

Forecasting of Childhood Pneumonia in Semarang City

Widya Hary Cahyati
Public Health Science Department
Universitas Negeri Semarang
Semarang, Indonesia
widyahary27@mail.unnes.ac.id

Merlinda Permata Sari
Public Health Science Department
Universitas Negeri Semarang
Semarang, Indonesia
merlindapermata@gmail.com

Abstract— The number of cases of under-five pneumonia in Semarang City recorded at Semarang City Health Office in 2018 was 6,477 cases, a decrease compared to 2017 at 9,586 cases. This study used under-five pneumonia cases in Semarang City as one of the variables to determine the results of ARIMA (*Autoregressive Integrated Moving Average*) time series forecasting. This study aimed to determine the estimation result for number of under-five pneumonia patients in Semarang City from 2019-2021 using the ARIMA method. This study was a non-reactive study (secondary data). Study population were all children under five years old with pneumonia Semarang City during 2012-2018. This study used secondary data, so the sample was all members of the population. Toddler pneumonia data during 2019-2021 was forecasted using ARIMA method that was applied to univariate data with time series identification models. The results of this study indicated that the number of under-five pneumonia patients in Semarang city can be estimated using the ARIMA model (2.0.2) with a MAPE error rate of 25%, meaning that the forecasting ability was fair, feasible, and adequate for practical use. The result of 2019 forecast for toddlers with pneumonia was 4,479 patients, 4,477 patients for 2020, and 4,477 cases in 2021. Conclusion, the ARIMA model (2.0.2) could be used to forecast the number of toddler pneumonia patients in Semarang city.

Keywords: *ARIMA, Forecasting, Time Series, Toddler Pneumonia*

I. INTRODUCTION

Pneumonia is an infectious disease that attacks all ages. Pneumonia is one of the unresolved health problems. According to the Republic of Indonesia's Ministry of Health [1], pneumonia is an acute infection that affects the lung tissue (alveoli). Pneumonia can be caused by viruses, bacteria, parasites or fungi. The most common bacterial cause of childhood pneumonia are *Streptococcus pneumoniae* and *Haemophilus influenzae*.

Pneumonia is the cause of 15-16% of under-five deaths in the world in 2015 and 2016. This disease attacks all ages in all regions, but most occur in South Asia and sub-Saharan Africa. Pneumonia has killed around 2,400 children per day with a large 16% of 5.6 million toddler deaths or around 880,000 toddler in 2016 and 920,136 toddler in 2015[2].

In the Health Profile of the Republic of Indonesia data for 2017 based on the Sub-directorate of Acute Respiratory Infection Routine Report data for 2017, it was

found that the incidence rate in Indonesia was 20.54 per 1000 toddlers. The number toddler pneumonia cases in Indonesia during 2013 to 2017 was fluctuating. In 2013, there were 571,547 cases of toddler pneumonia which increased to 657,490 cases in 2014. The number of cases decreased to 554,650 in 2015. However, it increases again to 568,146 in 2016 before decreasing to 511,434 cases in 2017.

Central Java Province ranked at the third highest cases of toddlers pneumonia in Indonesia from 2013 to 2017 after West Java and East Java Provinces. Semarang City was in third place during 2016 and 2017. The incidence of toddler pneumonia in Semarang City fluctuated during 2012 to 2017. There were 9,586 cases in 2017, 4,173 cases in 2016 and 7,759 cases in 2015. Based on the gender, 46% of toddler pneumonia cases in Semarang City during 2017 were female while 54% were males [3].

The high prevalence of pneumonia among toddlers in Semarang City suggests that it is very important to prevent and control the disease. In this study, efforts will be made to prevent and control pneumonia in infants by predicting the number of pneumonia patients in Semarang city. The results of accurate forecast of pneumonia cases will provide a scientific basis for formulation of appropriate planning and control measures.

This research was conducted using the ARIMA (Autoregressive Integrated Moving Average) method with data on children with pneumonia in Semarang. This ARIMA method can be used to predict the number of under-five pneumonia patients in Semarang city even though the data obtained is not stationary. Forecast of toddler pneumonia is important because estimates on the number of childhood pneumonia based on previous empirical data is needed.

II. METHODS AND MATERIALS

A. Source of Data

The data source of this study was secondary data on toddlers with pneumonia at the Semarang City Health Office from January 2012 to December 2018.

B. Method Description

This research was a secondary data research (nonreactive research). Non-reactive research is a study where the subject under study does not feel or realize that he is being researched. In other words, non-reactive research has no reaction from research subjects. The design of this

research is descriptive time series with ARIMA forecasting method. This research is classified as applied research because it tries to apply forecasting statistical methods. In this case, the forecasting method in the health sector, namely the number of toddler patients with pneumonia.

This study did not conduct sampling because this study used secondary data, in which the sample were all members of the population. The sample was determined by recording data obtained from the Semarang City Health Office.

The ARIMA method is a method developed by George Box and Gwilyn Jenkins so that their names are often used with the ARIMA process or the Box-Jenkins method. ARIMA is applied for the analysis and forecast of time series data. ARIMA modeling uses only one univariate time series to produce accurate short and medium term forecasts.

ARIMA modeling uses an iterative approach in the process of identifying an existing model. The selected model is tested again with past data to see the accuracy of the data. A model is said to be appropriate or appropriate if the residuals between the model and historical data points are of small value, randomly distributed, and free from each other or *white noise*.

The model that meets the requirements is then compared to the error value; the smaller the error value, the better the model. Determination of the best model can be done by comparing the value of Mean Square Error (MSE) because the smaller the MSE value is generated, the better the model.

The MAE value is expressed by the equation:

$$MAE = \frac{\sum |\bar{Z}_t - Z_t|}{n}$$

In addition there are indicators to calculate model errors using MAPE. MAPE (*Mean Absolute Percentage Error*) is used to measure the error of the estimated value with a model expressed in the form of an absolute percentage of absolute errors. The MAPE equation is stated by:

$$MAPE = 100 \frac{\sum \left| \frac{\bar{Z}_t - Z_t}{Z_t} \right|}{n}$$

If the best model has been obtained, this model can be used for forecasting. For data that has been transformed, the forecast results obtained are converted according to the original data.

III. RESULT

In this study, data obtained from the Semarang City Health Office was selected to fit into the data needed. The results of this selection were then used as data processed using the ARIMA method to find out the prediction results of toddlers with pneumonia in Semarang City during 2019-2021. During the data processing stage, it was processed using the help of Minitab 17 software.

The process of identifying models formed a time series pattern in toddler pneumonia data for 2012-2018 in the following figure:

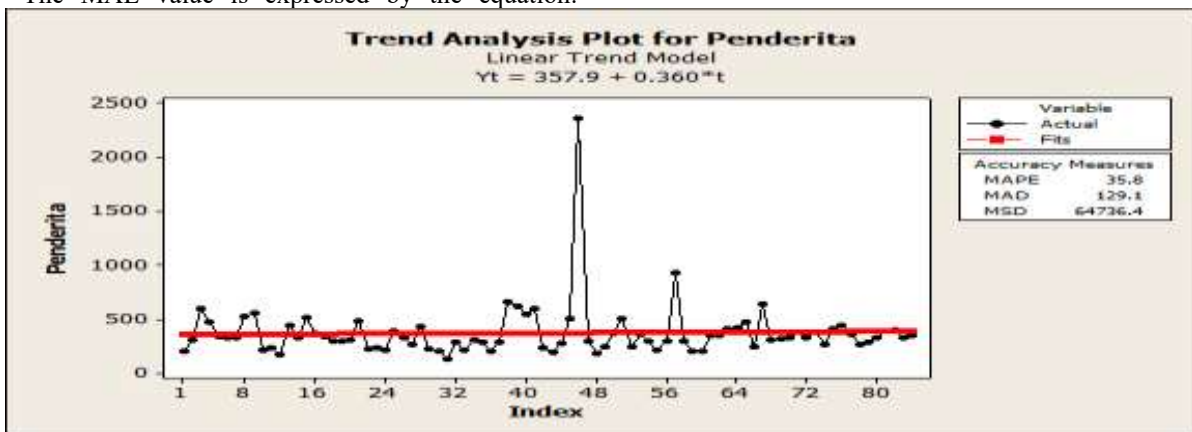


Figure 1. Toddler Pneumonia Patients from January 2012 to December 2018

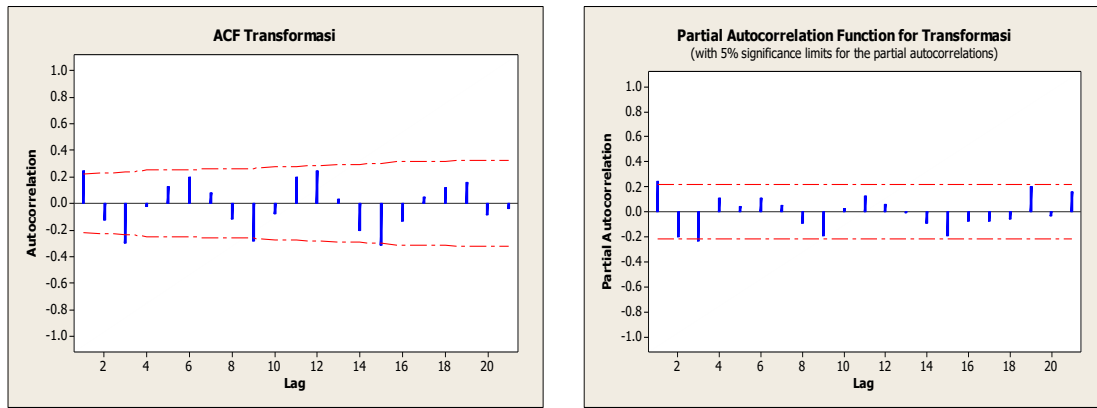


Figure 2. Auto correlation Function (ACF) and Partial Autocorrelation Function (PACF)

In figure 1 the pattern of toddler pneumonia data showed that there were 84 historical data points of toddler pneumonia in Semarang City which were obtained from 84 months (January 2012-December 2018) worth of data. The data plot can be seen if the data pattern formed was not stationary to the variant or the mean. The patients' data plot showed the highest point in the 46th data (October 2015) it was the highest increase in the number of toddler pneumonia cases during the period of 2012-2018. Meanwhile, data in the other months fluctuated from 100 to 600 cases.

As seen from the Box-Cox Plot graph of the patients above, rounded value of the estimated lambda was -0.05. This showed that the data was not stationary because $\lambda \neq 1$. Non-stationary in variants could be eliminated through data transformation to stabilize variance. After data transformation and the lambda value changes into 1, the data has now been stationary against the variant.

The ACF and PACF data plot showed 21 lags with the calculation of the number of lags obtained from $n / 4$. There were 84 data points for the number of toddlers with

pneumonia from January 2012 to December 2018, so there are $84/4 = 21$ lags in the data. Based on ACF and PACF from the previously transformed data, it could be seen that there was cut-off in ACF and PACF after lag 1. Therefore, it could be concluded that the data above is already stationary against the mean.

In the plot graph of ACF and PACF, the data on toddlers with pneumonia in Semarang city did not go through differencing to be stationary, so $d = 0$. There is a cut-off value for the 1st ACF lag plot so that the Moving Average (q) value is 1. The cut-off value for the PACF plot in 1st lag so that the Autoregressive (p) value is 1. The temporary model that had been obtained from the ACF and PACF plots were ARIMA (0,0,1), ARIMA (1,0,0), and ARIMA (1,0,1). Identifying the temporary model besides looking at the ACF and PACF plots could be done by adding the values of AR (p) and MA (q). Then the other temporary models were obtained, namely ARIMA (2,0,2). So that the temporary model that could be concluded was ARIMA (0,0,1), ARIMA (1,0,0), ARIMA (1,0,1), or ARIMA (2,0,2).

Table 1. Parameter Test of Temporary Model

ARIMA Model		Parameter	T Count	P Value	Description
ARIMA (0,0,1)	MA(1)	-0.0944	-0.86	0.393	Insignificant
	Constant	373.05	12.17	0.000	Significant
ARIMA (1,0,0)	AR(1)	0.0745	0.68	0.501	Insignificant
	Constant	345.28	12.32	0.000	Significant
ARIMA (1,0,1)	AR(1)	-0.3418	-0.36	0.718	Insignificant
	MA(1)	-0.4356	-0.48	0.631	Insignificant
	Constant	500.70	12.39	0.000	Significant
ARIMA (2,0,2)	AR(1)	1.0203	50.49	0.000	Significant
	AR(2)	-1.0044	-84.96	0.000	Significant
	MA(1)	1.0163	18.01	0.000	Significant
	MA(2)	-0.9846	-20.02	0.000	Significant
	Constant	365.92	14.07	0.000	Significant

Table 2. White Noise Test Result

Temporary Model	Ljung Box Chi Square Statistic				Description
	Lag 12	Lag 24	Lag 36	Lag 48	
	P-Value	P-Value	P-Value	P-Value	
ARIMA (0,0,1)	0,265	0,575	0,815	0,921	White Noise
ARIMA (1,0,0)	0,238	0,542	0,790	0,902	White Noise

ARIMA (1,0,1)	0,238	0,541	0,789	0,911	White Noise
ARIMA (2,0,2)	0,467	0,379	0,783	0,965	White Noise

Table 3. Residual Normality Test

Model ARIMA	Significance Test	White Noise Test	Residual Normality Test	Description
ARIMA (0,0,1)	Significant	White Noise	Normal Distribution	Acceptable
ARIMA (1,0,0)	Significant	White Noise	Normal Distribution	Acceptable
ARIMA (1,0,1)	Significant	White Noise	Normal Distribution	Acceptable
ARIMA (2,0,2)	Significant	White Noise	Normal Distribution	Acceptable

Diagnostic test in the form of significance test, white noise test, and residual normality test on ARIMA model (0,0,1), ARIMA (1,0,0), ARIMA (1,0,1) and ARIMA (2,0,2) met the requirements to be used in the forecast, namely the four models were significant, were white noise, and normally distributed. However, in a forecast we had to use the best model. We could determine the best model from the four temporary ARIMA models by looking for the smallest MSE (Mean Square Error) value among the four models which means the least estimated value error.

IV. DISCUSSION

The time series data used in this study was data on children with pneumonia in Semarang City from 2012 to 2018 recorded in the Semarang City Health Office. Toddler pneumonia data was available every month, therefore 84 historical data points were obtained from January 2012 to December 2018. This data included all pneumonia patients from 37 healthcare centers in Semarang City. These data indicated a change in the number of pneumonia patients in Semarang city every month of each year.

The number of pneumonia patients in Semarang city fluctuates every month. The lowest number of patients was at June 2014, 134 cases. The highest number was in 2,355 cases October 2015. This was a very significant surge compared to other data which fluctuated in the hundreds during January 2012 to December 2018. Compared to the adjacent months, September 2015 had 509 patients while November 2015 had only 296 patients. This showed the variability in the number of pneumonia patients in Semarang city on a monthly basis.

The results of the ARIMA model (2,0,2) had the smallest MSE value, 61957, compared to the MSE values of other models such as 65918 for ARIMA Model (0,0,1), 66025 for ARIMA (1,0,0), and 66550 for ARIMA (1,0,1). Therefore, ARIMA model (2,0,2) was the best model and was used to forecast of Pneumonia Patients in Semarang City during 2019-2021.

Comparison between actual data and forecast data for 2016-2018 showed that the forecast was the closest to the actual data in February 2017 with a deviation of 0%. While the furthest deviation was 100% in April 2016, with a forecast value of 497 patients and an actual data as many as 248 patients.

The ARIMA model (2,0,2) was the best model used to test the deviation results and comparison of results between the actual data with forecasts for 2016 to 2018. In

table 4.9, the total deviation result from January 2016 to December 2018 was 914% with monthly average deviation or MAPE (Mean Absolute Percentage Error) value of 25%.

In the best model, ARIMA model (2.0.2), the deviation value (MAPE) was 25%, which means that the model is sufficient for forecasting. The results of the forecast for toddlers with pneumonia was 4,479 patients in 2019, 4,477 patients in 2020, and 4,477 cases in 2021.

The estimated tuberculosis morbidity in Xinjiang China based on data from January 2004 to June 2014 with an incidence rate of 1,525 per 100,000 population produced the best method, namely ARIMA (1,1,2) (1,1,1) model and the combination of ARIMA models (1,1,2) (1,1,1) ARCH (1). From the two models obtained, the smallest error value was found in the ARIMA ARCH combination model using error value indicators RSME, MAE, and MAPE. Therefore, the model comparative analysis showed that the combination of ARIMA and ARCH models was more effective [4]. Compared to the same study in different countries, there were fewer cases of tuberculosis in Iran, namely 34,012 patients with an incidence rate of 9.8 per 100,000 population [5].

In Iran, the incidence of tuberculosis in 2005 to 2011 did not fluctuate. The highest tuberculosis cases was 6,579 cases and the lowest was 4,579. The data plot generated from 2005 to 2011 showed a seasonal data pattern, so the method used is the seasonal ARIMA model. The best model obtained was used to predict the incidence of tuberculosis in Iran in 2012-2014, namely the seasonal ARIMA model (0,1,1) (0,1,1). The estimated incidence of tuberculosis in 2012 was 14.69 per 100,000 population, in 2013 it was 15.66 per 100,000 population, and in 2014 it was 16.75 per 100,000 population. Therefore, it was known that tuberculosis incidence in Iran from 2012 to 2014 has increased slowly [6].

The diagnosis of HIV patients in the Republic of Korea from 2013 to 2017 was estimated at 14,724 cases and more than 15,000 cases in 2018. This was based on a research by [7] with data on HIV patients from 1985 to 2012. ARIMA model (1,2,3) was used to predict HIV patients in the Republic of Korea with a MAPE value of 13.7%. In 2012, the 30-39 age group became the highest age group infected with HIV, which amounted to 29%.

Leishmaniasis is a disease caused by the leishmania parasite. Prediction of leishmaniasis cases number in South Fars Province of Iran used the best model of the seasonal ARIMA model (4,1,4) (0,1,0). Data used in

case prediction was monthly data from January 2010 to March 2016. Information obtained included temperature, humidity, rainfall, dry season, rainy season, and evaporation affects the number of leishmania cases. Therefore, the pattern of the data generated showed a pattern of seasonal data from 2010 to 2016. Therefore, it was necessary to intervene and procure special programs in the seasons that triggered the increase in the number of leishmania cases in South Fars Province, Iran [8].

Based on research conducted by [9], it was stated that the most important risk factors in the incidence of childhood pneumonia were gender, type of residence, education, family economic level/ownership quintile, separation of kitchens from other rooms, availability/habit of opening room windows and adequate room ventilation with a sample of 82,666 people.

According to a study involving 138 subjects by [10], there were 4 significant risk factors, namely age, history of breastfeeding, nutritional status, and family smoking habits. The etiology of pneumonia varied with the two most common etiologies being viruses and bacteria. The [11] study stated that parasitic infections during pregnancy have no effect on antibody responses to early vaccination, including *Streptococcus pneumoniae*. Subanada and Purtini (2010) stated that the temperature and number of leukocytes in infants is associated with the presence of bacterial pneumonia. From [9], [10], as well as [12], there are several possible risk factors which contributed to childhood pneumonia cases in Semarang City.

An increase in the incidence of pneumonia could also be caused by resistance [13]. However, now there are vaccines to prevent pneumonia [14], and research into the treatment of pneumonia is still being developed including on African Green Monkeys [15]. Based on the research of [16], there was a correlation between age and antibiotics use among children in Southeast Asia.

V. CONCLUSIONS

Forecast of patients with toddlers pneumonia in Semarang City used the best model, namely the ARIMA model (2,0,2) with a MAPE value of 25%. The toddler pneumonia forecast predicted 4,479 patients in 2019, 4,477 patients in 2020, and 4,477 cases in 2021. The highest estimated number of toddlers with pneumonia is predicted to occur in October 2021 with an estimate of 501, while the smallest number is in July 2021 with 243 patients. In the upcoming years, the monthly pneumonia forecast data for each month is fluctuating; since there is no trend, it couldn't be predicted which months would face a surge in the number of pneumonia patients.

REFERENCES

- [1] Kemenkes, "Tatalaksana Pnemuonia Balita di Fasilitas Pelayanan Kesehatan Tingkat Pertama," Jakarta, 2018.
- [2] UNICEF, "Pneumonia Claims The Lives of The World's Most Vulnerable Children," 2018. .
- [3] Dinkes Kota Semarang, "Profil Kesehatan Kota Semarang 2017," Semarang, 2018.
- [4] Y. Zheng, "Forecast Model Analysis for the Morbidity of Tuberculosis in Xinjiang, China," *J. Plos One*, vol. 10, no. 3, pp. 1–13, 2015.
- [5] M. Moozasadeh, N. Khanjani, M. Nasehi, and A. Bahrampour, "Predicting the Incidence of Smear Positive Tuberculosis Cases in Iran Using Time Series Analysis," *Iran J. Public Heal.*, vol. 44, no. 11, pp. 1526–1534, 2015.
- [6] M. Moosazadeh, M. Nasehi, A. Bahrampour, N. Khanjani, S. Sharafi, and S. Ahmadi, "Forecasting Tuberculosis Incidence in Iran Using Box-Jenkins Models," *Iran Red Crescent Med. J.*, vol. 16, no. 5, pp. 1–6, 2014.

- [7] H.-K. Yu, N.-Y. Kim, S.-S. Kim, C. Chu, and M.-K. Kee, "Forecasting the Number of Human Immunodeficiency Virus Infections in the Korean Population Using the Autoregressive Integrated Moving Average Model," *J. Osong Public Heal.*, vol. 4, no. 6, pp. 358–362, 2013.
- [8] M. Sharafi, H. Ghaem, H. R. Tabatabaee, and H. Faramarzi, "Forecasting the Number of Zoonotic Cutaneous Leishmaniasis Cases in South of Fars Province, Iran Using Seasonal ARIMA Time Series Method," *Asian Pasific J. Trop. Med.*, vol. 10, no. 1, pp. 79–86, 2016.
- [9] A. Anwar and I. Dharmayanti, "Pneumonia Anak Balita di Indonesia," *J. Kesehat. Masy. Nas.*, vol. 8, 2014.
- [10] S. Hartati, N. Nurhaeni, and D. Gayatri, "Faktor Risiko Terjadinya Pneumonia Pada Anak Balita," *J. Keperawatan Indones.*, vol. 15, pp. 13–20, 2012.
- [11] N. D. McKittrick *et al.*, "Parasitic Infections During Pregnancy Need not Affect Infant Antibody Responses to Early Vaccination Against *Streptococcus pneumoniae*, Diphtheria, or *Haemophilus Influenzae* Type B," *PLoS Negl. Trop. Dis.*, 2019.
- [12] I. B. Subanada and N. P. . Purniti, "Faktor-Faktor yang Berhubungan dengan Pneumonia Bakteri pada Anak," *Sari Pediatr.*, vol. 2, 2010.
- [13] L. J. Quinton and J. P. Misgerd, "Dynamics of Lung Defense in Pneumonia: Resistance, Resilience, and Remodeling," *Annu. Rev. Physiol.*, vol. 77, pp. 407–430, 2015.
- [14] N. J. Croucher, A. Løchen, and S. D. Bentley, "Pneumococcal Vaccines: Host Interactions, Population Dynamics, and Design Principles," *Annu. Rev. Microbiol.*, vol. 72, pp. 521–549, 2018.
- [15] R. C. Layton *et al.*, "Levofloxacin cures experimental pneumonic plague in african green monkeys," *PLoS Negl. Trop. Dis.*, 2011.
- [16] S. J. Salter *et al.*, "A Longitudinal Study of the Infant Nasopharyngeal Microbiota: The Effects of Age, Illness and Antibiotic Use in a Cohort of South East Asian Children," *PLoS Negl. Trop. Dis.*, 2017.