Implementation of Least Mean Square Adaptive Algorithm on Covid-19 Prediction

Sri Arttini Dwi Prasetyowati[1], Munaf Ismail[2], Badieah[3]

[1,2,3]Universitas Islam Sultan Agung Semarang, Indonesia

[1]arttini@unissula.ac.id, [2]munaf@unissula.ac.id, [3]badieah.assegaf@unissula.ac.id

Abstract - This study used Corona Virus Disease-19 (Covid-19) data in Indonesia from June to August 2021, consisting of data on people who were infected or positive Covid-19, recovered from Covid-19, and passed away from Covid-19. The data were processed using the adaptive LMS algorithm directly without pre-processing cause calculation errors, because covid-19 data was not balanced. Z-score and min-max normalization were chosen as pre-processing methods. After that, the prediction process can be carried out using the LMS adaptive method. The analysis was done by observing the error prediction that occurred every month per case. The results showed that data pre-processing using min-max normalization was better than with Z-score normalization because the error prediction for pre-processing using min-max and z-score were 18% and 47%, respectively.

Keywords: LMS, min-max, z-score, error prediction

I. INTRODUCTION

The Covid-19 pandemic was still ongoing for more than a year. Many predictive processes were conducted for decision making. They were conducted with reliable methods. The prediction was the most challenging process in data mining. Some of the research on the prediction of Covid-19 can be explained below.

Research about the transmission process of Corona Virus Disease based on real data modeling concluded that forward prediction and backward inference analysis of epidemic situations helped in decision making, because the difference between the model and real data was quite small [1].

Research that implements the SIR (Susceptible, Infected, and Removed) model to predict the situation of the Covid-19 outbreak in Indonesia based on data as of July 18, 2020, estimates that the peak number of infected in October 2020 is approximately 14% of the total infected population, and a relative MSE of 18.42. against the actual data period [2].

The prediction of covid-19 using the exponential smoothing method through a time series approach has been carried out and the best parameter values have been determined [3]. The prediction of daily Covid-19 data for the South Sulawesi region by comparing the ARIMA, Holt Winters and NAR-NN models has been carried out. Prediction results have a fairly large level of accuracy. The NAR-NN model has better predictive accuracy than the ARIMA and Holt Winters models [4].

Prediction using the LMS adaptive algorithm has been performed on very few previous studies. Several studies showed that most of the adaptive LMS algorithm studies always performed data pre-processing or modifying the algorithm. Several studies that showed collaboration between adaptive LMS algorithms and other algorithms for pre-processing can be explained as follows.

Research showed that for weak and monotonous ECG signals, the basic LMS algorithm was quite reliable to use. However, in the presence of interference, the LMS algorithm had to be modified to eliminate interference in the ECG signal so that the signal reading was more accurate [5]. Other research also proved that for random or monotonous data, the LMS algorithm must be modified further to improve the results [6], for handling imbalanced data and also detecting outliers for kNN classifier can also use hybrid pre-processing technique [7].

The Adaptive LMS algorithm was an algorithm that was relatively reliable in achieving variable targets, because the algorithm was simple and the achievement of the target becomes sooner or later depending on the step size value [8]. The advantage of the Adaptive LMS method was simple and robust in the calculation [9-11]. This simplicity in the calculation made the LMS algorithm run slowly but was quite reliable in its calculation results. The weakness of the LMS algorithm was the simplicity of the calculation which made the algorithm take a long time to process. However, this weakness can be anticipated by performing pre-processing or modifying the algorithm, as has been done by several previous researchers.

This research implemented adaptive LMS algorithms for Covid-19 data using z-score and min-max normalization as a pre-processing. Therefore, the process

could be run quickly and quite reliable in achieving variable targets.

II. METHOD

The LMS algorithm was not carried out alone because Covid-19 data cannot be processed directly. It needed a process that can make the Covid-19 data become balanced data. All steps can be explained in Fig. 1. Starting with the Covid-19 data input. Before running the adaptive LMS Algorithm, the data pre-processing was first conducted with z-score or min-max normalization (in Fig. 1 used z-score normalization and could be replaced by min-max normalization). The data consisted of the data on the infected people (positively confirmed with Covid-19), people recovering from Covid–19, and death rate on Covid–19 patients. After that, the adaptive LMS process was executed and errors were recorded to analyse the average error per month.
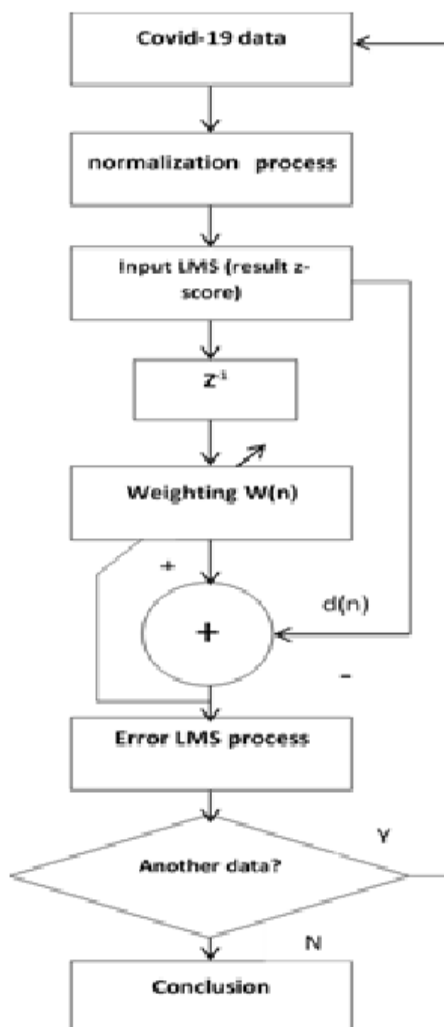

Fig. 1 Flowchart of research

A. *Z-score dan Min-max normalization*

The z-score method referred to the normal curve and involved only the mean and standard deviation. The formula for the Z-score method is (1).

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

where x is the original data and z is the resulting z-score, μ is the mean, and σ is the standard deviation.

The min-max normalization method determines the new minimum and new maximum values first so that the data were collected in the range of these values. The min-max normalization method formula is (2).

$$v' = \frac{v - min}{max - min}(new\_max - new\_min) \tag{2}$$

v is the data to be normalized and v′ is the result of normalization.

Min dan max are the smallest and the greatest data, new_min and new_max are desired maximum and minimum values.

B. *Adaptive Least Mean Square (LMS) Algorithm*

The block diagram of the predictive adaptive LMS algorithm is shown in Fig. 2.

The values of d and x are target and input, respectively. Both are the same data but input x was a target that was delayed by one sample [4][5]. The adaptive LMS algorithm usually used with the linear combiner shown in Fig. 3. With the input vector $x_k = [x_{1k} x_{2k} \cdots x_{Lk}]^T$ and the desired response is $d_k \in R$. The weight, output and error equations are expressed in the equation, respectively (3), (4) and (5) the following [12]:

$$w_k = [w_{1k} w_{2k} \dots w_{Lk}]^T \tag{3}$$

where wk is the weight vector and y_k is the output.

$$y_k = x_k^T w_k \tag{4}$$

and *error* ε_k:

$$\varepsilon_k = d_k - y_k \tag{5}$$

Mean Square Error (MSE) defined as (6)

$$\xi_k = [\varepsilon_k^2] \tag{6}$$

Optimum Weight $w*$ is (7) and (8)

$$w^* = R^{-1}p \tag{7}$$

where

$$R = E[x_1 x_k^T] \tag{8}$$

is a autocorrelation matrix, whereas (9)
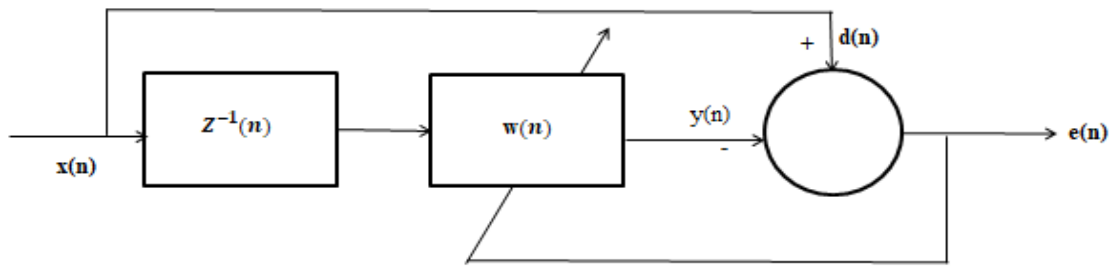
$$p = E[x_1 d_k] \tag{9}$$

Fig. 2 LMS adaptive block diagram

is a cross-correlation matrix and the weight formula is stated simply in (10).

$$w_{k+1} = w_k + 2\mu\varepsilon_k x_k \qquad (10)$$

The value of the step size μ and the initial weight w_0 could be determined first. The MSE value series ξ_k corresponds to w_k forming a learning curve. The adaptive linear combiner is shown in Fig. 3 [9], [13-15].

III. RESULTS AND DISCUSSION

The Results and Discussion section begins with the spread of Covid-19 data from June to August 2021, followed by the normalization and adaptive LMS processes. Analysis of the results obtained will determine the best data pre-processing carried out before the core process, namely the prediction process with Adaptive LMS.

A. Covid-19 Data Distribution

Data was obtained from the Web [16] from June until August 2021. The data had a fairly varied distribution. The data on the people infected with Covid-19 was in the range that was almost the same as data on the people recovering from Covid-19. The death rate data were very small and significantly different from the two previous data. Fig. 4 shows the overall data to be processed.
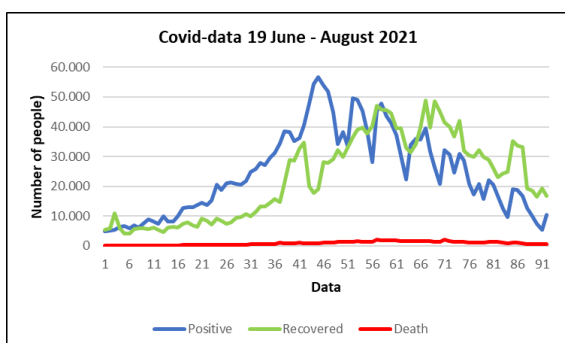

Fig. 4 Covid-19 data for three months

Fig. 5 showed data on the people infected with Covid–19, recovering from Covid–19 and the death rate which had been normalized with Z-Score. From Fig. 5. The normalized z-score data was easier to compare than the original data because it was located in the same data range. The data showed a very high increase on June 3, 2021. Apart from that date the data showed a gradual increase. In July, the data was very oscillating on positive cases. Meanwhile, the death cases often exceeded for the recovered cases. There was a significant increase in July 2021. In August, the conditions were inversely proportional from July 2021, there was a significant decrease even though the death cases still exceeded those of recovered cases at the beginning to the middle of the month, but near the end of the month, the recovered cases had exceeded the death cases.

Fig. 6 showed the results of min-max normalization method. Graphically, the min-max normalization was the same as the z-score normalization. The difference data in min-max normalization was collected in intervals zero until one, while for z-score normalization the data was in the negative, zero, and positive values. The increase and decrease are the same as the normalized z-score. However, even though it looked the same, the number of people represented by the normalization value can affect the predictive process. The prediction process was carried out using an adaptive LMS algorithm and discussed in section B for z-score normalization and section C for min-max normalization.

B. Prediction process using Adaptive LMS for z-score normalized data.

Figures of prediction results for positive, recovered, and death cases can be seen in Fig. 7, 8, and 9 respectively.
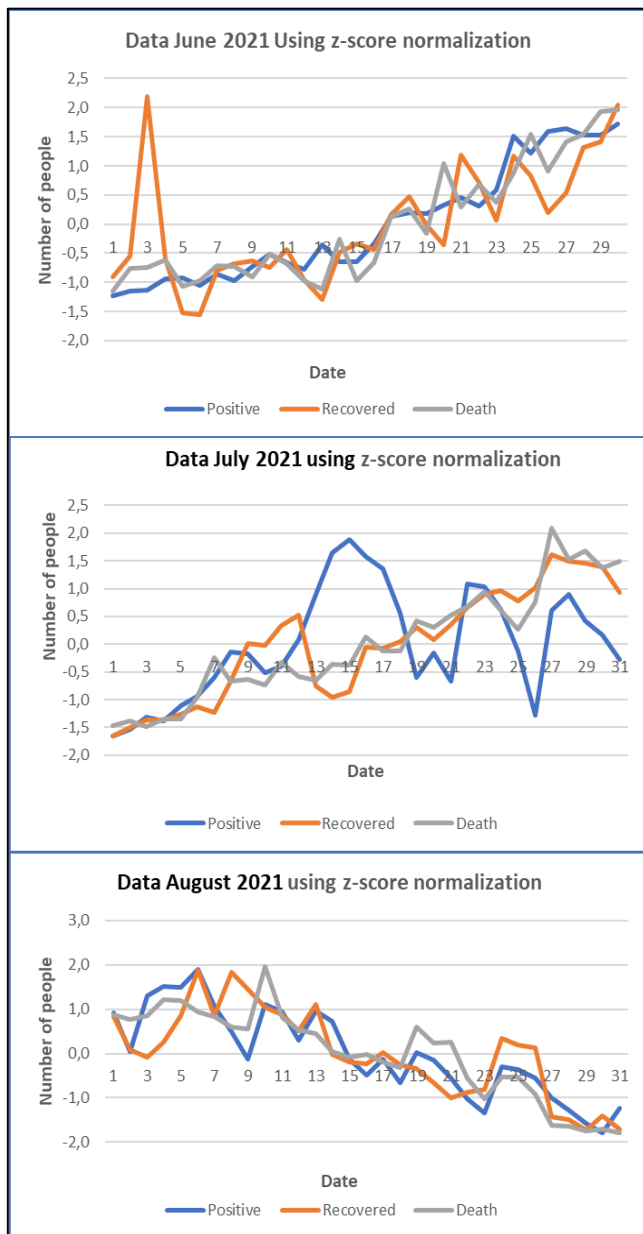
Fig. 5 Results from z-score for each month for three cases

Fig. 6 Results from min-max for each month for three cases

In Fig. 7, the curve obtained after the adaptive LMS tried to converge to the z-score curve. In July 2021, the adaptive LMS process looked very difficult to converge to the z-score data, but the robust LMS algorithm can still form a curve similar to the z-score curve.

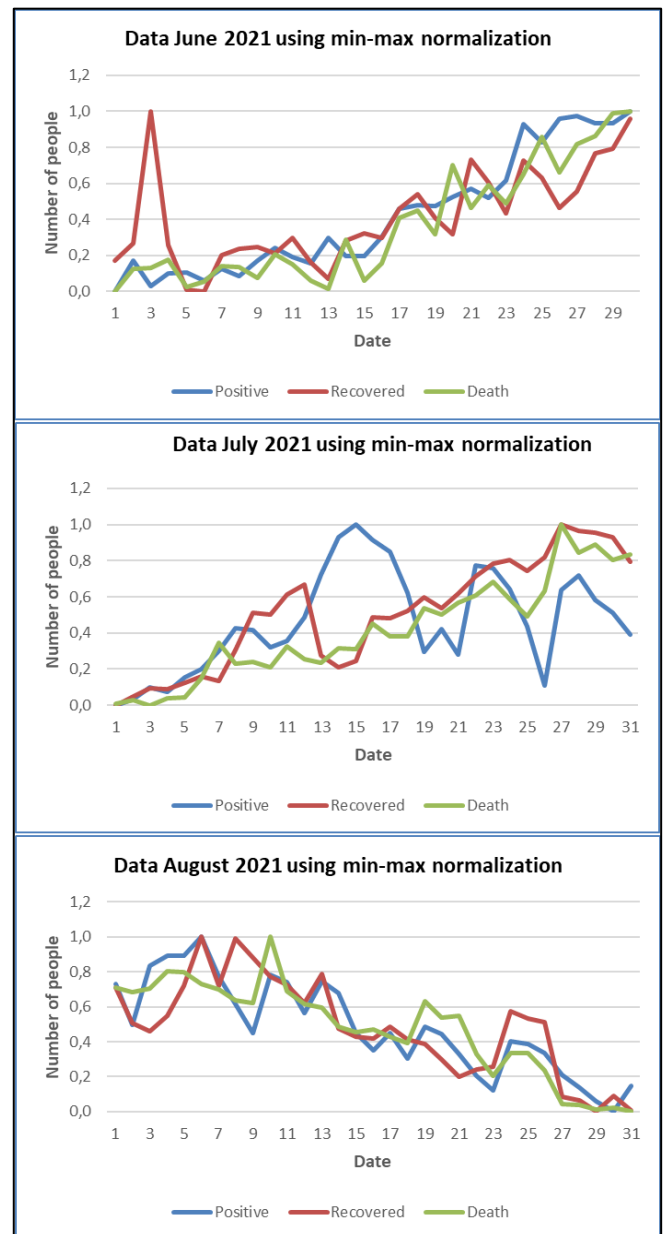The average error was still better in July and August for recovered cases when compared to Fig. 7 (positive cases) and Fig. 8 (recovered cases). However, in June, the recovered cases had a much higher error than the positive infected cases, so the total average error was still better for the infected cases.
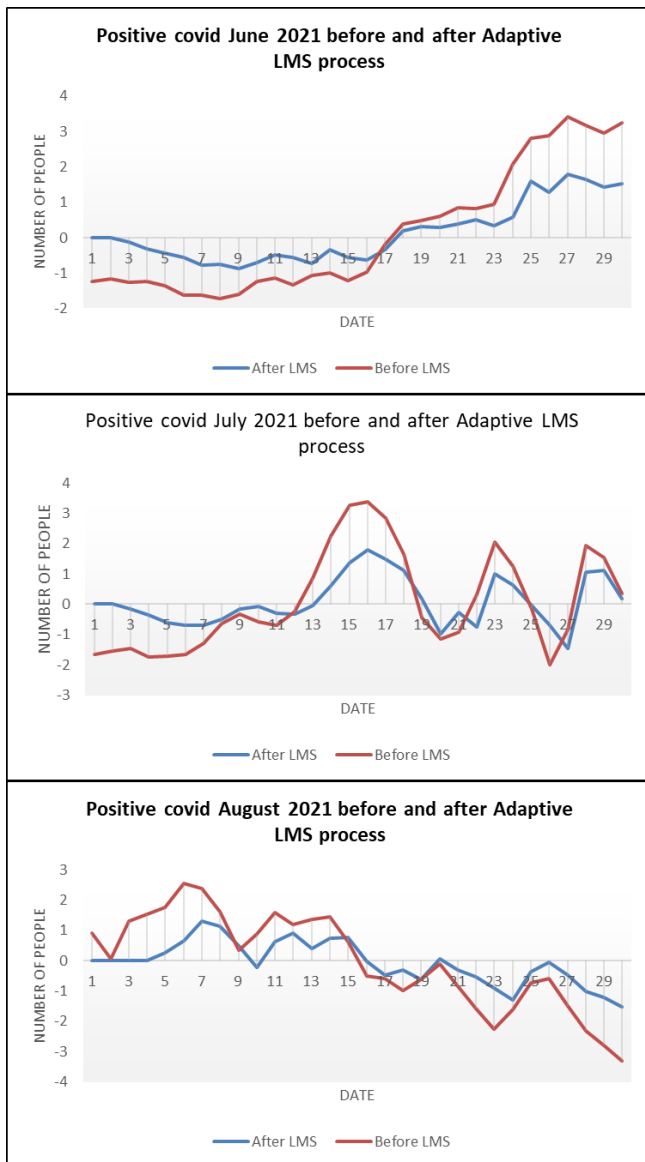
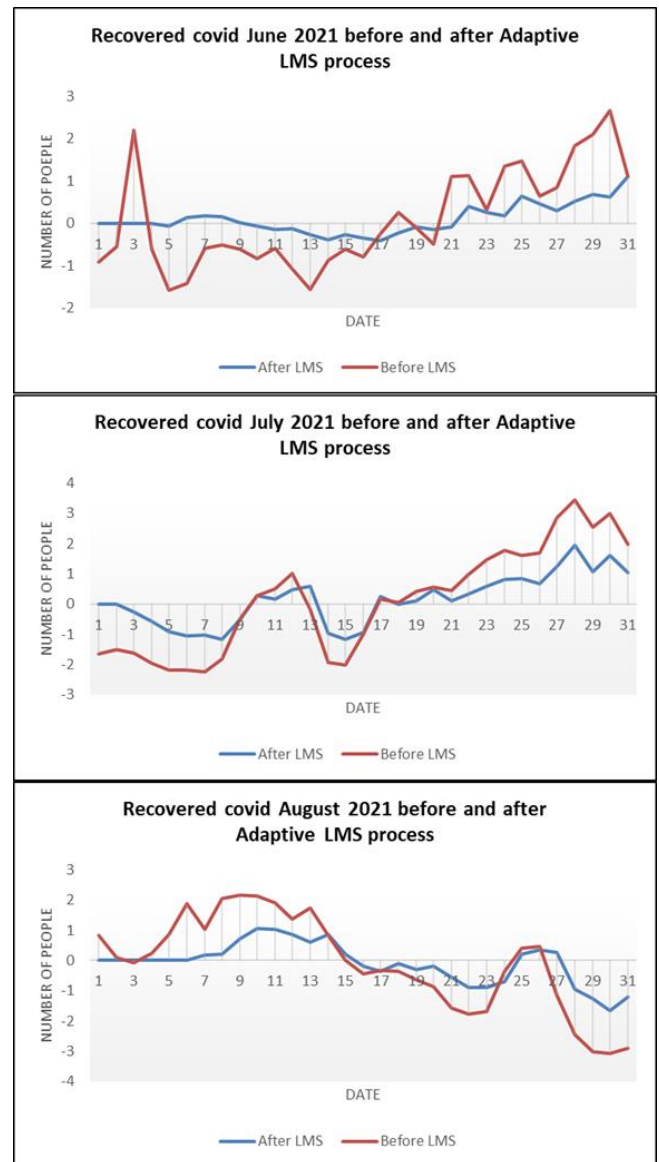**Fig. 7 Positive case before and after LMS adaptive (z-score)**

Fig. 8 Recovered case before and after LMS adaptive (z-score)

For death cases (Fig. 9), the largest error occurred in June 2021, but when compared with positive cases and recovered cases, death cases had the smallest average error. However, these three cases still have relatively high errors, far from the acceptable error value below 20%.

C. Predictive LMS adaptive process for min-max normalized data.

The same process was carried out for data that has been processed with min-max normalization. The results can be seen in Fig. 10, 11, and 12, respectively, for positive, recovered, and death cases. The three prediction result curves (after the LMS process) in Fig. 10 tried to converge to the original data, but in July 2021, the curve still oscillates following the original curve. Furthermore, for recovered cases (Fig. 11), the prediction curve will coincide at the end of the month in June and August. Hopefully, it will coincide with the curve before the LMS process in the next month. For Death case, the process towards convergence has started since the beginning of the process.
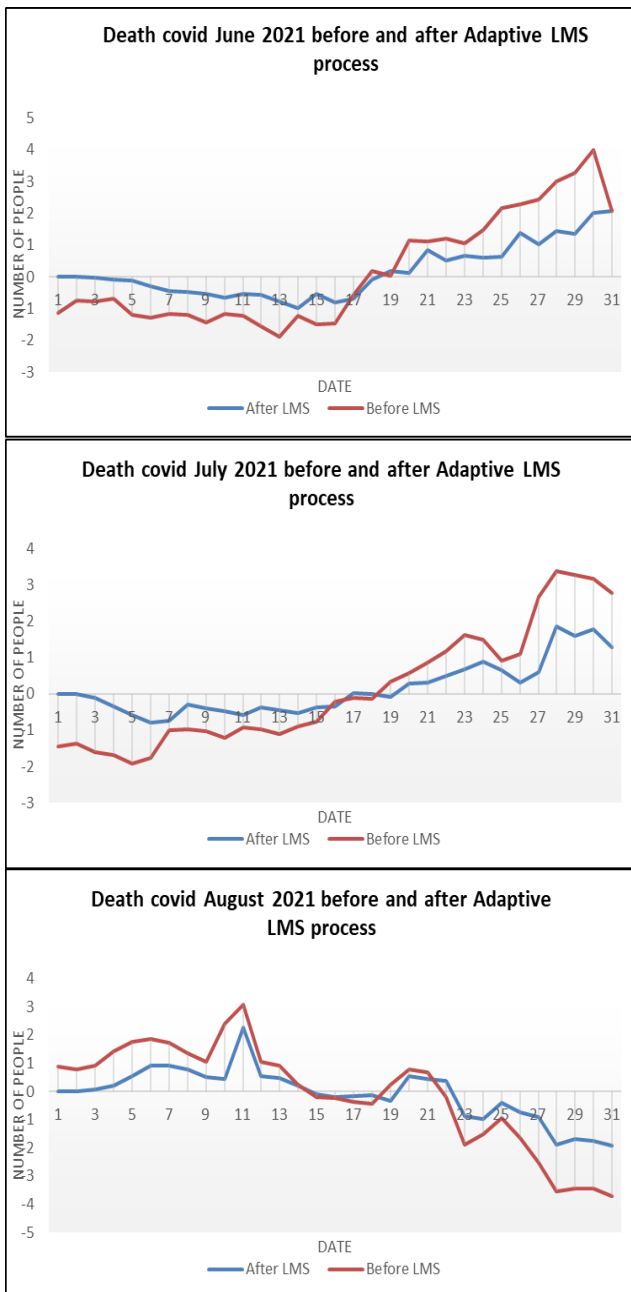
Fig. 9 Death case before and after LMS adaptive (z-score)

Fig. 10 Positive case for min-max

D. *Comparison of predictive LMS Adaptive Processing with z-score and min-max normalization pre-processing*.

From the results of the prediction process using adaptive LMS, it was found that the average error difference was quite significant between predictive results with data pre-processing using z-score normalization and using min-max normalization. These differences can be observed in Table I and II.
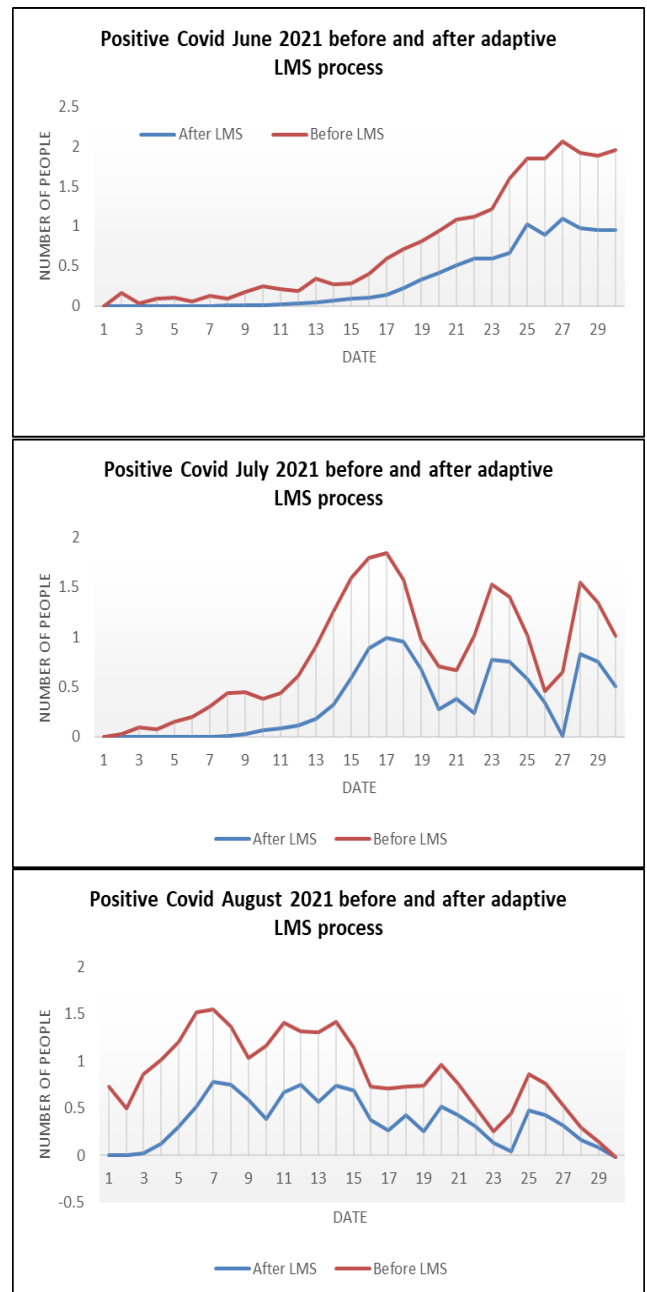
TABLE I
PREDICTION RESULTS WITH Z-SCORE
NORMALIZATION

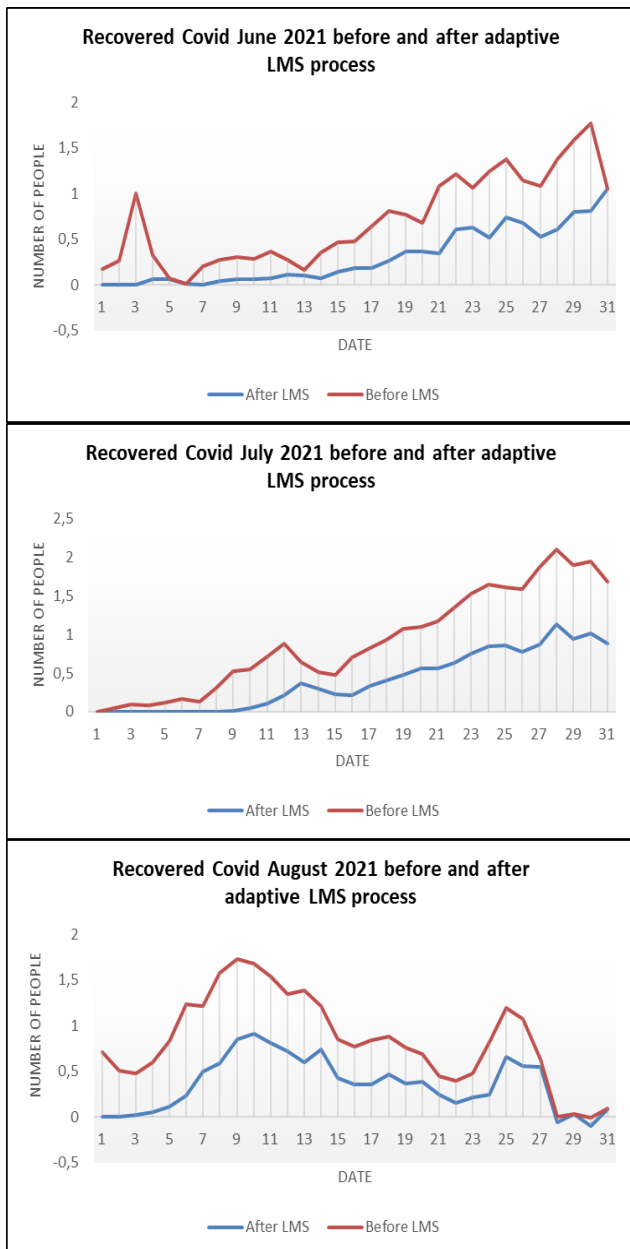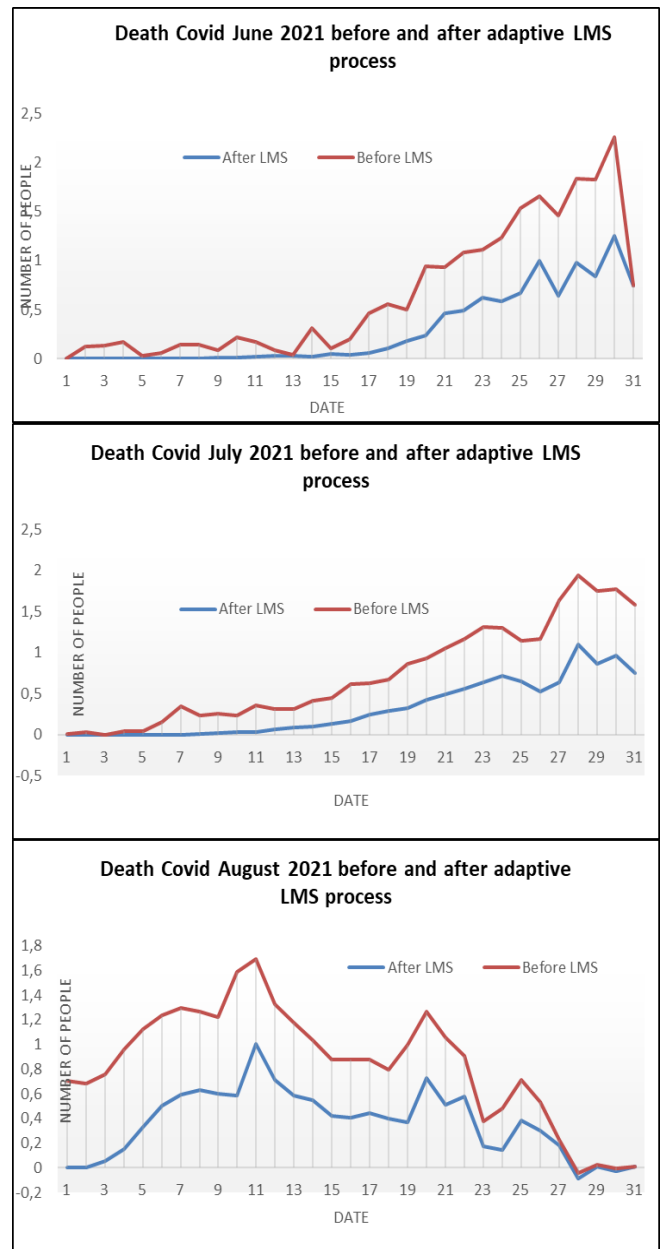| Month	Mean error using z-score normalization (%)			Average (%)
	Positive	**Recovered**	**Death**	
Jun-21	29	71	49	50
Jul-21	55	36	38	43
Aug-21	59	51	37	49
Average (Mean)	48	53	41	47

Fig. 11 Recovered case with min-max

Fig. 12 Death case with min-max

TABLE II
PREDICTION RESULTS WITH MIN-MAX
NORMALIZATION

Month	Mean error using min-max normalization (%)			Average (%)
	Positive	**Recovered**	**Death**	
Jun-21	13	20	17	17
Jul-21	25	16	16	19
Aug-21	20	20	16	19
Average (Mean)	19	19	16	18

Table I and II show that the average error for LMS adaptive prediction using min-max normalization was better than LMS adaptive prediction using z-score normalization, with the difference reaching 29%. In each case, whether it was positive, recovered, or death cases, the LMS adaptive predictive process using min-max normalization was better than using z-score normalization. Positive, recovered, and death cases with normalized z-score have a mean error of 48%, 53%, and 41%, respectively. While positive, recovered, and death cases for min-max normalization had a mean error of 19%, 19% and 16%, respectively. Therefore, the mean

error for all cases using min-max normalization as a data pre-processing was 18%, while z-score normalization reached 47%.

IV. CONCLUSION

From this research, it can be concluded that the prediction process using the LMS adaptive algorithm required data pre-processing to be applied to Covid-19 data with various conditions. Z-score and min-max normalization were used as data pre-processing. Research showed that min-max normalization was better than z-score because the mean error values were 18% and 47%, respectively for all cases.

REFERENCES

[1] L. Li, Z. Yang, Z. Dang, C. Meng, J. Huang, H. Meng, D. Wang, G. Chen, J. Zhang, H. Peng, and Y. Shao "Propagation analysis and prediction of the COVID-19," *Infect. Dis. Model.*, vol. 5, pp. 282–292, 2020.

[2] B. Wahyudi and I. Palupi, "Prediction of the peak Covid-19 pandemic in Indonesia using SIR model," *J. Teknol. dan Sist. Komput.*, vol. 0, 2021.

[3] C. M. Gibran, S. Setiyawati, and F. Liantoni, "Prediksi Penambahan Kasus Covid-19 di Indonesia Melalui Pendekatan Time Series Menggunakan Metode Exponential Smoothing," *J. Inform. Univ. Pamulang*, vol. 6, no. 1, p. 112, 2021.

[4] A. Asrirawan, A. Seppewali, and N. Fitriyani, "Model Time Series untuk Prediksi Jumlah Kasus Infeksi Coronavirus (Covid-19) di Sulawesi Selatan," *J. MSA (Mat. dan Stat. serta Apl.)*, vol. 8, no. 2, p. 78, 2020.

[5] A. Ren, Z. Du, J. Li, F. Hu, X. Yang, and H. Abbas, "Adaptive interference cancellation of ECG signals," *Sensors (Switzerland)*, vol. 17, no. 5, pp. 1–15, 2017.

[6] S. A. D. Prasetyowati and A. Susanto, "Multiple Processes for Least Mean Square Adaptive Algorithm on Roadway Noise Cancelling," *Int. J. Electr. Comput. Eng.*, vol. 5, no. 2, pp. 355–360, 2015.

[7] P. Nair and I. Kashyap, "Hybrid Pre-processing Technique for Handling Imbalanced Data and Detecting Outliers for KNN Classifier," *Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. Trends, Prespectives Prospect. Com. 2019*, pp. 460–464, 2019.

[8] S. A. D. Prasetyowati, H. Abbas, and N. B. S. Eka, "Solution for Vehicles Noise Cancellation With Modification of LMS Adaptive Algorithm," *Int. J. Comput. Sci. Eng. (IJCSE*, vol. 4, no. 5, pp. 929–937, 2012.

[9] A. Flores and B. Widrow, "Assesment of the efficiency of the LMS algorithm based on spectral information," *Conf. Rec. - Asilomar Conf. Signals, Syst. Comput.*, vol. 1, no. 1, pp. 120–124, 2004.

[10] T. L. and T. K. S. B. Jalal, X. Yang, Q. Liu, "Fast and Robust Variable-Step-Size LMS Algorithm for Adaptive Beamforming," *IEEE Trans. Instrum. Meas.*, vol. 19, no. 7, pp. 1206–1210, 2020.

[11] M. K. Eghbal and G. Alipoor, "LMSK: A robust higher-order gradient-based adaptive algorithm," *IET Signal Process.*, vol. 13, no. 5, pp. 506–515, 2019.

[12] S. A. D. Prasetyowati, A. Susanto, and I. Widihastuti, "Investigation of diesel's residual noise on predictive vehicles noise cancelling using LMS adaptive algorithm," in *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 2016.

[13] S. Gazor, "Prediction in LMS-type adaptive algorithms for smoothly time varying environments," *IEEE Trans. Signal Process.*, vol. 47, no. 6, pp. 1735–1739, 1999.

[14] J. R. Zeidler, "Performance Analysis of LMS Adaptive Prediction Filters," *Proc. IEEE*, vol. 78, no. 12, pp. 1781–1806, 1990.

[15] S. Prasetyowati, "Hybrid FLC-LMS Algorithm for Predicting Sediment Volume in the River," *Int. J. Intell. Eng. Syst.*, vol. 14, no. 2, pp. 395–409, 2021.

[16] Center for Data and Information of the Ministry of Indonesia, "The Pusdatin of the Ministry of HealthCenter for Data and Information of the Ministry of Health of the Republic of Indonesia," 2021.