Comparative analysis of time series prediction model for forecasting COVID-19 trend

Sri Ngudi Wahyuni[1], Eko Sediono[2], Irwan Sembiring[2], Nazmun Nahar Khanom[3]
[1]Department of Informatics Management, Universitas Amikom, Yogyakarta, Indonesia
[2]Department of Information Technology, Universitas Kristen Satya Wacana, Salatiga, Indonesia
[3]Department of Information and Communication Technology, Bangladesh University of Professionals, Dhaka, Bangladesh

Article Info

Article history:

Received Dec 30, 2021
Revised Jul 28, 2022
Accepted Aug 10, 2022

Keywords:

COVID-19 trends
Forecasting
Long short term memory
Multiple linear regression
Ridge regression

ABSTRACT

The outbreak of the COVID-19 pandemic occurred some time ago, making the world a pandemic. Based on this condition is important to predict early to prevent the COVID-19 disease if someday pandemic occurs. The aim of the study is to compare the analysis result of cumulative cases of COVID-19 using multiple linear regression (MLR), ridge regression (RR), and long short term memory (LSTM) models for cases study Java and Bali islands. We chose both islands as a case study because they have very dense populations. These three models are the most widely used time series-based prediction models and have relatively high accuracy values. The predictive variables used are the number of cumulative cases, the daily cases, and population density. The research data was taken from Kaggle and processed using google collabs. Data was taken from January 20, 2020, to August 8, 2020, and data training was carried out for 12 days. The results show the accuracy of LSTM is better than other models. it can be seen in the accuracy value (99.8%) of the model test result. The testing model uses R2, mean square error (MSE), and root mean square error (RMSE).

Corresponding Author:

Sri Ngudi Wahyuni
Department of Informatics Management, Universitas Amikom
St. North Ringroad, Depok, Condongcatur, Sleman, Yogyakarta, Indonesia
Email: yuni@amikom.ac.id

1. INTRODUCTION

Since COVID-19 occurred in Indonesia in February 2019, all activities have changed into regular habits. In 2020, Java and Bali islands had the highest spread of COVID-19 cases in Indonesia. Java and Bali's islands are Indonesia's most significant business and tourism destinations. Judging from the number of positive cases, both of the islands contributed 67.76% of the total national cases. In the following order, Sumatra, Kalimantan, Sulawesi, Nusa Tenggara, and Maluku-Papua are followed in the last order regarding the number of positive cases. This is because Java and Bali dominate the population in Indonesia and there is a capital city in it. Indonesia is a country with the 5th largest population in the world [1]. Recently the population in Indonesia has reached more than 270 million people and spread throughout the islands of Indonesia. Currently, the Indonesian government is using all efforts to suppress the rate of positive COVID-19 cases by boosting vaccinations. There was also chaos caused by the surge in COVID-19 patients in Indonesia. This condition should have been predictable before and taken the strategic way for dealing with the COVID-19 patients in Java and Bali islands. Prediction is one significant way of helping the government and others, mapping out and preparing health services in the pandemic era. Prediction and machine learning have a relationship in the process-relational approach, which improves processes, data quality, and model quality [2]. Prediction algorithms have been implemented by many researchers before. One of the machine learning algorithms for prediction is regression. Regression functions to predict the future by analyzing past events. It can be to minimize the risk or impact that will occur in the future, both short-term and long-term [3]. The data of the COVID-19 cumulative cases is time-series data. Because it has patterns depending on the previous time trend [4]. The research aim is to help the government anticipate the COVID-19 outbreak. Information and communication between researchers and the government are essential for decision-making and strategic planning for outbreak handlers. Strategic planning is for preparing hospital facilities, the immune system of the infected person socialization, steps taken to combat the proliferation of the virus, and so on to make it completely informative [5]. Several researchers compared linear regression and support vector machine (SVM) in machine learning-based for a number of cases prediction [6], [7] and the result is linear regression has better accuracy than the SVM algorithm. But linear regression only contains one variable, meanwhile MLR is able to contain two or more variables in the prediction. Based on it, several studies use multiple linear regression (MLR) for prediction, there are Rath *et al*. [8] compared a linear regression algorithm with MLR for the prediction of COVID-19 cases in India. The model was evaluated using R2 and the result was 0.995. However, this study did not explain the case prediction results for some time in the future, so it did not reveal the accuracy of the detailed historical model. Wahyuni *et al*. [9] use the MLR model for COVID-19 cases prediction in Indonesia. The model accuracy test used R2 and the results were 0.999. The other study that uses MLR is [10]. However, MLR is considered less suitable for prediction because it is often overfitting, so some studies use the ridge regression (RR) algorithm to avoid overfitting. Because RR applies regularization to the predictive variable coefficients, and in this way selects the coefficients in a way that is kept as low as possible. The effect predictive variable does not have a major effect on the outcome variable, based on it some studies use the RR for predicted cases [11]–[13]. Liu, compared linear regression, logistic, and recurrent neural network (RNN) models to predict the trend of COVID-19 in the US. The comparison results show that RNN is more accurate than the other two models [14].

RNN is a generalization of a feed-forward neural network that has internal memory. RNN is iterative because it performs the same function for each data input while the output of the current input depends on previous calculations. After generating output, it is copied and sent back to the network over and over. To make a decision, it considers the current input and the output that has been learned from the previous input. RNN in machine learning and deep learning is widely used to make various predictions, including weather predictions, stock prices [15] electrical diagram (ECG) recording [16] etc. RNN is also considered capable of predictions based on time series data. One of the problems with RNN is that the gradient disappears [17], [18]. To solve this problem, long short term memory (LSTM) is considered suitable for predicting time series [19]. LSTM is an RNN model development by adding one cell state, which functions to store time-series data for a long time [20], [21]. LSTM is often used to predict infectious diseases, prediction of dengue disease [22]–[24], mouth and foot disease [25], Hepatitis [26], [27], chickenpox [28], Malaria [29]–[31], Tubercolusis [32] and other infectious diseases like COVID-19 prediction. several studies on LSTM for COVID-19 prediction are Indriani *et al*. [33]. conducted research on LSTM models for COVID-19 prediction in Indonesia using epoch 50 and lookback 8 stating that the LSTM model is suitable for the forecasting model. Bedi *et al*. [34] compared the LSTM model with susceptible-exposed-infected-recovered-deceased (SEIRD) for the short-term prediction of the outbreak of COVID-19 cases in India. The data training is carried out for the next 30 days, the results show that the LSTM model is better and in accordance with COVID-19 cases in India than other models. Iqbal *et al*. [35] used the LSTM model for the prediction of COVID-19 cases in Bangladesh, and the results show that the LSTM model is recommended to be used as a prediction model because it has a high accuracy value, especially in the regression task. Rauf *et al.* [36] developed the LSTM model for prediction and the results showed that the LSTM has a high accuracy of 99,525 compared to other models. The other studies that used LSTM for COVID-19 prediction are [37]–[43].

Based on the literature above, no one has directly compared the accuracy of the MLR, Ridge, and LSTM models by adding the population density parameter. Because in some of the studies above no one has discussed the RR model for COVID-19 predictions. In this study, predictions of COVID-19 will be made using the RR, MLR, and LSTM models based on machine learning. These three models are considered to have very high accuracy values in terms of prediction. However, we will compare these three models for the predicted case of COVID-19 in developing countries such as Indonesia by using the population density variable on the islands of Java and Bali. The prediction results will be explained in detail, as a recommendation for decision-making in the future. The population density variable is a time series variable so it is suitable for the three prediction models. The next sections of this study will discuss the research and method in section 2, the results and discussion in section 3, the conclusion in section 4, and the references in the last part.

2. METHOD

2.1. Data set selection and processing

The data was taken by Kagle and taken from January 20, 2020, to August 1, 2020. The data training was carried out for 7 days to see the trend in the number of COVID-19 cases on the islands of Java and Bali in the next 4 days. The data is processed using google collabs using three variables as machine learning models, namely cumulative cases, daily new cases and population density on the islands of Java and Bali. We divide the data set into two, namely 80% as training data and 20% for model testing.

2.2. Multiple linear regression

The linear regression model is a simple prediction model in machine learning. This model just predicts using two parameters. With the development of linear regression, MLR can be used to expect more than two parameters [44], [45]. The MLR model can be shown in (1).

$$y = \beta_o x_1 + \beta_1 x_2 + \beta_2 x_3 + \varepsilon \tag{1}$$

Where in (3) is equivalent to (2).

$$E(y) = \beta_o x_1 + \beta_1 x_2 + \beta_2 x_3 + \beta_p x_p \tag{2}$$

Where p is the number of independent variables, y is a predictor x is the independent variable, β is the coefficient, and ε is constant [10], [46]. And the coefficient is shown in (3) and (4).

$$\beta_0 = \frac{n\sum XY-(\sum X)\sum Y}{n\sum X^2-(\sum X)^2} \tag{3}$$

$$\beta_p = \frac{\sum Y-\beta_0 \sum X}{n} \tag{4}$$

2.3. Ridge regression

RR is a technique to develop and stabilize the regression coefficient value because of multicollinearity. Multicollinearity is a strong correlation or relationship between two or more independent variables in a multiple regression model situation. This method is intended to overcome the bad conditions because of the high correlation between several independent variables in the regression model. In this case, the matrix to be nearly singular generates the estimated value of the model parameter unstable regression [12]. RR is a modification of the method of least squares which produces a biased estimator of the regression coefficient [13]. RR is a special algorithm of regression for multilinear regression information that has multicollinearity. The RR formula is presented in (5).

$$\cos(a) = \frac{1}{2}\sum_{p-1}^{p-b}\left(z_p - \hat{z}_i\right)^2 |\gamma \sum_q^a a_q^2 \tag{5}$$

Where $\gamma$is the slope and if **lambda** = 0, the RR is equal to least squares regression and when **lambda** = infinity, all coefficients shrink to zero.

2.4. LSTM

Hochreiter and Schmidhuber [28] have proposed LSTM to overcome the vanishing and exploding gradients problem [47]. The memory of the LSTM cell will be stored and converted from input to output in cell state. long short-term memory is a particular type of RNN. It has a better effect on time series prediction [48]. Entering the LSTM will first pass through the forgetting gate and then through the sigmoid layer, called the forgotten gate layer, indicating whether sort forgot to store in the last cell the decision on C_{t-1}, 0 means completely forgotten, 1 means completely ordered. Taking the current state h_{t-1} and the new input x_t as the input of this layer, the output is the value (0,1) [49], [50]. LSTM has four layers, namely forget gate (1), input gate (2), new cell state candidate (3), and output gate (4) in the model loop as shown in Figure 1.

In Figure 2, LSTM is defined in the following formula 8.

$$ft = (Wf.[h_{t-1}, x_t] + bf) \tag{6}$$

This decides which information can be transferred to the cell. The information from the previously ignored memory input is resolved by the forget gate and is defined as (7).

$$it = (Wi.[h_{t-1}, x_t] + bi) \tag{7}$$

$$\bar{C}t = \tanh\,(Wc.[h_{t-1}, x_t] + bc) \tag{8}$$

$$Ct = ft * C_{t-1} + it * \bar{C}t \tag{9}$$

$$ot = (Wo.[h_{t-1}, x_t] + bo \tag{10}$$

$$ht = ot * \tanh\,(Ct) \tag{11}$$

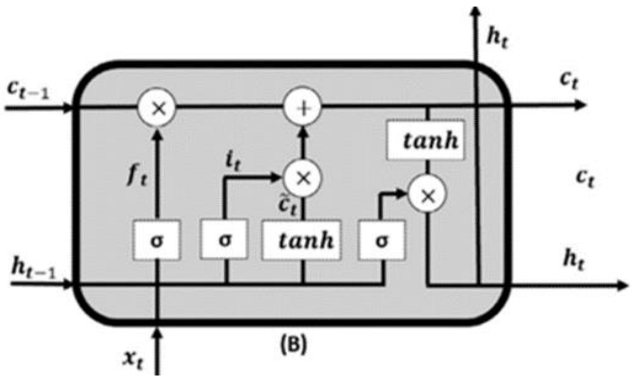

Figure 1. Looping in LSTM model

2.5. Proposed framework

In this study, three-time series prediction algorithms have been compared to obtain short-term predictions. Then the accuracy results of the three algorithms have been reached in order to get the best accuracy results among the three. The research was begun by collecting the dataset and the second process is to clean the data by eliminating the value 0. The data have been split into two parts, namely 80% training for data and 20% testing data. The next step was to predict using multiple linear regression, RR, and LSTM. The last process was parameter evaluation using the R^2, mean square error (MSE), and root mean square error (RMSE). The proposed framework is described in Figure 2.

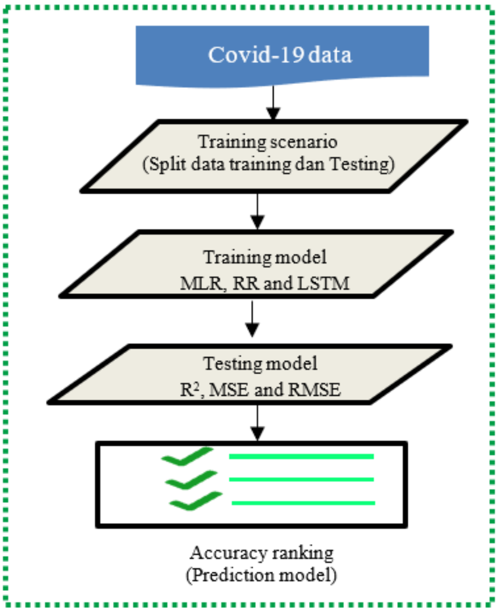

Figure 2. Proposed framework

3. RESULT AND DISCUSSION

3.1. Data set

The data set were taken from Kaggle and already used in the [9], [33]. Data is updated by the national disaster management agency (BNPB-Indonesia National Disaster Management Authority). This study only uses three parameters for predictive analysis, namely cumulative cases, daily new cases, and population density.

3.2. Evaluation metric

The R^2 score measures the relationship between the independent and dependent variables using the regression model [51]. The R^2 formula is:

$$R^2 = \frac{\text{SS}_{\text{REGRESSION}}}{\text{SS}_{\text{TOTAL}}} \tag{14}$$

the SSRegression is the sum of squares in the regression results, and SSTotal is the total number of all data. The second evaluation parameter used is mean square error or MSE is the following equation functions.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(yi - y)^2 \tag{15}$$

3.3. Prediction result in Java and Bali Island

We use the data from the Java and Bali islands because this area is the top cumulative new case of COVID-19 in Indonesia. Java and Bali are the largest tourist and business destinations in Indonesia and from outside Indonesia. The data was taken 12 days s for data train and we forecast for four days, from July 6, 202, until July 9, 2020. The initial process of experimentation is to perform feature extraction as input data. Then it is entered into the model to get the time series output. The final step is to compare the results of the accuracy of the three models. Cumulative case prediction results using RR are presented in Figure 3. Based on Figure 3 above, the prediction results using the RR model appear to have decreased from the actual data. But at the beginning of the prediction, the RR model has results that are close to the actual value. Short-term predictions are using RR model prediction results are presented in Table 1.

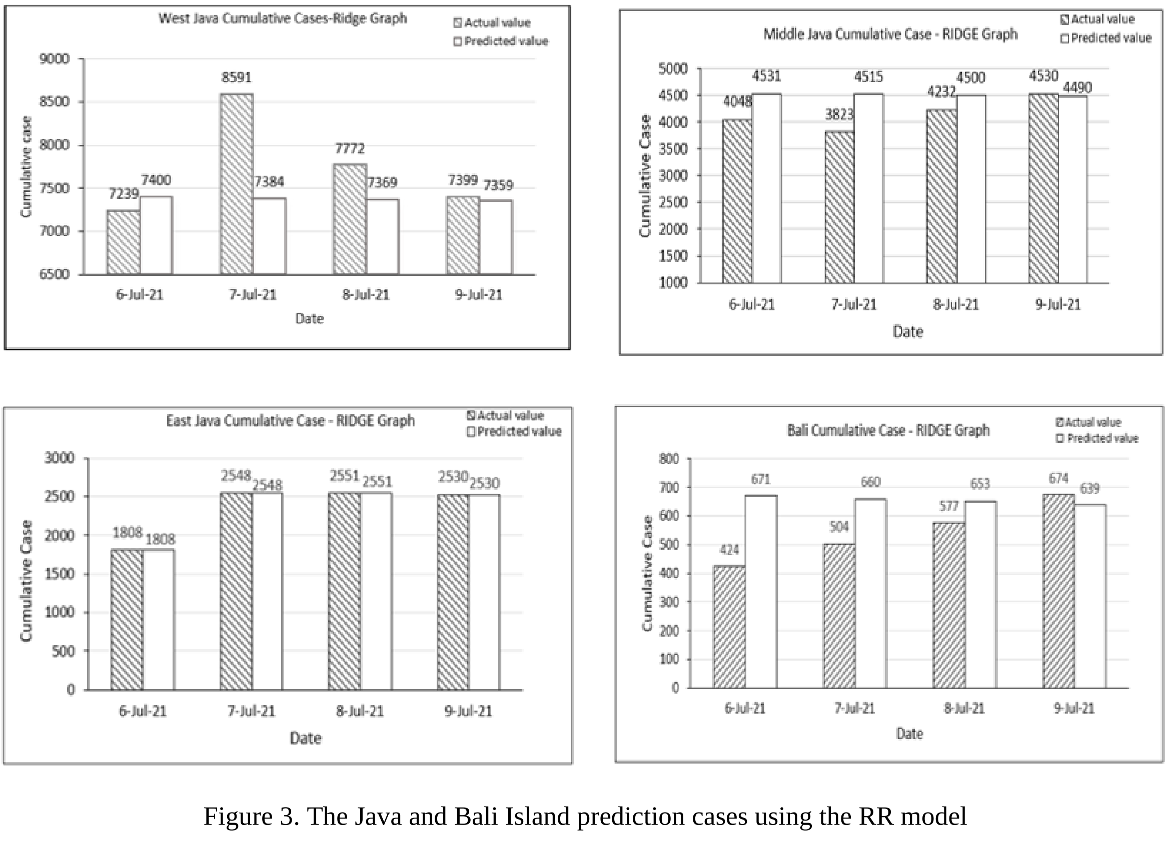

Figure 3. The Java and Bali Island prediction cases using the RR model

Table 1 is the results of the short-term prediction of 4 days using the RR model. The experiment results show, the difference of the actual and the predicted values are relatively small. However, the prediction results on the Bali Island have the difference value is very significant. This explain the RR model is not suitable for predicting the cumulative case on the Bali Island. Meanwhile, the prediction results using MLR are presented in Figure 4. Figure 4 is the cumulative cases predicting the result of Java and Bali islands using MLR. This result shows of the prediction value has a fairly good accuracy than RR prediction results. The prediction results detailed are described in Table 2.

Based on the predicted values in Table 2, it is shown of the predicted values are closer the actual values, especially the prediction results in 6th and 7th July 2021 for the Bali Island. This value is closer than the RR prediction results in the same dates. Based on this result, we can conclude that the multiple linear reression has a better validation level than the RR model. Next, we will compare the predicted results using the LSTM model presented in Figure 5.

Table 1. Prediction results using the RR model

Region	Date	Actual value	RR Predicted value
West Java	6th July 2021	7,239	7,400
	7th July 2021	8,591	7,384
	8th July 2021	7,772	7,369
	9th July 2021	7,399	7,359
Middle Java	6th July 2021	4,048	4,531
	7th July 2021	3,823	4,515
	8th July 2021	4,232	4,500
	9th July 2021	4,530	4,490
East Java	6th July 2021	1,808	1,808
	7th July 2021	2,548	2,548
	8th July 2021	2,551	2,551
	9th July 2021	2,530	2,530
Bali Island	6th July 2021	424	671
	7th July 2021	504	660
	8th July 2021	577	653
	9th July 2021	674	639

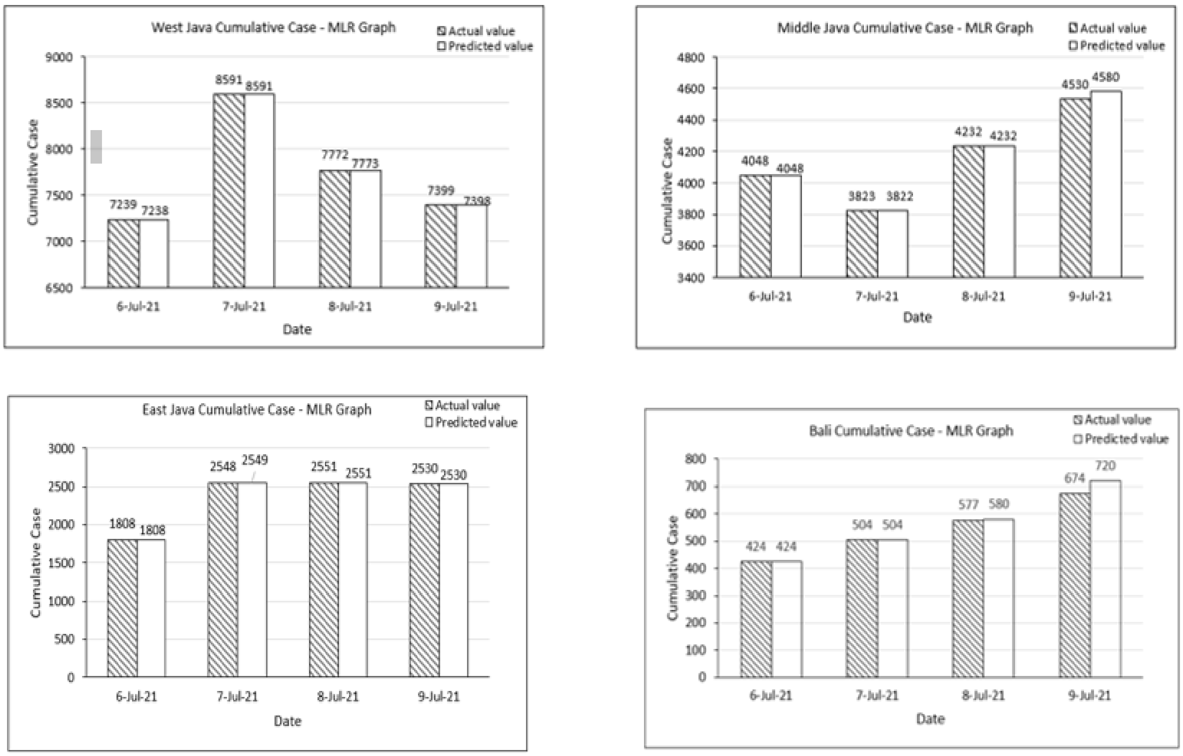

Figure 4. The Java and Bali Island prediction cases using the MLR model

Table 2. Prediction results of cumulative COVID-19 cases using the MLR model

Region	Date	Actual value	MLR Predicted value
West Java	6th July 2021	7,239	7,238
	7th July 2021	8,591	8,591
	8th July 2021	7,772	7,773
	9th July 2021	7,399	7,398
Middle Java	6th July 2021	4,048	4,048
	7th July 2021	3,823	3,822
	8th July 2021	4,232	4,232
	9th July 2021	4,530	4,580
East Java	6th July 2021	1,808	1,808
	7th July 2021	2,548	2,549
	8th July 2021	2,551	2,551
	9th July 2021	2,530	2,530
Bali Island	6th July 2021	424	424
	7th July 2021	504	504
	8th July 2021	577	580
	9th July 2021	674	720

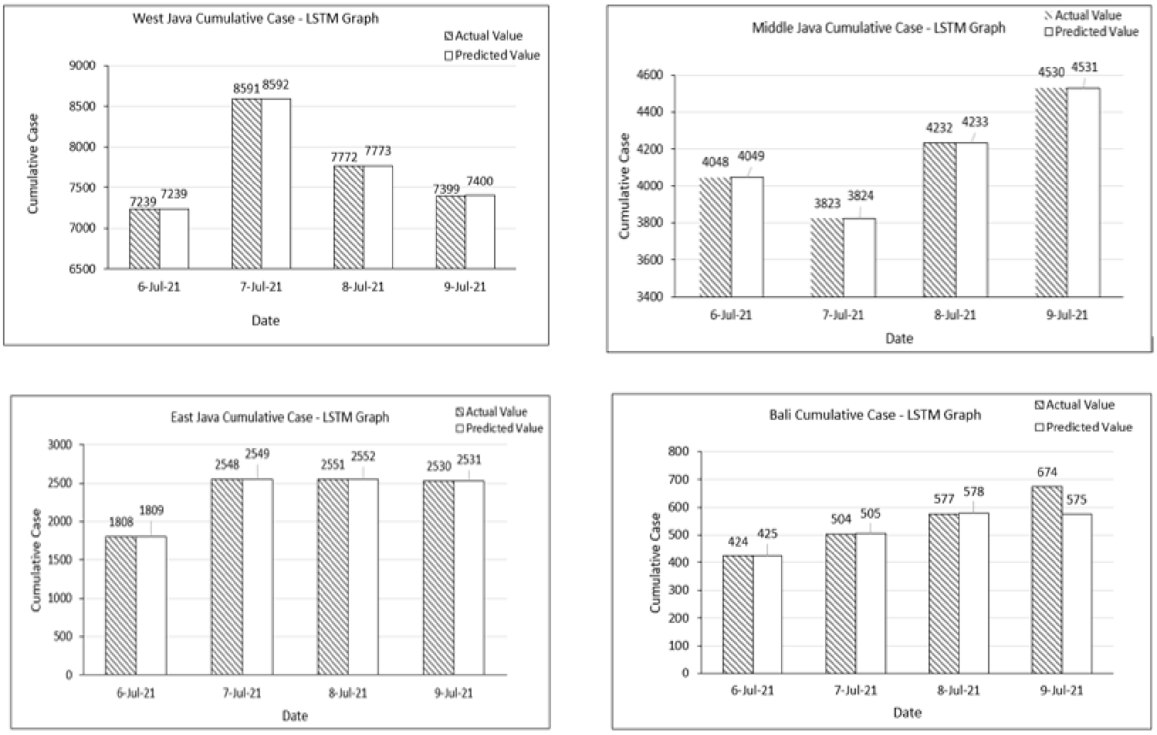

Figure 5. The Java and Bali Island prediction cases using the LSTM model

Figure 5 is prediction results using the LSTM model. We find the trend line in Figure 5 is more significant than the RR and MLR Models. The trend line is increased and closer actual value. The prediction results can be seen clearly on the graphs of East Java and Bali Island, where the prediction results are almost the same as the actual values. This means of the LSTM model is recommended to be used in predicting cumulative cases COVID-19 in Java and Bali Island in Indonesia. The details of the prediction results are shown in Table 3.

Based on Table 3, it is shown that the prediction results using the LSTM model are closer to the actual value. It can be seen in the predicted value on Bali Island on 8th and 9th July 2020. The predicted value of this date is more increased than the result in Table 1 and Table 2. This is clearly seen in the predicted results of Bali Island, which are better than the RR and MLR models. The detail of the error value comparison is shown in Table 4.

The error value in Table 4 is taken from the difference between the predicted value and the actual value. Based on the result in Table 4 can be seen, the LSTM model is recommended to be used to predict the cumulative cases in Java and Bali. The results of the model test are presented in Table 5.

Table 3. The comparison of the prediction result using LSTM model

Region	Date	Actual data	Predicted
West Java	6th July 2021	7,239	7,239
	7th July 2021	8,591	8,592
	8th July 2021	7,772	7,773
	9th July 2021	7,399	7,400
Middle Java	6th July 2021	4,048	4,049
	7th July 2021	3,823	3,824
	8th July 2021	4,232	4,233
	9th July 2021	4,530	4,531
East Java	6th July 2021	1,808	1,809
	7th July 2021	2,548	2,549
	8th July 2021	2,551	2,552
	9th July 2021	2,530	2,531
Bali Island	6th July 2021	424	425
	7th July 2021	504	505
	8th July 2021	577	578
	9th July 2021	674	575

Table 4. Comparison of error values in the RR, MLR, and LSTM models

Region	Date	RR Error %	MLR Error %	LSTM Error %
West Java	6th July 2021	-0.0222	0.0001	0
	7th July 2021	0.1405	0	-0.0001
	8th July 2021	0.0519	-0.0001	-0.0001
	9th July 2021	0.0054	0.0001	-0.0001
Middle Java	6th July 2021	-0.1193	0	-0.0002
	7th July 2021	-0.181	0.0003	-0.0003
	8th July 2021	-0.0633	0	-0.0002
	9th July 2021	0.0088	-0.011	-0.0002
East Java	6th July 2021	0	0	-0.0006
	7th July 2021	0	-0.0004	-0.0004
	8th July 2021	0	0	-0.0004
	9th July 2021	0	0	-0.0004
Bali Island	6th July 2021	-0.5825	0	-0.0024
	7th July 2021	-0.3095	0	-0.002
	8th July 2021	-0.1317	-0.0052	-0.0017
	9th July 2021	0.0519	-0.0682	0.1469

Table 5. The evaluation result of model

Model	Evaluation		
	R2	MSE	RMSE
MLR	0.989	2.89375×10^{2}	17.11×10^{3}
RR	0.988	3.97232×10^{2}	3.972×10^{2}
LSTM	0.998	0.00937×10^{2}	96.82×10^{-2}

The model test values using R^2, MSE, and RMSE in Table 5 showed varying results for the RR, MLR, and LSTM models. But the best test value of the three models is the LSTM model. This is clearly seen in the largest R2 value among the three models. Furthermore, judging from the MSE and RMSE values, the LSTM model also has a better level of accuracy compared to other models. The LSTM model is more recommended in predicting infectious diseases, especially COVID-19 on the islands of Java and Bali.

4. CONCLUSION

Based on the results of experiments conducted to test the prediction model of RR, MLR and LSTM, very significant results were obtained. Several things were produced, among others, the prediction results using LSTM were closer to the actual value compared to other models. This can be seen in the value of the difference between the predicted value and the actual value. Furthermore, the predicted error value, the LSTM model has a lower error rate than other models. Then the results of the model test using R2, MSE and RMSE show that the LSTM model has a better value than other models, so this model is recommended to be used for prediction of infectious diseases, especially COVID-19 in Java and Bali. In future research, the LSTM model will be improved to predict COVID-19 cases globally so that the model can be used in all countries.

ACKNOWLEDGEMENTS

This work was supported by Universitas Amikom Yogyakarta, Indonesia.

REFERENCES

[1] D. Cucinotta and M. Vanelli, "WHO declares COVID-19 a pandemic," *Acta bio-medica : Atenei Parmensis*, vol. 91, no. 1, pp. 157–160, 2020, doi: 10.23750/abm.v91i1.9397.

[2] P. Wang, X. Zheng, J. Li, and B. Zhu, "Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics," *Chaos, Solitons & Fractals*, vol. 139, p. 110058, Oct. 2020, doi: 10.1016/j.chaos.2020.110058.

[3] W. Guo, L. Che, M. Shahidehpour, and X. Wan, "Machine-learning based methods in short-term load forecasting," *The Electricity Journal*, vol. 34, no. 1, p. 106884, Jan. 2021, doi: 10.1016/j.tej.2020.106884.

[4] K. E. ArunKumar, D. V. Kalaga, C. M. S. Kumar, M. Kawaji, and T. M. Brenza, "Comparative analysis of gated recurrent units (GRU), long short-term memory (LSTM) cells, autoregressive integrated moving average (ARIMA), seasonal autoregressive integrated moving average (SARIMA) for forecasting COVID-19 trends," *Alexandria Engineering Journal*, vol. 61, no. 10, pp. 7585–7603, Oct. 2022, doi: 10.1016/j.aej.2022.01.011.

[5] R. Sujath, J. M. Chatterjee, and A. E. Hassanien, "A machine learning forecasting model for COVID-19 pandemic in India," *Stochastic Environmental Research and Risk Assessment*, vol. 34, no. 7, pp. 959–972, Jul. 2020, doi: 10.1007/s00477-020-01827-8.

[6] A. K. Gupta, V. Singh, P. Mathur, and C. M. Travieso-Gonzalez, "Prediction of COVID-19 pandemic measuring criteria using support vector machine, prophet and linear regression models in Indian scenario," *Journal of Interdisciplinary Mathematics*, vol. 24, no. 1, pp. 89–108, Jan. 2021, doi: 10.1080/09720502.2020.1833458.

[7] J. Musulin *et al.*, "Application of artificial intelligence-based regression methods in the problem of COVID-19 spread prediction: a systematic review.," *International journal of environmental research and public health*, vol. 18, no. 8, p. 4287, Apr. 2021, doi: 10.3390/ijerph18084287.

[8] S. Rath, A. Tripathy, and A. R. Tripathy, "Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model.," *Diabetes & metabolic syndrome*, vol. 14, no. 5, pp. 1467–1474, Sep. 2020, doi: 10.1016/j.dsx.2020.07.045.

[9] S. N. Wahyuni, E. Sediyono, and I. Sembiring, "Indonesian Covid-19 future forecasting based on machine learning approach," in *2021 3rd International Conference on Electronics Representation and Algorithm (ICERA)*, 2021, pp. 104–108, doi: 10.1109/ICERA53111.2021.9538672.

[10] G. Ciulla and A. D'Amico, "Building energy performance forecasting: a multiple linear regression approach," *Applied Energy*, vol. 253, p. 113500, Nov. 2019, doi: 10.1016/j.apenergy.2019.113500.

[11] P. Exterkate, P. J. F. Groenen, C. Heij, and D. J. C. van Dijk, "Nonlinear forecasting with many predictors using kernel ridge regression," *SSRN Electronic Journal*, 2011, doi: 10.2139/ssrn.1738192.

[12] E. Bas, E. Egrioglu, U. Yolcu, and C. Grosan, "Type 1 fuzzy function approach based on ridge regression for forecasting," *Granular Computing*, vol. 4, no. 4, pp. 629–637, Oct. 2019, doi: 10.1007/s41066-018-0115-4.

[13] J. Haworth, J. Shawe-Taylor, T. Cheng, and J. Wang, "Local online kernel ridge regression for forecasting of urban travel times," *Transportation Research Part C: Emerging Technologies*, vol. 46, pp. 151–178, Sep. 2014, doi: 10.1016/j.trc.2014.05.015.

[14] Z. Chen, Y. Liu, and S. Liu, "Mechanical state prediction based on LSTM neural netwok," in *2017 36th Chinese Control Conference (CCC)*, 2017, pp. 3876–3881, doi: 10.23919/ChiCC.2017.8027963.

[15] P. Kang-Lin, W. Chib-Hung, and J. G. Yeong-Jia, "The development of a new statistical technique for relating financial information to stock market returns," *International Journal of Management*, vol. 21, no. 4, pp. 492–505, 2004.

[16] M. P. S. Chawla, H. K. Verma, and V. Kumar, "A new statistical PCA-ICA algorithm for location of R-peaks in ECG.," *International journal of cardiology*, vol. 129, no. 1, pp. 146–8, Sep. 2008, doi: 10.1016/j.ijcard.2007.06.036.

[17] K. E. ArunKumar, D. V. Kalaga, C. M. S. Kumar, M. Kawaji, and T. M. Brenza, "Forecasting of COVID-19 using deep layer recurrent neural networks (RNNs) with gated recurrent units (GRUs) and long short-term memory (LSTM) cells.," *Chaos, solitons, and fractals*, vol. 146, p. 110861, 2021.

[18] R. Casado-Vara, A. M. del Rey, D. Pérez-Palau, L. De-La-fuente-valentín, and J. M. Corchado, "Article web traffic time series forecasting using LSTM neural networks with distributed asynchronous training," *Mathematics*, vol. 9, no. 4, pp. 1–22, Feb. 2021, doi: 10.3390/math9040421.

[19] A. Sagheer and M. Kotb, "Time series forecasting of petroleum production using deep LSTM recurrent networks," *Neurocomputing*, vol. 323, pp. 203–213, Jan. 2019, doi: 10.1016/j.neucom.2018.09.082.

[20] K. Pawar, R. S. Jalem, and V. Tiwari, "Stock Market Price Prediction Using LSTM RNN," in *Advances in Intelligent Systems and Computing*, vol. 841, 2019, pp. 493–503.

[21] N. K. Manaswi, *Deep Learning with Applications Using Python*. Berkeley, CA: Apress, 2018.

[22] E. Mussumeci and F. C. Coelho, "Large-scale multivariate forecasting models for dengue - LSTM versus random forest regression.," *Spatial and spatio-temporal epidemiology*, vol. 35, p. 100372, Nov. 2020, doi: 10.1016/j.sste.2020.100372.

[23] J. Xu *et al.*, "Forecast of dengue cases in 20 chinese cities based on the deep learning method.," *International journal of environmental research and public health*, vol. 17, no. 2, p. 453, Jan. 2020, doi: 10.3390/ijerph17020453.

[24] C. Edussuriya, S. Deegalla, and I. Gawarammana, "An accurate mathematical model predicting number of dengue cases in tropics.," *PLoS neglected tropical diseases*, vol. 15, no. 11, p. e0009756, Nov. 2021, doi: 10.1371/journal.pntd.0009756.

[25] J. Gu *et al.*, "A method for hand-foot-mouth disease prediction using geodetector and LSTM model in Guangxi, China.," *Scientific reports*, vol. 9, no. 1, p. 17928, Dec. 2019, doi: 10.1038/s41598-019-54495-2.

[26] K. Lee, M. Lee, and I. Na, "Predicting regional outbreaks of hepatitis a using 3D LSTM and open data in Korea," *Electronics*, vol. 10, no. 21, p. 2668, Oct. 2021, doi: 10.3390/electronics10212668.

[27] Y. Guo, Y. Feng, F. Qu, L. Zhang, B. Yan, and J. Lv, "Prediction of hepatitis E using machine learning models," *PLOS ONE*, vol. 15, no. 9, p. e0237750, Sep. 2020, doi: 10.1371/journal.pone.0237750.

[28] S. Chae, S. Kwon, and D. Lee, "Predicting infectious disease using deep learning and big data.," *International journal of environmental research and public health*, vol. 15, no. 8, p. 1596, Jul. 2018, doi: 10.3390/ijerph15081596.

[29] T. Santosh, D. Ramesh, and D. Reddy, "LSTM based prediction of malaria abundances using big data.," *Computers in biology and medicine*, vol. 124, p. 103859, Sep. 2020, doi: 10.1016/j.compbiomed.2020.103859.

[30] I. A. Clark, L. M. Alleva, A. C. Budd, and W. B. Cowden, "Understanding the role of inflammatory cytokines in malaria and related diseases.," *Travel medicine and infectious disease*, vol. 6, no. 1–2, pp. 67–81, Jan. 2008, doi: 10.1016/j.tmaid.2007.07.002.
[31] A. K. Verma and V. Kuppili, "Data-oriented neural time series with long short-term memories (LSTM) for Malaria incidence prediction in goa, India," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2019, pp. 1–6, doi: 10.1109/ICCCNT45670.2019.8944749.
[32] E. Yang, H. Zhang, X. Guo, Z. Zang, Z. Liu, and Y. Liu, "A multivariate multi-step LSTM forecasting model for tuberculosis incidence with model explanation in liaoning province, China," *BMC Infectious Diseases*, vol. 22, no. 1, p. 490, Dec. 2022, doi: 10.1186/s12879-022-07462-8.
[33] C. A. Indriyani, C. R. Wijaya, and N. N. Qomariyah, "Forecasting COVID-19 total daily cases in Indonesia using LSTM networks," in *2022 5th International Conference on Computing and Informatics (ICCI)*, 2022, pp. 385–391, doi: 10.1109/ICCI54321.2022.9756062.
[34] P. Bedi, S. Dhiman, P. Gole, N. Gupta, and V. Jindal, "Prediction of COVID-19 trend in India and its four worst-affected states using modified SEIRD and LSTM Models," *SN Computer Science*, vol. 2, no. 3, p. 224, May 2021, doi: 10.1007/s42979-021-00598-5.
[35] M. I. Iqbal, M. I. Leon, and S. M. Azim, "Analysing and predicting coronavirus infections and deaths in Bangladesh using machine learning algorithms," *SSRN Electronic Journal*, pp. 1–16, 2020, doi: 10.2139/ssrn.3671978.
[36] H. T. Rauf, J. Gao, A. Almadhor, M. Arif, and M. T. Nafis, "Enhanced bat algorithm for COVID-19 short-term forecasting using optimized LSTM," *Soft Computing*, vol. 25, no. 20, pp. 12989–12999, Oct. 2021, doi: 10.1007/s00500-021-06075-8.
[37] B. Nikparvar, M. M. Rahman, F. Hatami, and J.-C. Thill, "Spatio-temporal prediction of the COVID-19 pandemic in US counties: modeling with a deep LSTM neural network.," *Scientific reports*, vol. 11, no. 1, p. 21715, Dec. 2021, doi: 10.1038/s41598-021-01119-3.
[38] Z. E. Rasjid, R. Setiawan, and A. Effendi, "A comparison: prediction of death and infected COVID-19 cases in Indonesia using time series smoothing and LSTM neural network," *Procedia Computer Science*, vol. 179, pp. 982–988, 2021, doi: 10.1016/j.procs.2021.01.102.
[39] S. Kola, V. Veena, and K. Guntoju, "Forecasting COVID-19 cases in Saudi Arabia using machine learning SEIR and LSTM," *Journal of Green Engineering*, vol. 10, no. 5, pp. 2087–2102, 2020.
[40] R. K. Pathan, M. Biswas, and M. U. Khandaker, "Time series prediction of COVID-19 by mutation rate analysis using recurrent neural network-based LSTM model," *Chaos, Solitons & Fractals*, vol. 138, p. 110018, Sep. 2020, doi: 10.1016/j.chaos.2020.110018.
[41] P. Wang, X. Zheng, G. Ai, D. Liu, and B. Zhu, "Time series prediction for the epidemic trends of COVID-19 using the improved LSTM deep learning method: case studies in Russia, Peru and Iran," *Chaos, Solitons & Fractals*, vol. 140, p. 110214, Nov. 2020, doi: 10.1016/j.chaos.2020.110214.
[42] A. Dairi, F. Harrou, A. Zeroual, M. M. Hittawe, and Y. Sun, "Comparative study of machine learning methods for COVID-19 transmission forecasting.," *Journal of biomedical informatics*, vol. 118, p. 103791, Jun. 2021, doi: 10.1016/j.jbi.2021.103791.
[43] F. Shahid, A. Zameer, and M. Muneeb, "Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM," *Chaos, Solitons & Fractals*, vol. 140, p. 110212, Nov. 2020, doi: 10.1016/j.chaos.2020.110212.
[44] C. Zhu, J. Wang, and J. Tian, "Grain consumption forecasting: one modified MLR model combined with time series forecasting theory," *International Journal of Mathematics and Mathematical Sciences*, vol. 2020, pp. 1–11, May 2020, doi: 10.1155/2020/3956717.
[45] B. Pirouz, H. J. Nejad, G. Violini, and B. Pirouz, "The role of artificial intelligence, MLR and statistical analysis in investigations about the correlation of swab tests and stress on health care systems by COVID-19," *Information*, vol. 11, no. 9, p. 454, Sep. 2020, doi: 10.3390/info11090454.
[46] J. Garcia-Guiliany, E. De-la-hoz-Franco, A.-D. Rodríguez-Toscano, J.-D. De-la-Hoz-Hernández, and H. G. Hernández-Palma, "Multiple linear regression model applied to the projection of electricity demand in Colombia," *International Journal of Energy Economics and Policy*, vol. 10, no. 1, pp. 419–422, Jan. 2020, doi: 10.32479/ijeep.7813.
[47] H. Chung and K. Shin, "Genetic algorithm-optimized long short-term memory network for stock market prediction," *Sustainability*, vol. 10, no. 10, p. 3765, Oct. 2018, doi: 10.3390/su10103765.
[48] B. Yan *et al.*, "An improved method for the fitting and prediction of the number of COVID-19 confirmed cases based on LSTM," *Computers, Materials & Continua*, vol. 64, no. 3, pp. 1473–1490, May 2020, doi: 10.32604/cmc.2020.011317.
[49] V. K. R. Chimmula and L. Zhang, "Time series forecasting of COVID-19 transmission in Canada using LSTM networks.," *Chaos, solitons, and fractals*, vol. 135, p. 109864, Jun. 2020, doi: 10.1016/j.chaos.2020.109864.
[50] A. Tomar and N. Gupta, "Prediction for the spread of COVID-19 in India and effectiveness of preventive measures.," *The Science of the total environment*, vol. 728, p. 138762, Aug. 2020, doi: 10.1016/j.scitotenv.2020.138762.
[51] F. Rustam *et al.*, "COVID-19 future forecasting using supervised machine learning models," *IEEE Access*, vol. 8, pp. 101489–101499, 2020, doi: 10.1109/ACCESS.2020.2997311.

BIOGRAPHIES OF AUTHORS

Sri Ngudi Wahyuni, ST., M.Kom. (Dr. Cand) is senior lecture in Universitas Amikom Yogyakarta Indonesia. Has received her Dr. Candidate degree from the Satya Wacana Christian University of Indonesia in 2020, She received her Bachelor's Degree in Engineering received in 2005 from Ahmad Dahlan University and her Master's Degree in Computer Science received in 2014 from Indonesian Islamic University. Currently working as a senior lecturer in the Informatics Management Study Program, Faculty of Computer Science, Amikom University Yogyakarta Indonesia. Research interest in Data Science, Data mining, and Computer Science. She can be contacted at email: yuni@amikom.ac.id.

Prof. Dr. Ir. Eko Sediono, M.Kom, completed his professorship in 2008. He is a member of IEEE Computer Society #41605422. He completed his undergraduate studies in 1985 at the Bogor Institute of Agriculture with a major in Statistics. Received a master's degree in 1994 from the University of Indonesia and received a Doctorate degree in 2006 from the University of Indonesia with a major in Computer Sciences. Currently working as a lecturer at the Satya Wacana Christian University Salatiga and serving as Manager of the Information Systems and Technology Bureau-Satya Wacana Christian University-Indonesia. Research interest in Data Sciences, Algorithms, and Image Processing. Completed his professorship in 2008. He completed his undergraduate studies in 1985 at the Bogor Institute of Technology. Received a master's degree in 1994 from the University of Indonesia and received a Doctorate degree in 2006 from the University of Indonesia. Currently working as a lecturer at the Satya Wacana Christian University Salatiga and serving as Manager of the Information Systems and Technology Bureau-Satya Wacana Christian University-Indonesia. Research interest in statistics and computer science. He can be contacted at email: sediyono@uksw.edu.

Dr. Irwan Sembiring, ST., M.Kom, CHFI, CIHFI completed his Bachelor's degree in 2001 at UPN "Veteran" Yogyakarta Indonesia. His Master degree is completed in 2004 from Gadjah Mada University Yogyakarta Indonesia, and Doctorate degree completed in 2016 at Gadjah Mada University Indonesia. Research interest in Network Security, Information System and Digital Forensic. Now he is a lecturer at faculty of Information Technology Satya Wacana Christian University, Salatiga, Indonesia. He can be contacted at email: irwan@uksw.edu.

Nazmun Nahar Khanom, M.Sc. Student received her BSc in CSE degree from East West University of Bangladesh in 2020. She is currently pursuing MSc in Information and Communication technology at Bangladesh University of Professionals, Bangladesh. Her research interest includes machine learning, artificial intelligence, data mining, data science and computer science. She can be contacted at email: tithi.khan0@gmail.com.