Journal of Computers for Society
Journal homepage: https://ejournal.upi.edu/index.php/JCS
Vol. 1 No. 1, June-2020, pp. 61-79

Prediction of Diarrhea Sufferers in Bandung with Seasonal Autoregressive Integrated Moving Average (SARIMA)

Cacuk Jati Pangestu[a], Erna Piantari[b], Munir[c]

[a, b, c] *Department of Computer Science Education, Universitas Pendidikan Indonesia, Bandung, Indonesia*
[a]cacukjati@student.upi.edu, [b]erna.piantari@upi.edu, [c]munir@upi.edu

Abstract

Diarrhea is the second disease that causes death in children in the world. Every year, around 1.7 million cases of diarrhea are found and cause around 525,000 deaths in children under the age of five in the world. Proper analysis of health service data can help predict epidemics, cure, and disease, and improve quality of life and avoid preventable deaths. This research is aimed at predicting diarrhea sufferers in the future by using Seasonal Autoregressive Integrated Moving Average (SARIMA) and Seasonal Autoregressive Integrated Moving Average with explanatory X (SARIMAX) by involving climate factors in the form of average temperature and average humidity. The data used are data of diarrhea sufferers and climate in 2010-2019 in the city of Bandung. The result shows that there is not significant relation between temperature or humidity and the diarrhea cases. However, the SARIMA model had performed better than the SARIMAX model with the addition of climate factors to predict the diarrhea case in Bandung. The predictive accuracy of the SARIMA model obtained is 78.6%.

Keywords: Diarrhea, Time Series, SARIMA, SARIMAX

First draft received: June 12, 2020
Date Accepted: June 16, 2020
Final proof received: June 25, 2020

1. Introduction

Diarrhea is the second disease that causes death in children in the world. Every year, around there are 1.7 million cases of diarrhea and cause around 525,000 deaths in children under the age of five in the world [1]. Unclean water and poor sanitation are the main factors of diarrheal disease in children under the age of five years [2].

Diarrhea diseases are classified based on critical level to two shorts [3]. The Diarrhea on the lasting less than two weeks is classified as acute diarrhea, whereas diarrhea lasting two weeks or more is classified as chronic diarrhea. Patients' feces can be accompanied by mucus, blood, or pus. Some of the symptoms that can be experienced are nausea, vomiting, abdominal pain, heartburn, tenesmus, fever, and signs of dehydration. The risk factors for acute diarrhea are personal hygiene, bad sanitary, the history of lactose intolerance, and the sexually transmitted infections such as HIV infection. However, personal hygiene and bad sanitary conditions contribute for the most number of the diarrhea cases.

Previous studies have reported a relationship between weather and diarrhea in several places, such as in Taiwan by using the Spearman's Correlation and Regression Analysis method to conclude a relationship between temperature and rainfall in diarrhea cases [4], in Sub-Saharan Africa by using the Ordinary Least method Square (OLS) Regression concludes there is a relationship between temperature and diarrhea cases [5]. Then in Bangladesh using Statistical Analysis concluded the case of diarrhea increased when temperatures were higher [6]. By analyzing and knowing the relationship between weather and diarrhea, we can predict or predict future diarrhea cases.

There are so many previous researches about Forecasting or prediction using time series analysis. One of the methods that are considered to have good accuracy for time series analysis is the Autoregressive Integrated Moving Average method which is briefly ARIMA [7]. This ARIMA method has been implemented for forecasting such as forecasting in food retailing [8], forecasting on short term load electricity [9], and forecasting fuel demand in Turkey [10]. For models that have a seasonal trend, Seasonal ARIMA (SARIMA) has been implemented in leishmaniasis disease forecasting in Iran [11] and short-term traffic flow forecasting in India [12]. The ARIMA method with explanatory variable X (ARIMAX) has also been implemented in export forecasting in Thailand [13], forecasting demand for children's Muslim clothes [14], and non-Nigerian oil exports [15]. Then the ARIMA seasonal method with explanatory variable X (SARIMAX) was implemented in the prediction of photovoltaic power plants [16].

Appropriate analytical tools enable health specialists to collect and analyze patient data, which can be used by insurance agents and administrative organizations. Beside of that, the proper analysis of health service data could help to predict epidemics, cure, and disease as well as improve quality of life and avoid preventable deaths [17].

This study was conducted to build a prediction model for Diarrhea Sufferers in Bandung, Indonesia. The Pearson Correlation was use to figure out the correlation between the temperature, humidity and the diarrhea cases in Bandung. Moreover, the SARIMA and SARIMAX were used to build the prediction model. If the temperature or humidity has good correlation with the number of diarrhea sufferers, then forecasting using SARIMAX with temperature or humidity can be expected to produce a better model.

2. Methods

There are four main stages to build a prediction model (see **Figure 1**): preprocessing data, build SARIMA model, build SARIMX model and model validation. All steps are explained in the next subsection.

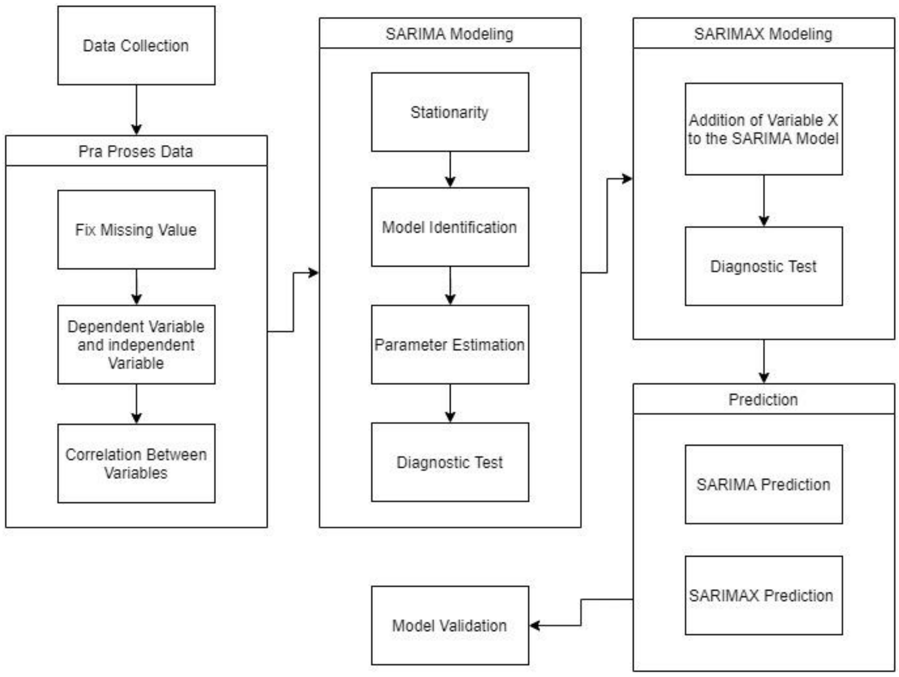

Figure 1. *Design of prediction research for diarrhea sufferers using the SARIMA and SARIMAX methods*

2.1 Data set and Pre-processing Data

In this study we consider the diarrheal diseases case in Bandung to be predicted. There are two kinds of data set that we use in this study: diarrheal diseases and weather. In accordance to the study objectives which is to predict the diarrheal case monthly, all the data set are present in monthly period. Diarrheal case data was carried out by health ministry observation of Bandung for 5 years form 2010-2019. The weather data was downloaded from Meteorological, Climatology, and Geophysical agency (BMKG) Bandung, and it consists temperature and humidity data.

Before building the model prediction, there were two initial steps to prepare the data to be processed. These stages becomes very important because it determines the quality of the data which will greatly affect the predicted results of the model being built. The three stages of data preparation were the process of missing value handling, and the search for correlation values between the dependent variable and the independent variable.

The missing value handling in sessional data was done by filling it with the mean of the values in the same phase in others period. For example, we have a missing value for diarrheal case number in January 2019. This missing value was filled by the average of diarrheal case number in January for 5 years before. From the collected data, there were 15 missing values for diarrhea case data, 3 missing value for the temperature and humidity. All the missing value was filled with the mean data of value from the same phase in others year. The sample data can be seen in **Table 1**.

The next step in the preprocessing was finding the correlation between dependent and independent variable. In this case, the dependent variable is the diarrhea and the independent variables are temperature and humidity. Pearson correlation was used to find the correlation value between diarrhea and temperature, diarrhea and humidity. Pearson correlation value could explain the relationship between two linear variables. Variables have a positive correlation if the value of the Pearson correlation approaches 1, and has a negative correlation if the value approaches -1.

__Table 1.__ Data of diarrhea, temperature, and humidity.

Data in 2010			
Month	Diarrhea	Temperature	Humidity
January	5023	23,48	79,30
February	4830	23,61	80,22
March	4595	23,72	79,85
April	4851	23,97	80,39
May	4737	23,85	79,34
June	4530	23,44	77,24
July	4980	22,90	74,75
August	4739	23,29	70,95
September	4511	23,70	70,11
October	5423	24,04	73,05
November	5879	23,50	81,40
December	5258	23,81	79,70

While the variable does not correlate it has a value of 0. The equation to find the Pearson correlation value is expressed by **Eq. 1** and **Eq. 2** [19].

$$r_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \underline{x})(y_i - \underline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \underline{x})^2 \sum_{i=1}^{n}(y_i - \underline{y})^2}} \quad (1)$$

$$r_{x,y} = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\sum x_i{}^2 - \frac{(\sum x_i)^2}{n}} \sqrt{\sum y_i{}^2 - \frac{(\sum y_i)^2}{n}}} \quad (2)$$

where r is the Pearson correlation value between the variables x and y, and n is the amount of data.

2.2 SARIMA Modelling

SARIMA modelling was done in four stages: stationary test, identify model function based on the plot of the Autocorrelation Function (ACF) and Partial Autocorrelation

Function (PACF), estimation the model parameters and conduct diagnostic test on the model.

- Stationary Test

Stationary test is a vital step in build a prediction model. Using non-stationary data would lead the result of the prediction model become poor. However, to build a prediction model using seasonal data must be transform the data into the stationary data form. Therefor the stationary test must be performed to make sure that the data was stationer. This test was done by observing the Autocorrelation Function (ACF) plot and histogram data distribution. ACF plot would show the stationary of mean data. If the ACF plot shows that data decline too fast and tend to 0, then it can be ascertained that the data is non-stationary, on contrary if the decline is very slow or tends not to approach 0 the data is stationary. ACF value was calculated using **Eq. 3** [20].

$$\rho_k = \frac{\sum_{t=k+1}^{n}\left(Z_t - \underline{Z}\right)\left(Z_{t-k} - \underline{Z}\right)}{\sum_{t=1}^{n}\left(Z_t - \underline{Z}\right)^2} \quad k = 1,2,\ldots \tag{3}$$

Where:

ρ_k = correlation coefficient -lag

Z_t = observation value at

Z_{t-k} = observation value at $(t + k)$

$\underline{Z}$ = average observation value

- Specify the Sessional Model

The second step was identifying the model function to specify the sessional model by observing the ACF plot and Partial Autocorrelation Function (PACF) plot. PACF partially calculate the correlation between value t and value t+1, without being influenced by the value among them. While ACF calculate the correlation between value t and value t+1 and still consider the value among them overall. The Partial Autocorrelation Function (PACF) is expressed by **Eq. 4** [20].

$$\emptyset_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1}\emptyset_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{k-1}\emptyset_{k-1,j}\rho_j} \tag{4}$$

Where:

$\emptyset_{kk}$ = partial autocorrelation coefficient lag-k.

$$\emptyset_{kj} = \emptyset_{k-1,j} - \emptyset_{kk}\emptyset_{k-1,j-1} \qquad j = 1,2, \dots k-1$$

The result of these plots observation will determine the sessional model that would build. For example, the plots will determine what session or how many phase would be used in one session or one period of data. It would be one session consists of 6 phases or 12 phases.

- Parameter Estimation for SARIMA

After the session model was determine, the next step was building the SARIMA model. SARIMA is an ARIMA model by adding seasonal components. In general, the SARIMA model (p, d, q) (P, D, Q) in **Eq. 5.** [18].

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Z_t = \theta_q(B)\Theta_Q(B^s)e_t \qquad (5)$$

Where $\phi_p(B) = \left(1 - \phi_1 B - \cdots - \phi_p B^p\right)$ is an AR process(p), $\theta_q(B) = \left(1 - \theta_1 B - \cdots - \theta_q B^q\right)$ is a MA process(q), $\Phi_P(B) = (1 - \Phi_1 B^s - \cdots - \Phi_P B^{s\,x\,P})$ is a seasonal AR(P), $\Theta_Q(B) = \left(1 - \Theta_1 B^s - \cdots - \Theta_Q B^{s\,x\,Q}\right)$ is a seasonal MA process(Q), $(1-B)^d$ is a differencing process, $(1-B^s)^D$ is a seasonal differencing process, Z_t is the time series and e_t is error value.

In SARIMA modelling, some parameters were needed. Building some models with different parameters then identified the best model by conducting diagnostic test were did to get the best prediction model.

- Diagnostic test

The diagnostic test plays an important role to decide the best prediction model. Therefore, the last step to build the SARIMA model is conducting diagnostic tests on model residuals using the Ljung-box test **Eq. 6** [21].

$$Q_M = n(n+2)\sum_{k=1}^{M}\frac{r_k^2}{n-k} \qquad (6)$$

Where Q_M is the chi-square distribution M-lag and r_k is the value of the ACF at the k-lag. Furthermore, the model is considered feasible if it has residuals that are white noise.

The best model could be known from the smallest AIC value and the largest log-likelihood value. Akaike's Information Criterion (AIC) is expressed by **Eq. 7** [20].

$$AIC = -2\ log\ log\ (maximum\ likehood)\ + 2k \qquad (7)$$

where $k = p + q + 1$.

2.3 SARIMAX Modelling

The extension of the SARIMA model is the SARIMA model with explanatory variable X, also called SARIMAX (p, d, q) (P, D, Q) s. Both models were used to build a sessional model. The comparisons of SARIMA and SARIMAX was carried out for build a forecasting power plants model [16]. In some data period, SARIMAX had better performance, while in another data period, SARIMA showed better prediction [16]. SARIMA model which was originally univariate becomes multivariate. In general, the SARIMAX model (p, d, q) (P, D, Q) in **Eq. 8.** [16].

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Z_t = \beta_1 X_{1,t} + \cdots + \beta_k X_{k,t} + \theta_q(B)\Theta_Q(B^s)e_t \qquad (8)$$

$X_{k,t}$ is the k independent variable or the k-explanatory variable at time t with $k = 1,2,3,\dots,k$.

All results from the SARIMA modelling steps before the estimating SARIMA parameter, would be used to build SARIMAX model. In **Eq 8** The X variable is the independent variable that have been identified in the previous steps. In this study, Temperature and humidity were the independent variables. Both independent variables could be used for the SARIMAX modelling or choose one of them that has better correlation with dependent variable.

- Diagnostic Test

Diagnostic test were always done to find the best model. As was done when building the SARIMA model, diagnostic test process was conducted using Ljung-box test and Akaike's Information Criterion (AIC).

2.4 Model Validation

Model validation was done by calculating the value of the mean absolute percentage error (MAPE) (see **Eq. 9)** and root mean square error (RMSE) (see **Eq. 10**). MAPE will visual the absolute error occurs which is it robust to the outlier errors, while the RMSE will be more sensitive if any outlier error. However, both validation model could be used for considering the best prediction model.

$$MAPE = \frac{\sum_{t=1}^{n}\left|\frac{x_t - f_t}{x_t}\right|}{n} x100\% \tag{9}$$

Where:

n = amount of data

x_t = actual value in the t-period

f_t = forecast value in the t-period

$$RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_{t=1}^{n}(x_t - f_t)^2}{n}} \tag{10}$$

3. Results and Discussion

From the experiments that have been carried out, the following are the results and explanation for each stage that has been carried out. Some of them are pre-processing stages, SARIMA modeling, SARIMAX modelling and validation model.

3.1 Pre-processing Data

The objective of pre-processing data is providing high quality data which was ready to train for data modelling. As explained before (see section 2.1), in this study some missing values was found and has been solved by filling with the mean value. Furthermore, it has been found that Diarrhea and temperature sufferers have Pearson correlation value = -0.166, diarrhea and humidity sufferers data = 0.181, temperature and humidity = -0.162. It can be concluded that the incidence of diarrhea is not affected by temperature or humidity from our Data. This result was different with the previous result in Taiwan [4], Africa [5], and Bangladesh [6]. This might happen due to the climate in Bandung is stable, while in the previous research those country has various seasons.

3.2 SARIMA and SARIMAX Modeling

Based on the ACF plot in **Figure 2**, the diarrhea sufferer data is not stationary in the average and has a seasonal component of 6 or 12. Because the data is not stationary, non-seasonal differencing needs to be done. After performing the first non-seasonal differencing, the ACF plot in **Figure 3** shows that the data is stationary.

Next, identify the model by observing the ACF and PACF plots after doing seasonal differencing 6 and 12. P orders can be identified through a significant lag in PACF and P orders can be identified through significant seasonal lags in PACF. Whereas order q can be identified through significant lags in ACF and order Q can be known through significant seasonal lags in ACF.

The ACF sufferers of diarrhea after differencing 1 and seasonality 6 plot can be seen in **Figure 4**, and the PACF plot can be seen in **Figure 5**.

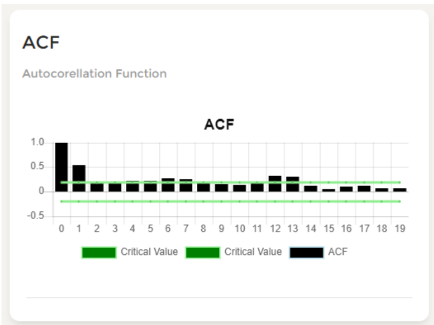

Figure 2. *ACF diarrhea sufferer.*

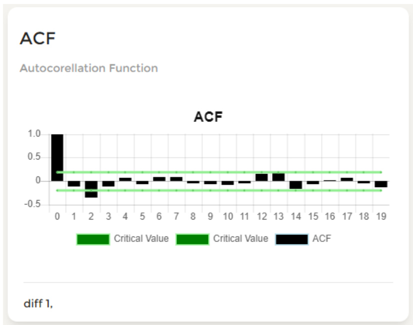

Figure 3. *ACF diarrhea sufferers after differencing 1.*

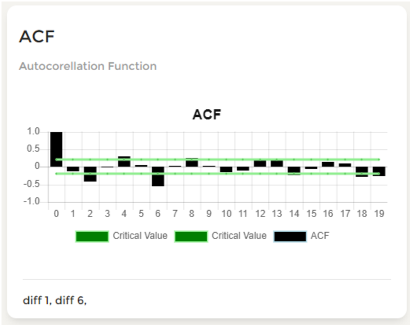

Figure 4. *ACF sufferers of diarrhea after differencing 1 and seasonality 6.*

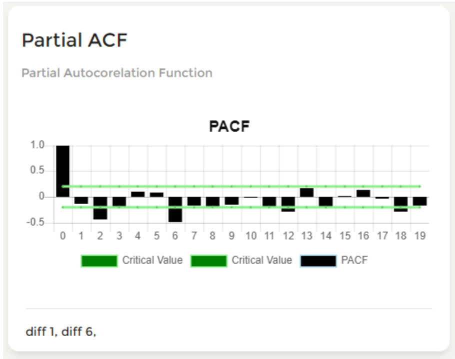

Figure 5. *PACF diarrhea sufferers after differencing 1 and seasonality 6.*

By observing seasonal ACF and PACF plots 6 in **Figure 4** and **Figure 5**, a significant ACF plot at lags 0, 2, and 6 then plot PACF significant at lags 0, 2, 6, 12, and 18, then seasonal 6 whereas for seasonal 12 in **Figure 6** and **Figure 7**, the ACF plot is significant at lags 0, 2, and 12 then the PACF plot is significant at lags 0, 2, 5, and 12, then seasonal 12 can be performed.

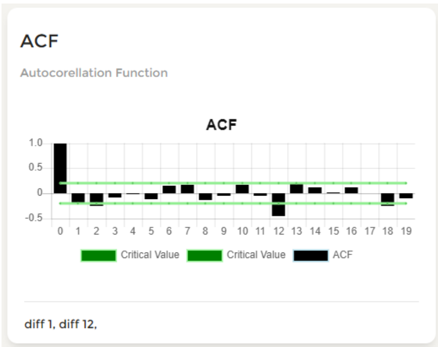

Figure 6. *ACF sufferers of diarrhea after differencing 1 and seasonal 12.*

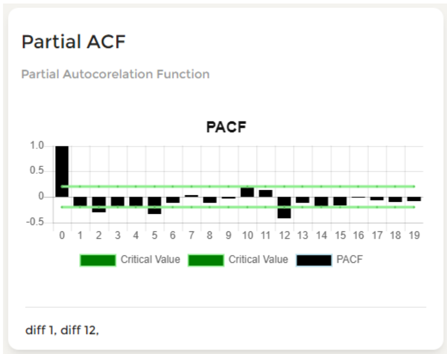

Figure 7. *PACF sufferers of diarrhea after differencing 1 and seasonal 12.*

Based on ACF and PACF seasonal plots 6 and 12, the authors obtain several estimated model parameters and perform diagnostic tests using ljung-box tests on residuals, calculating the AIC and log-likelihood values. The diagnostic test results of the SARIMA model can be seen in **Table 2.**

__Table 2.__ SARIMA diagnostic test results.

Model SARIMA	Log likelihood	AIC	White noise
SARIMA(2,1,2)(0,1,0)6	-823,202	1656,404	No
SARIMA(2,1,2)(1,1,1)6	-741,883	1497,766	Yes
SARIMA(2,1,0)(1,1,0)6	-771,365	1550,729	Yes
SARIMA(2,1,2)(0,1,0)12	-764,608	1539,215	No
SARIMA(2,1,2)(1,1,1)12	-635,092	1284,124	Yes
SARIMA(2,1,0)(1,1,0)12	-664,333	1336,665	No

The best model is a model that has the smallest AIC value, has the largest log-likelihood value, and residuals are white noise. From the diagnostic tests of each model in Table 2, the best model in season 6 is obtained by the SARIMA model (2,1,2) (1,1,1) 6 and the best model in season 12 is the SARIMA model (2,1, 2) (1,1,1) 12.

Then add the independent variable or explanatory variable X to the best SARIMA model to build the SARIMAX model and conduct diagnostic tests on each model. The diagnostic test results in the SARIMAX model can be seen in **Table 3.**

__Table 3.__ SARIMAX diagnostic test results.

Model SARIMAX	Log like-lihood	AIC	*White noise*
SARIMAX(2,1,2)(1,1,1)6 -Temperature	-740,673	1497,345	Yes
SARIMAX(2,1,2)(1,1,1)6 -Humidity	-745,3	1506,601	Yes
SARIMAX(2,1,2)(1,1,1)6 -Temperature & Humidity	-743,625	1505,25	Yes
SARIMAX(2,1,2)(1,1,1)12 -Temperature	-637,162	1290,324	Yes
SARIMAX(2,1,2)(1,1,1)12 -Humidity	-633,151	1282,302	Yes
SARIMAX(2,1,2)(1,1,1)12 -Temperature & Humidity	-636,139	1290,278	Yes

3.3 Simple Prediction

Simple prediction using the average in the same month is done by taking the average value every month from the last 5 years (2014-2018). The average value of each month in 2014-2018 can be seen in **Table 4.**

Table 4. *Average diarrhea sufferers in 2014-2018.*

Month	Diarrhea sufferers
January	4618
February	4759
March	4386
April	4316
May	4222
June	3915
July	4427
August	4756
September	4923
October	5059
November	5167
December	4796

Figure 8 is showing the SARIMA (2, 1, 2)(1, 1, 1) prediction graph, meanwhile **Figure 9** is showing SARIMA (2, 1, 2)(1, 1, 1)12 prediction graph. For SARIMA (2, 1, 2)(1, 1, 1)6 with a temperature prediction graph see **Figure 10**, and the one with humidity prediction see **Figure 11**. **Figure 12** is SARIMAX (2, 1, 2) (1, 1, 1)6 with temperature and humidity prediction graph.

Figure 13 shows SARIMA (2, 1, 2)(1, 1, 1)12 with a temperature prediction graph, for the one with humidity prediction sees **Figure 14**. The prediction of SARIMA (2, 1, 2)(1, 1, 1)12 with temperature and humidity prediction see **Figure 15**.

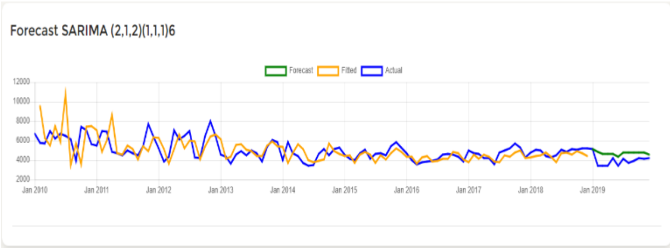

Figure 8. *SARIMA(2, 1, 2)(1, 1, 1)$_6$ prediction graph.*

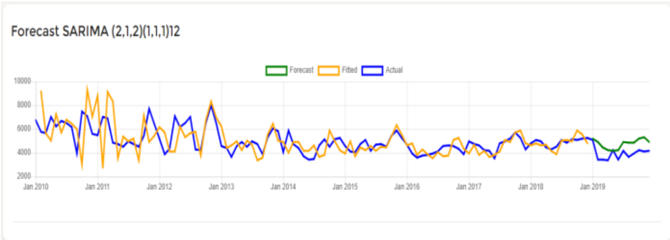

Figure 9. *SARIMA(2, 1, 2)(1, 1, 1)$_{12}$ prediction graph.*

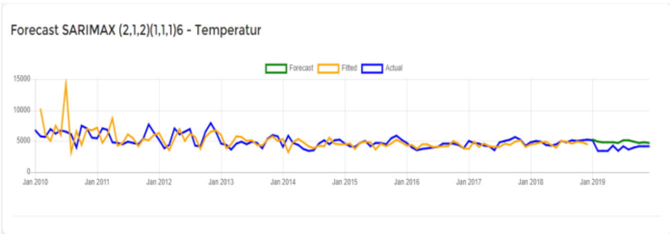

Figure 10. *SARIMAX(2,1,2)(1,1,1)$_6$ with temperature prediction graph.*

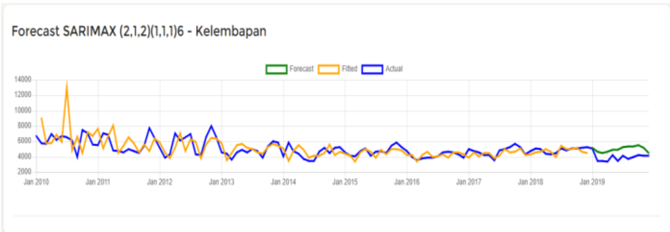

Figure 11. *SARIMAX(2,1,2)(1,1,1)$_6$ with humidity prediction graph.*

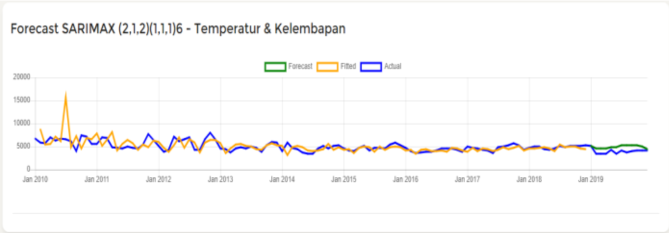

Figure 12. *SARIMAX(2,1,2)(1,1,1)$_6$ with temperature & humidity prediction graph.*

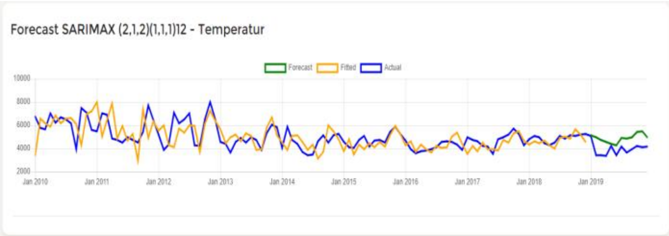

Figure 13. *SARIMAX(2,1,2)(1,1,1)$_{12}$ with temperature prediction graph.*

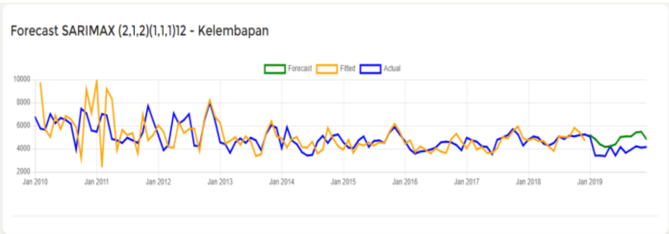

Figure 14. *SARIMAX(2,1,2)(1,1,1)$_{12}$ with humidity prediction graph.*

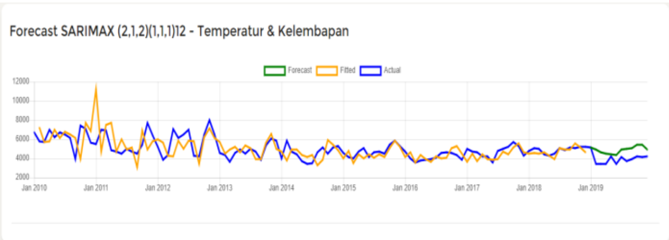

Figure 15 *SARIMAX(2,1,2)(1,1,1)$_{12}$ with temperatur & humidity prediction graph.*

3.4 Validation Results

Model accuracy validation was done using MAPE and RMSE calculations on the 2019 prediction results of the actual data in 2019. Furthermore, analyzing the ability of the model to follow the pattern of actual data. Validation results can be seen in **Tables 4** and **Table 5.**

Table 4. *Prediction accuracy.*

Model	Prediction	
	RMSE	MAPE
SARIMA(2,1,2)(1,1,1)$_6$	884,38	21,40
SARIMA(2,1,2)(1,1,1)$_{12}$	935,44	22,37
SARIMAX(2,1,2)(1,1,1)$_6$ -Temperature	1058,71	26,01
SARIMAX(2,1,2)(1,1,1)$_6$ -Humidity	1144,95	27,97
SARIMAX(2,1,2)(1,1,1)$_6$ -Temperature & Humidity	1072,14	26,07
SARIMAX(2,1,2)(1,1,1)$_{12}$ -Temperature	1042,56	25,17
SARIMAX(2,1,2)(1,1,1)$_{12}$ -Humidity	1016,13	24,16
SARIMAX(2,1,2)(1,1,1)$_{12}$ -Temperature & Humidity	1046,05	25,41
Simple prediction	833,52	19,80

Simple prediction using the average in the same month produces smaller error values compared to the SARIMA and SARIMAX models, but in the ability to predict results follow the actual data patterns, the SARIMA and SARIMAX models show better results.

Table 5. *The ability of the model to follow the pattern.*

Model	Prediction
SARIMA(2,1,2)(1,1,1)6	72,73%
SARIMA(2,1,2)(1,1,1)12	54,55%
SARIMAX(2,1,2)(1,1,1)6 -Temperature	54,55%
SARIMAX(2,1,2)(1,1,1)6 -Humidity	72,73%
SARIMAX(2,1,2)(1,1,1)6 -Temperature & Humidity	36,36%
SARIMAX(2,1,2)(1,1,1)12 -Temperature	72,73%
SARIMAX(2,1,2)(1,1,1)12 -Humidity	63,64%
SARIMAX(2,1,2)(1,1,1)12 -Temperature & Humidity	63,64%
Simple prediction	54,54%

4. Conclusions

After conducting this research, authors reached several conclusions. From the data collection, temperature and humidity do not have a significant relationship with the case of diarrhea in Bandung. Nonetheless, we build a model to predict the diarrhea sufferers using SARIMA and SARIMAX model. In comparison with the model SARIMAX, SARIMA models produce better accuracy. The best model for prediction is the SARIMA model (2,1,2) (1,1,1) 6 with MAPE = 21,40 and the ability to follow the pattern of 72.73%. For the future research we propose to use more diarrhea data set and consider spatial information in building the model prediction. We propose that future work, based on the allegation that the spatial information might has more role for sanitary condition rather than temperature or humidity.

References

[1] World Health Organization. (2017). *Diarrhoeal disease*. World Health Organization.

[2] Troeger, C., Blacker, B. F., Khalil, I. A., Rao, P. C., Cao, S., Zimsen, S. R., ... & Alvis-Guzman, N. (2018). Estimates of the global, regional, and national morbidity, mortality, and aetiologies of diarrhoea in 195 countries: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Infectious Diseases*, *18*(11), 1211-1228.

[3] Schiller, L. R., Pardi, D. S., Spiller, R., Semrad, C. E., Surawicz, C. M., Giannella, R. A., ... & Sellin, J. H. (2014). Gastro 2013 APDW/WCOG Shanghai working party report: chronic diarrhea: definition, classification, diagnosis. *Journal of gastroenterology and hepatology*, *29*(1), 6-25.

[4] Chou, W. C., Wu, J. L., Wang, Y. C., Huang, H., Sung, F. C., & Chuang, C. Y. (2010). Modeling the impact of climate variability on diarrhea-associated diseases in Taiwan (1996–2007). *Science of the Total Environment*, *409*(1), 43-51.

[5] Bandyopadhyay, S., Kanji, S., & Wang, L. (2012). The impact of rainfall and temperature variation on diarrheal prevalence in Sub-Saharan Africa. *Applied Geography*, *33*, 63-72.

[6] Chowdhury, F. R., Ibrahim, Q. S. U., Bari, M. S., Alam, M. J., Dunachie, S. J., Rodriguez-Morales, A. J., & Patwary, M. I. (2018). The association between temperature, rainfall and humidity with common climate-sensitive infectious diseases in Bangladesh. *PLoS One*, *13*(6).

[7] Omane-Adjepong, M., Oduro, F. T., & Oduro, S. D. (2013). Determining the better approach for short-term forecasting of ghana's inflation: Seasonal ARIMA Vs holt-winters. *International Journal of Business, Humanities and Technology*, *3*(1), 69-79.

[8] Da Veiga, C. P., Da Veiga, C. R. P., Catapan, A., Tortato, U., & Da Silva, W. V. (2014). Demand forecasting in food retail: A comparison between the Holt-Winters and ARIMA models. *WSEAS transactions on business and economics*, *11*(1), 608-614.

[9] Abdullah, A. G., Hasan, B., Mulyadi, Y., Hakim, D. L., & Riza, L. S. (2017). Analysis on anomalous short term load forecasting using two different approaches. In *2017 3rd International Conference on Science in Information Technology (ICSITech)* (pp. 573-576). IEEE.

[10] Ediger, V. Ş., & Akar, S. (2007). ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy policy*, *35*(3), 1701-1708.

[11] Sharafi, M., Ghaem, H., Tabatabaee, H. R., & Faramarzi, H. (2017). Forecasting the number of zoonotic cutaneous leishmaniasis cases in south of Fars province, Iran using seasonal ARIMA time series method. *Asian Pacific journal of tropical medicine*, *10*(1), 79-86.

[12] Kumar, S. V., & Vanajakshi, L. (2015). Short-term traffic flow prediction using seasonal ARIMA model with limited input data. *European Transport Research Review*, *7*(3), 21.

[13] Kongcharoen, C., & Kruangpradit, T. (2013). Autoregressive integrated moving average with explanatory variable (ARIMAX) model for Thailand export. In *33rd International Symposium on Forecasting, South Korea*, 1-8.

[14] Anggraeni, W., Vinarti, R. A., & Kurniawati, Y. D. (2015). Performance comparisons between arima and arimax method in moslem kids clothes demand forecasting: Case study. *Procedia Computer Science*, *72*, 630-637.

[15] Victor-Edema, U. A., & Essi, I. D. (2016). Autoregressive integrated moving average with exogenous variable (ARIMAX) model for Nigerian non-oil export. *European Journal of Business and Management*, *8*(36), 29-34.

[16] Vagropoulos, S. I., Chouliaras, G. I., Kardakos, E. G., Simoglou, C. K., & Bakirtzis, A. G. (2016). Comparison of SARIMAX, SARIMA, modified SARIMA and ANN-based models for short-term PV generation forecasting. In *2016 IEEE International Energy Conference (ENERGYCON)*, 1-6.

[17] Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A., ... & Chiroma, H. (2016). The role of big data in smart city. *International Journal of Information Management*, *36*(5), 748-758.

[18] Wei, W. W. (2006). Time series analysis. In *The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2*.

[19] Salleh, F. H. M., Arif, S. M., Zainudin, S., & Firdaus-Raih, M. (2015). Reconstructing gene regulatory networks from knock-out data using Gaussian Noise Model and Pearson Correlation Coefficient. *Computational biology and chemistry*, *59*, 3-14.

[20] Cryer, J. D., & Chan, K. S. (2008). Time Series Analysis with Applications in R Second Edition Springer Science+ Business Media.

[21] Hassani, H., & Yeganegi, M. R. (2020). Selecting optimal lag order in Ljung–Box test. *Physica A: Statistical Mechanics and its Applications*, *541*, 123700.

[22] John, E. (2016). Short-Term Forecasting of Nigeria Inflation Rates Using Seasonal ARIMA Model. *Sci. J. Appl. Math. Stat.,* 4(3), 101.

[23] Milenković, M., Švadlenka, L., Melichar, V., Bojović, N., & Avramović, Z. (2018). SARIMA modelling approach for railway passenger flow forecasting. *Transport, 33*(5), 1113–1120, 2018.