International Seminar of Research Month 2021
Volume 2022
http://dx.doi.org/10.11594/nstp.2022.2428

NST PROCEEDINGS

Conference Paper

Weather Forecast Based on Daily Temperature Data Using Seasonal Auto-regressive Integrated Moving Average (SARIMA) Method: Surabaya – BMKG Juanda

Prismahardi Aji Riyantoko*

Department of Data Science, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Surabaya 60294, Indonesia

Corresponding author:
E-mail:
prismahardi.aji.ds@upnjatim.ac.id

ABSTRACT

Indonesia is a country that has two seasons in one period, namely the dry season and the rainy season. Problems often arise in the field of climatology, is the presence of missing data, which affects the quality of the prediction results. This study uses time-series results, which show the characteristics of statistical values with observed data and behave according to the probabilistic concept. In this research case study, we took data from Kaggle which contained Global Climate Change data for the Surabaya area from 1900 to 2012. In addition, we took BMKG online data from 2013 to 21 April 2021 based on observations from the Juanda Meteorological station for the Surabaya area, and surrounding. Specifically for this research, we focus on the rainfall temperature which is modeled using the ARIMA Seasonal model, to predict the weather in the Surabaya area and its surroundings. Therefore, we want to analyze the weather prediction using the ARIMA Seasonal time series model for a case study in the Surabaya area. It can be simplified that the temperature prediction in the Surabaya and surrounding areas has an accurate value with the selection of the best model SARIMA (3,0,0) (0,1,1,12).

Keywords: Forecast, temperature, SARIMA, daily

Introduction

Indonesia is a country that has two seasons in one period, namely the dry season and the rainy season. Seasonal changes are strongly influenced by several variables including changes in wind direction, rainfall, wind speed, humidity, duration of sunshine, and temperature of rainfall. Rainfall temperature is one of the common variables that is often used to study climate variability. It is the result of the interaction of various physical phenomena and is characterized by spatial and temporal variations. Thus, temperature analysis of rainfall data is very important for information prediction, planning, and management of meteorological data for water resources systems (Dastorani et al., 2016).

Problems often arise in the field of climatology, is the presence of missing data, which affects the quality of the prediction results. This case study requires a complete and accurate recording of rainfall temperature data. Therefore, several methods are needed to solve the problem of accuracy of the results, including the time series method on the stochastic model (Venkata et al., 2013; Li et al., 2017; Radhakrishnan & Dinesh, 2006; Wang et al., 2013). This study uses time-series results, which show the characteristics of statistical values with observed data and behave according to the probabilistic concept. Thus, the observed sequence is only one of the possible achievements of the stochastic process. Thus, an appropriate scenario is needed to observe the results of the time series. For example, the Autoregressive Moving Average Model (ARMA) is a

How to cite:
Riyantoko, P. A. (2022). Weather forecast based on daily temperature data using Seasonal Auto-regressive Integrated Moving Average (SARIMA) Method: Surabaya – BMKG Juanda. *International Seminar of Research Month 2021*. NST Proceedings. pages 170-175. doi: 10.11594/ nstp.2022.2428

stochastic model based on probabilistic theory, so it can represent data uncertainty (Chatfield & Xing, 2019). In addition, there are also several variations of autoregressive models, such as the Periodic Autoregressive Moving Average (PARMA) model; ARMA model with Auxiliary input, ARMAX; Autoregressive Integrated Moving Average (ARIMA) model; and the ARIMA model with a seasonal component or the Seasonal Auto-Regressive Integrative Moving Average (SARIMA), which allows users to generate synthetic time series taking into account cyclical variations in the observed series records (Chatfield & Xing, 2019).

In this research case study, we took data from Kaggle which contained Global Climate Change data for the Surabaya area from 1900 to 2012. In addition, we took BMKG online data from 2013 to 21 April 2021 based on observations from the Juanda Meteorological station for the Surabaya area, and surrounding. Some of the parameters in the data we take include changes in wind direction, rainfall, wind speed, humidity, duration of sunshine, and temperature of rainfall. Specifically for this research, we focus on the rainfall temperature which is modeled using the ARIMA Seasonal model, to predict the weather in the Surabaya area and its surroundings. Therefore, we want to analyze the weather prediction using the ARIMA Seasonal time series model for a case study in the Surabaya area.

Material and Methods

Literature review

For the literature review subsection, we took several previous references about solving time series problems using the ARIMA Seasonal model. Here is some related research including:

Luisa Martinez-Acosta, Juan Pablo Medrano-Barboza, Alvaro Lopes-Ramos, John Freddy Remolina Lopez, and Alvaro Alberto Lopes-Lambrano use the ARIMA Seasonal model to predict rainfall based on monthly data (Acosta et al., 2020). In this study, the best SARIMA model was obtained based on the autocorrelation function (ACF), partial autocorrelation function (PACF), and the minimum Akaike Information Criterion (AIC) value. The results obtained in this study are used to make policies and decisions to determine strategies, priorities, and appropriate use in managing Water Resources in the Sinu River Colombia.

Peng Chen, Aichen Niu, Duaanyang Liu, Wei Jiang, and Bin Ma conducted a study on temperature forecasting using the SARIMA model for a case study in Nanjing, China (Chen et al., 2018). They analyzed monthly average temperature data from 1951 to 2017. The results showed that the predictions were very accurate.

Methods

In completing this research, we divide it into two very important things, including data collection and data prediction using the SARIMA model. We took data on rainfall temperatures from 1900 to 2012 from Kaggle, while data from 2013 to 21 April 2021, we took data from the online BMKG website. The data is specific to the Surabaya area and its surroundings because we took the data based on the classification of the Juanda Meteorological Station. To make predictions, we use the ARIMA Seasonal model by analyzing the autocorrelation function (ACF), partial autocorrelation function (PACF), and RMSE.

At the research methodology stage, it is necessary to carry out research planning to describe the steps to be carried out in the form of a chart or roadmap as follows

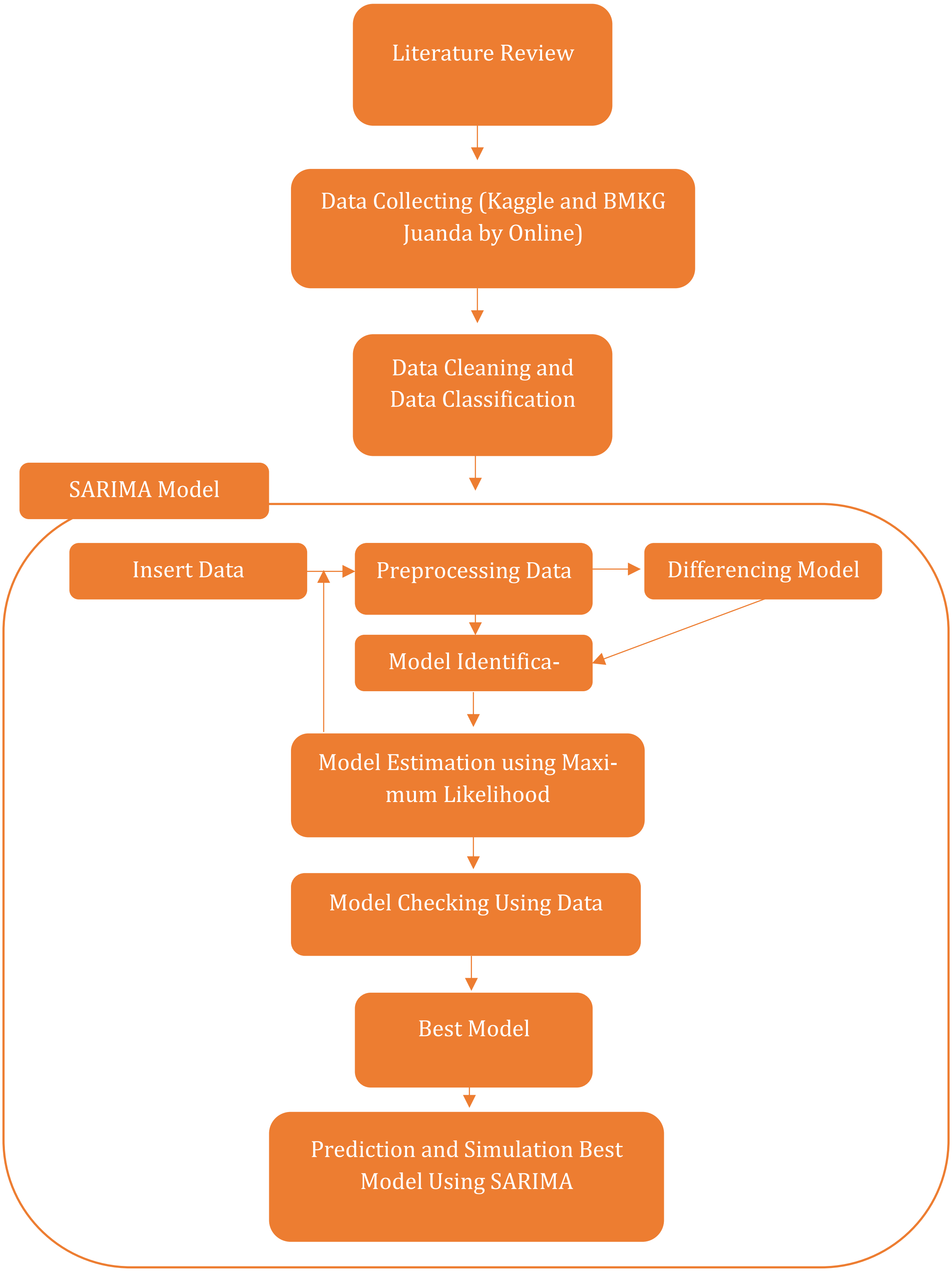

Figure 1. Scheme of processing data using SARIMA

Results and Discussion

In this section, we will provide the results of our research.

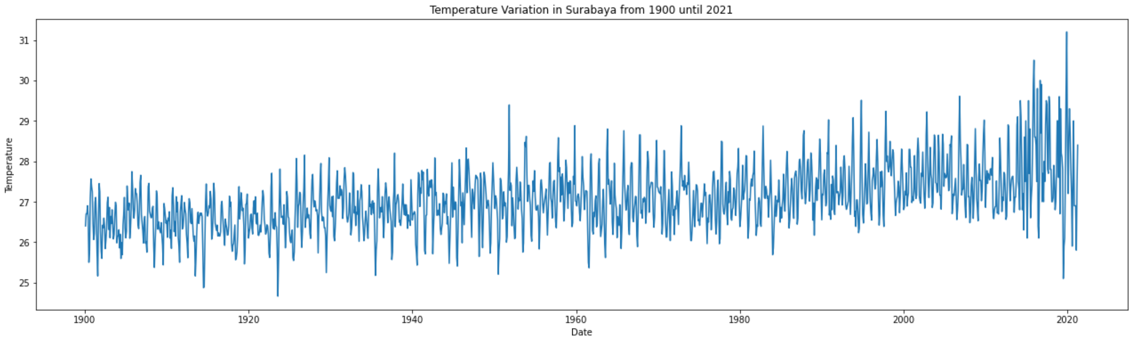

Figure 2. Temperature data along 100 years

The following plots a graph of temperature data in Surabaya and its surroundings from 1900 to 21 April 2021. It can be seen that the lowest temperature is around 24.8o C, around the 1920s, on the other hand, the highest temperature data is around 31.2o C in the 2020s. This shows a difference in the value of 6.4o C or 4.25% in 100 years.

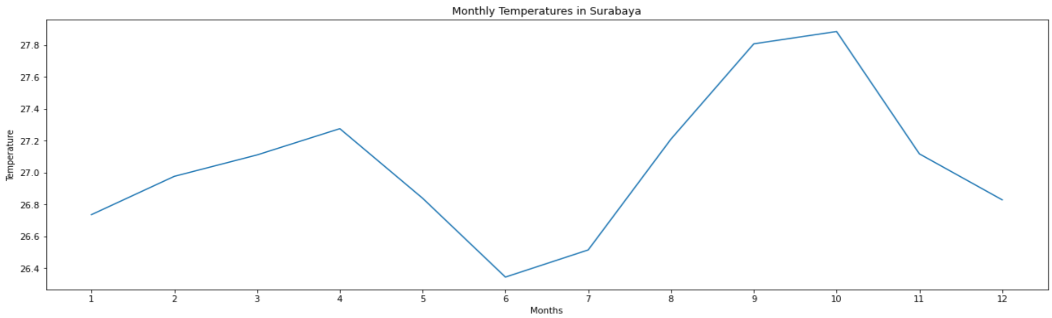

Figure 3. Temperature monthly data along 100 years

It can be seen that the lowest average temperature is in the 6th month of June, whereas the highest average temperature is found in October.

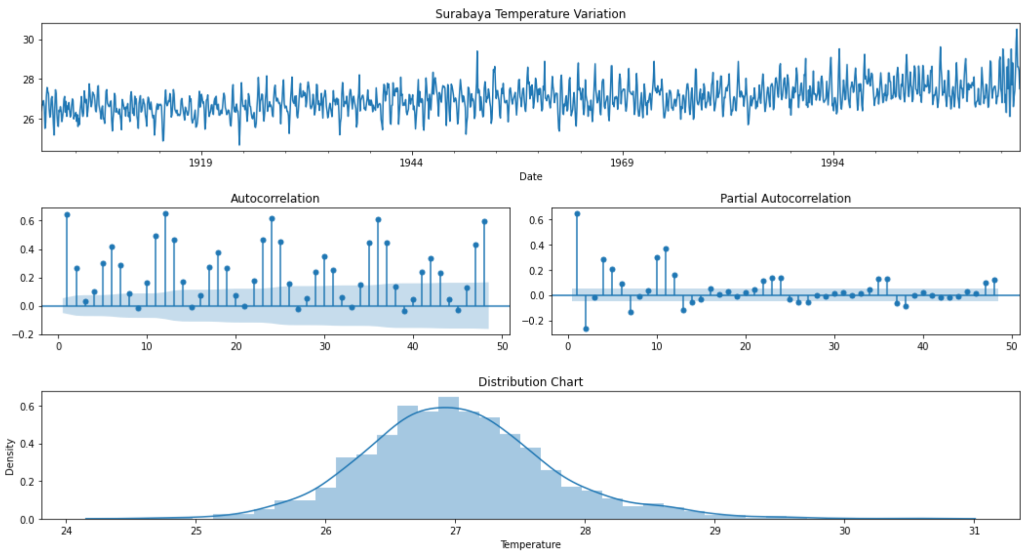

Figure 4. ACF and PACF Plot

Figure 4. is used to check for rationalization by observing the ACF and PACF plots. For ACF shows the correlation between the current temperature with the previous version. While the PACF shows the correlation between the current temperature and the previous version excluding the effect of lag, for example, this shows the effect of effectiveness in the current temperature. The best model of Seasonal ARIMA obtained is as follows SARIMA (3,0,0) (0,1,1,12).

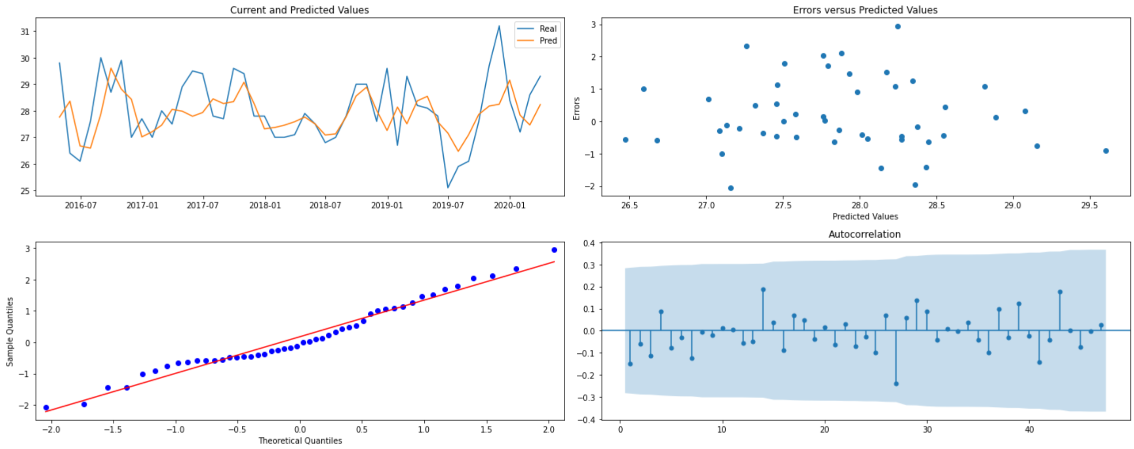

Figure 5. Error value and prediction value plot

Analyzing the plot of Figure 5 above, it can be seen that the prediction is very much in line with the current value. The error and prediction values have a linear distribution, where the error value is between -1.5 and +1.5 when the temperature increases. For Q-Q Plot shows a normal pattern with some small anomalies. The auto-correlation plot shows a positive graph spike during the confidence interval just above the second break, but going forward we are confident that there will be no further change.

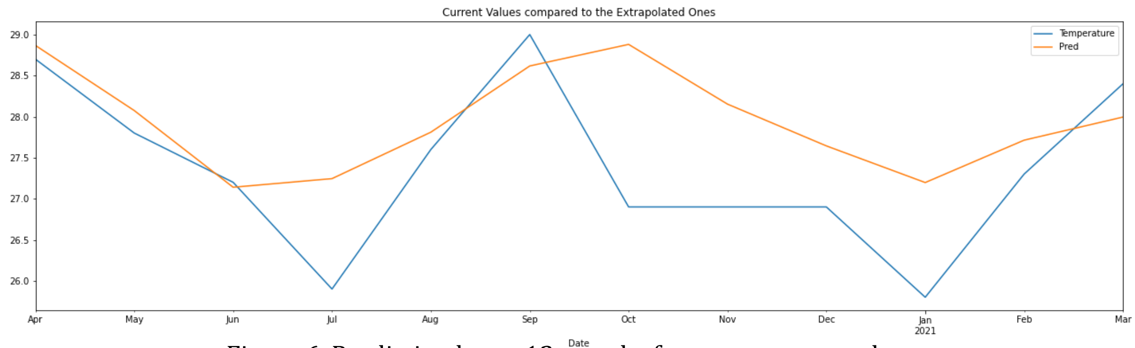

Figure 6. Prediction latest 12 months from temperature data

It seems that the SARIMA parameter is appropriate, the predicted value follows the exact value and also the seasonal pattern. This is supported by the evaluation of the RMSE value when testing data on extrapolation, with a baseline RMSE test of 1.18^0 C and an extrapolation test of 0.93^0 C, and can make improvements of 21.16%.

Conclusion

It can be simplified that the temperature prediction in the Surabaya and surrounding areas has an accurate value with the selection of the best model SARIMA (3,0,0) (0,1,1,12).

Acknowledgment

This work was financially supported by LP3M through "DIPA 2021". Therefore, we are grateful for this funding and support of this research.

References

Acosta, L. M., et al. (2020). SARIMA approach to generating synthetic monthly rainfall in the Sinu River Watershed Colombia. *Atmosphere, 11*(6), 602, doi:10.3390/atmos11060602

Chatfield, C., & Xing, H. (2019). *The analysis of time series an introduction with R, 7th ed.* London, UK: Chapman and Hall/CRC.

Chen, P., Niu, A., Liu, D., Jiang, W., & Ma, B. (2018). Time Series Forecasting of Temperature using SARIMA: An Example from Nanjing". ACMME 2018. *IOP Conf. Series: Material Science and Engineering, 394*(2018), 052024.

Dastorani, M., Mirzavand, M., Dastorani, M. T., & Sadatinejad, S. J. (2016). Comparative study among different time series model applied to monthly rainfall forecasting in semi-arid climate condition. *Nat. Hazards, 81,* 1811-1827.

Li, Y. P., Nie, S., Huang, C. Z., McBean, E. A., Fan, Y. R., & Huang, G. H. (2017). An integrated risk analysis method for planning water resource systems to support sustainable development of an arid region. *J. Environ. Inform., 29*(1), 1–15. Doi:10.3808/jei.201700359

Radhakrishnan, P., & Dinesh, S. (2006). An alternative approach to characterize time series data: Case study on Malaysian rainfall data. *Chaos Solitons Fractals, 7*(2), 511–518. https://doi.org/10.1016/j.chaos.2005.04.030

Venkata, R. R., Krishna, B., Kumar, S. R., & Pandey, N. G. (2013). Monthly rainfall prediction using wavelet neural network analysis. *Water Resour. Manag., 27,* 3697–3711. Doi: 10.1007/s11269-013-0374-4

Wang, S., Feng, J., & Liu, G. (2013). Application of seasonal time series model in the precipitation forecast. *Mathematical and Computer Modelling., 58*(3-4), 677–683. https://doi.org/10.1016/j.mcm.2011.10.034