1)^{168}$	2,518%
$(1,1,0)(0,0,1)^{168}$	7,678%
$(1,1,1)(0,0,1)^{168}$	2,302%
$(0,1,1)(0,0,1)^{8736}$	2,495%
$(1,1,0)(0,0,1)^{8736}$	6,569%
$(1,1,1)(0,0,1)^{8736}$	2,312%
$(0,1,1)(0,0,1)^{24}(0,0,1)^{168}$	2,437%
$(1,1,0)(0,0,1)^{24}(0,0,1)^{168}$	7,828%
$(1,1,1)(0,0,1)^{24}(0,0,1)^{168}$	2,092%
$(0,1,1)(0,0,1)^{24}(0,0,1)^{8736}$	2,437%
$(1,1,0)(0,0,1)^{24}(0,0,1)^{8736}$	6,928%
$(1,1,1)(0,0,1)^{24}(0,0,1)^{8736}$	2,125%
$(0,1,1)(0,0,1)^{168}(0,0,1)^{8736}$	2,166%
$(1,1,0)(0,0,1)^{168}(0,0,1)^{8736}$	6,313%
$(1,1,1)(0,0,1)^{168}(0,0,1)^{8736}$	2,111%
$(0,1,1)(0,0,1)^{24}(0,0,1)^{168}(0,0,1)^{8736}$	2,159%
$(1,1,0)(0,0,1)^{24}(0,0,1)^{168}(0,0,1)^{8736}$	6,105%
$(1,1,1)(0,0,1)^{24}(0,0,1)^{168}(0,0,1)^{8736}$	2,102%

Based on Table 8, it can be seen that the best value for predicting temperature data in Balikpapan City with a span of 2 weeks, the best model can be chosen, namely the DSARIMA model $(0,1,1)(0,0,1)^{24}(0,0,1)^{168}(0,0,1)^{8736}$ with the equation $\dot{Y}_t = e_t + 0{,}3756e_{t-1} + 0{,}21151e_{t-24} + 0{,}079455e_{t-25} + 0{,}18457e_{t-168} + 14112e_{t-169} + 0{,}039040e_{t-192} + 0{,}0146659e_{t-193} + 0{,}17937e_{t-8736} + 0{,}0673821e_{t-8737} + 0{,}037938e_{t-8760} + 0{,}01425199e_{t-8761} + 0{,}03310e_{t-8904} + 0{,}01243e_{t-8905} + 0{,}00700e_{t-8928} + 0{,}00263e_{t-8929}$ get MAPE values 2,057% because MAPE values are better than the model values compared to the Single SARIMA model and the TSARIMA model.

4. CONCLUSIONS

The short-term prediction accuracy results from the TSARIMA model and the alternative models of Single SARIMA and DSARIMA are as follows:

1) The best value for predicting temperature data within 2 meters in Balikpapan City with a span of 2 weeks is the Single SARIMA model $(0,1,1)(0,0,1)^{24}$ with an MAPE value is 1.780%.

2) The best value for predicting temperature data within 2 meters in Balikpapan City with a span of 1 month, is the DDSARIMA $(1,1,1)(0,0,1)^{24}(0,0,1)^{8736}$ model with a value of MAPE value 1.851%.

3) The best value for predicting temperature data within 2 meters in Balikpapan City with a span of 3 months is the DSARIMA $(1,1,1)(0,0,1)^{24}(0,0,1)^{168}$ model with a value of MAPE value 2.092%.

4) The best value for predicting temperature data within 2 meters in Balikpapan City with a span of 6 months is the TSARIMA model $(0,1,1)(0,0,1)^{24}(0,0,1)^{168}(0\ ,0,1)^{8736}$ with an MAPE value of 2.057%.

REFERENCES

[1] A. P. Natasuwarna. (2018), Prediksi Pengaruh Medsos terhadap Pilihan Mahasiswa pada Pilkada Serentak Menggunakan Data Mining, Sensitek, 12, July, 2018. Vol.1, No. 1, pp. 98-103.

[2] Ali, I. (2015). Big Data: Apa dan pengaruhnya pada perpustakaan?(What is Big Data and its Influence to Library). Media Pustakawan, 22(4), 15-19.

[3] Aswi Dan Sukarna. (2006). Analisis Deret Waktu: Teori Dan Aplikasi. Makasar: Andira Publisher.

[4] Azka, M., Wiradinata, S. A., & Faisal, M. (2020, May). Double Seasonal ARIMA for Forecasting Electricity Demand of Kuaro Main Gate in East Kalimantan. *In IOP Conference Series: Materials Science and Engineering* (Vol. 846, No. 1, p. 012064). IOP Publishing.

[5] Bowerman, Bruce L. And O'connell Richard T. Forecasting And Time Series: An Applied Approach, Third Edition. Duxbury Press, 1993.

[6] Elvierayani, R. R. (2017), "Prediksi Nilai Tukar (Kurs) Rupiah Terhadap Dolar Tahun 2017 dengan Menggunakan Metode Arima Box-Jenkins", 254.

[7] Fitria, I., Handayati, K. N., & Hasanah, P. (2021, March). The application of Geometric Brownian Motion in stock forecasting during the coronavirus outbreak in Indonesia. In Journal of Physics: Conference Series (Vol. 1821, No. 1, p. 012008). IOP Publishing.

[8] Katabba, Y. I. (2021). "Metode Seasonal Autoregressive Integrated Moving Average (SARIMA) Untuk Memprediksi Jumlah Penumpang Kereta Api Di Pulau Sumatera".

[9] Ningrum, K. P. (2021), Prediksi nilai impor alat telekomunikasi dan elektronik indonesia menggunkan model ARIMA-GARCH.

[10] Susanti, L., Hasanah, P., & Winarni, W. (2020). Peramalan Suhu Udara dan Dampaknya terhadap Konsumsi Energi Listrik di Kalimantan Timur. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, 14(3), 399-412.

[11] Sugiarsono, J. (2015). SWA sembada Media Bisnis.

[12] Suyanto. (2017), "Data Mining untuk Klasifikasi dan Klasterisasi Data", Informatika.

[13] Dinata, S. A. W., Azka, M., Hasanah, P., Suhartono, S., & Gamal, M. D. H. (2021). Comparison of Short-Term Load Forecasting Based on Kalimantan Data. Indonesian Journal of Statistics and Its Applications, 5(2), 243-259.

[14] Vuzi Fandiyah, Laylil. (2020), Peramalan Penjualan Beras Pada Ud. Elang Perkasa Ponorogo Sebagai Dasar Penyusunan Anggaran Komprehensif Tahun 2020. Skripsi (S1) thesis, Universitas Muhammadiyah Ponorogo.

[15] W. W. S. Wei. (2006), Time Series Analysis Univariate and Multivariate Methods. SECOND EDITION. 2006.

[16] Wibowo, H., Mulyadi, Y., & Abdullah, A. G. (2012). "Peramalan Beban Listrik Jangka Pendek Terklasifikasi Berbasis Metode Autoregressive Integrated Moving Average". Electrans, 11(2), 44–50

[17] Y.S. Nugroho. (2014), "Data Mining Menggunakan Algoritma Naive Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro", Dian Nuswantoro Fakultas Ilmu Komputer, November.