Prediction Modelling of COVID-19 on Provinces in Indonesia using Long Short-Term Memory Machine Learning

F W Wibowo¹, Wihayati²

¹ Informatics Department, Universitas Amikom Yogyakarta, Jl. Ring Road Utara, Condong Catur, Depok, Sleman, Yogyakarta 55283, Indonesia
² Astrophysics and Mathematics Group, Sekip Utara Bulaksumur, Yogyakarta 55281, Indonesia

Email: ferry.w@amikom.ac.id¹, wihayati@gmail.com²

Abstract. The COVID-19 is a dangerous virus that has been declared by the world health organization (WHO) as a pandemic. Many countries have taken policies to control the virus’s spread and have played an active role in overcoming this global pandemic, including Indonesia. Indonesia consists of many islands, so the level of distribution varies. Although the mortality rate is shallow than the cure rate, this virus’s spread must be controlled. This paper aims to model the prediction of infected cases, cases of recovery from COVID-19, and mortality for each province in Indonesia using the Long Short-Term Memory (LSTM) machine learning method. The results of the model evaluation of this method used the root mean squared error (RMSE) approach.

1. Introduction
Changes in world activities had had a significant impact when the World Health Organization (WHO) declared the pandemic coronavirus 2019 disease (COVID-19) as a global pandemic and caused a global health emergency on March 11, 2020. COVID-19 was caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-COV2), which became one of the spreads of potentially fatal diseases. The transmission is spread by droplets from person to person, so that the government made a policy of maintaining distance, wearing masks, washing hands as part of the health protocol. This virus attaches to the host cell and can cause an inflammatory reaction. The incubation period of up to 14 days can show fever, leucopenia, respiratory syndrome, thrombocytopenia, and multi-organ failure conditions, and it can cause death. An understanding of the characteristics of this virus is needed to control the mortality rate [1]. The high mortality rate is usually influenced by the presence of comorbidities that are owned by patients who have been infected with COVID-19, the vulnerable age factor, and inadequate health facilities [2]. This virus pandemic has drawn attention to global connectivity, which logically brings about changes globally, including Indonesia.

At the time of the spike in cases in Indonesia after the announcement, several regions in Indonesia imposed large-scale social restrictions, and agencies and schools were closed. On the other hand, the air quality in the environment is outstanding and significant because it is slightly contaminated by air pollution [3]. The spike in cases is increasing over time, so that various ways are taken to control this virus’s spread. This paper aims to model the prediction of the virus’s reach, the number of cured, and the number of mortality from Indonesia’s provinces. The modelling used in this paper is a long short-term memory (LSTM) machine learning model. Each case’s model in these provinces will be evaluated using the root mean squared error (RMSE). The model obtained in the
predictions can be used to examine broader implications for a region and trends in the short-term and the long-term [4]. These predictions are used to evaluate the resulting intensity of this pandemic. Public policy and individual awareness in this matter are interrelated to overcome this pandemic. The use of computational intelligence techniques is still believed to be quite good and is widely applied in most application models. It is because the performance can reach an acceptable level of use with measurable evaluation. It makes this technique and the resulting machine learning algorithm is a solution to future problems.

2. Related Works
Cases related to COVID-19 have caught the attention of many countries in the world. Scientists are trying to predict its spread as a means of mitigating this pandemic. It is used as a strategy for the public health system and an effort to minimize people’s death rate with COVID-19. The best model level was sequentially carried out by [5], who got the modeling results of Bi-Long Short-Term Memory (Bi-LSTM), Long Short-term memory (LSTM), Gated Recurrent Unit (GRU), Support Vector Regression (SVR), and Auto-Regressive Integrated Moving Average (ARIMA), respectively. The application of deep learning based on time series techniques has a position in short and medium-time dependence with adaptive learning. Long-term prediction modeling for this case, the researcher [6] implemented multivariate LSTM. The results that have been compared to obtain a reliable model show that the stacked LSTM algorithm has a higher accuracy value. Research [7] modeled this pandemic case using recurrent neural network (RNN) based LSTM variants consisting of bi-directional LSTM, convolutional LSTM, and stacked LSTM to predict points one month ahead. The best prediction model is generated from Convolutional LSTM with high accuracy and fewer errors. From several studies indicating cases of COVID-19 using the LSTM method, the authors observed the LSTM model to model predictions of the spread of COVID-19 infection, patients who recovered from COVID-19, and COVID-19 patients who died. It is due to the high level of COVID-19 disease in Indonesia [8].

3. Research Methodology
This section describes the dataset used in the study. The dataset used is open data, so it's easy to get it from the website. The data used in this paper consists of data plots per region/province marked by the longitude and latitude of each region in Indonesia. Descriptions of the dataset will be searched to determine the amount of data used in the study, the mean, standard deviation (std), and daily maximal case (max) used for each subject. The case data used is the number of everyday issues of cases infected with COVID-19, patients who have recovered from COVID-19 infection, and patients who have died after being infected with COVID-19. This dataset will be used as training data and test data with a ratio of 80:20. Meanwhile, the method employed in this paper is a nested ensemble model or machine learning model using deep learning methods based on the Long Short-Term Memory (LSTM).

3.1. Datasets
The datasets of the COVID-19 for this research were taken from the website of https://data.humdata.org/dataset/indonesia-covid-19-cases-recoveries-and-deaths-per-province [9]. This dataset is taken from 34 provinces in Indonesia. The points of longitude and latitude of each representation of these provinces are shown in Table 1.

| No. | Provinces     | Longitude  | Latitude  |
|-----|---------------|------------|-----------|
| 1.  | Aceh          | 96.90952275| 4.225081892|
| 2.  | Sumatera Utara| 99.05788785| 2.184779215|
| 3.  | Sumatera Barat| 100.466805 | -0.851914348|
| 4.  | Riau          | 101.812892 | 0.510525849|
| 5.  | Jambi         | 103.59932  | -1.61775  |
The positions of each longitude and latitude of these provinces are shown in Figure 1 with plotting spatial data in an Indonesia’s map. Indonesia consists of thousands of islands and is divided into 34 provinces. Some of these islands are separated by a vast ocean.

| Province                      | Longitude  | Latitude   |
|-------------------------------|------------|------------|
| 6. Sumatera Selatan           | 104.1704194| -3.213346822|
| 7. Bengkulu                   | 102.26312  | -3.79116   |
| 8. Kepulauan Bangka Belitung  | 105.9869481| -2.252956411|
| 9. Lampung                    | 105.0183131| -4.918073087|
| 10. Kepulauan Riau            | 108.2022289| 3.915947241|
| 11. DKI Jakarta               | 106.82649  | -6.17148   |
| 12. Jawa Barat                | 107.6028079| -6.919980216|
| 13. Jawa Tengah               | 110.2011125| -7.258970383|
| 14. Daerah Istimewa Yogyakarta| 110.445367 | -7.895013815|
| 15. Jawa Timur                | 112.7336788| -7.720113019|
| 16. Banten                    | 106.1091971| -6.456181599|
| 17. Bali                      | 115.131651 | -8.369743261|
| 18. Nusa Tenggara Barat      | 117.5068625| -8.606652084|
| 19. Nusa Tenggara Timur       | 121.1418292| -8.604843332|
| 20. Kalimantan Barat          | 111.1207624| -0.085901234|
| 21. Kalimantan Tengah         | 113.1810479| -1.634340045|
| 22. Kalimantan Selatan        | 115.3113078| -2.449779137|
| 23. Kalimantan Timur          | 116.4708584| 0.461046227 |
| 24. Kalimantan Utara          | 116.2167183| 2.888820961 |
| 25. Sulawesi Utara            | 124.5353789| 1.286097687 |
| 26. Sulawesi Tengah           | 120.4495925| -1.459214954|
| 27. Sulawesi Selatan          | 120.1611577| -3.747115712|
| 28. Sulawesi Tenggara         | 122.0786781| -4.132803971|
| 29. Gorontalo                 | 123.06184  | 0.53707    |
| 30. Sulawesi Barat            | 119.3434714| -2.461902015|
| 31. Maluku                    | 129.390425 | -3.12619343 |
| 32. Maluku Utara              | 128.0159294| 0.8693517001|
| 33. Papua                     | 140        | -5         |
| 34. Papua Barat               | 132.9683417| -2.041723282|

![Figure 1. Plotting spatial data on 34 provinces in Indonesia.](image-url)

The spread of this virus in big cities is relatively high because human interaction there is quite significant. So that the policies made between a town and another are somewhat different. The distribution of confirmed cases of COVID-19, patients who have recovered from COVID-19, and deaths from COVID-19 patients are shown in Figure 2.
Figure 2. Plotting spatial data on 34 provinces in Indonesia.

The data used in this paper is daily and not the sum per day. The data for COVID-19 infected cases used in this paper were taken from March 15, 2020, to October 29, 2020. The data used for COVID-19 patients who had recovered were taken from March 21, 2020, to October 29, 2020. Meanwhile, the number of patients COVID-19 who died was also taken from March 21, 2020, to October 29, 2020. Blank data in the dataset, which means there are no case reports, will be replaced with zero. This research is only based on the dataset obtained, while unreported case data is not included in this paper. So it could be that the number of people infected with this virus is enormous, but because they do not report themselves to the authorized officers, their data is not recorded [9]. This paper seeks to discuss predictions obtained from the LSTM model only so that the proposed prediction model can be used to anticipate future pandemics. Each province data used in this paper is written in table 2.

Table 2. Dataset descriptions of infected, recovered, and death cases of COVID-19 in each province of Indonesia.

| No. | Provinces | Dataset Description | Infected | Recovered | Death |
|-----|-----------|---------------------|----------|-----------|-------|
| 1.  | Aceh      | Count               | 229      | 223       | 223   |
|     |           | Mean                | 32.197   | 22.587    | 1.193 |
|     |           | Standard Deviation  | 49.571   | 61.394    | 2.189 |
|     |           | Max                 | 212      | 399       | 12    |
| 2.  | Sumatera Utara | Count               | 229      | 223       | 223   |
|     |           | Mean                | 56.432   | 47.287    | 2.368 |
|     |           | Standard Deviation  | 53.409   | 55.753    | 2.290 |
|     |           | Max                 | 249      | 266       | 10    |
| 3.  | Sumatera Barat | Count               | 229      | 223       | 223   |
|     |           | Mean                | 61.066   | 39.610    | 1.126 |
|     |           | Standard Deviation  | 98.157   | 78.260    | 2.030 |
|     |           | Max                 | 484      | 550       | 13    |
| 4.  | Riau      | Count               | 229      | 223       | 223   |
|     |           | Mean                | 17.031   | 11.565    | 1.457 |
|     |           | Standard Deviation  | 30.384   | 28.787    | 2.592 |
|   |   |   |
|---|---|---|
|   | Max | Median | Standard Deviation |
| 5. Jambi | 163 | 229 | 9.253 |
|   | 286 | 223 | 6.586 |
|   | 12 | 0.103 | 0.406 |
| 6. Sumatera Selatan | 5.323 | 33.830 | 26.361 |
|   | 2.659 | 28.031 | 30.234 |
|   | 0.103 | 1.830 | 2.094 |
| 7. Bengkulu | 46 | 119 | 11 |
|   | 54 | 181 | 11 |
|   | 3 | 0.215 | 0.510 |
| 8. Kepulauan Bangka Belitung | 229 | 229 | 2.524 |
|   | 223 | 2.296 | 4.046 |
|   | 223 | 0.036 | 0.186 |
| 9. Lampung | 229 | 7.677 | 12.382 |
|   | 223 | 4.605 | 9.035 |
|   | 223 | 0.332 | 0.793 |
| 10. Kepulauan Riau | 62 | 62 | 63.402 |
|   | 63 | 11.565 | 28.788 |
|   | 5 | 0.417 | 1.131 |
| 11. DKI Jakarta | 229 | 229 | 455.873 |
|   | 223 | 409.193 | 446.625 |
|   | 223 | 9.852 | 8.178 |
| 12. Jawa Barat | 1398 | 1398 | 154.209 |
|   | 40.9193 | 1896 | 1446.625 |
|   | 37 | 1896 | 8.178 |
| 13. Jawa Tengah | 229 | 154.209 | 193.116 |
|   | 223 | 111.547 | 174.269 |
|   | 223 | 3.103 | 4.833 |
| 14. Daerah Istimewa Yogyakarta | 571 | 136.192 | 136.192 |
|   | 2223 | 198.967 | 198.967 |
|   | 70 | 10.219 | 10.219 |
| 15. Jawa Timur | 229 | 144.524 | 144.524 |
|   | 223 | 123.354 | 123.354 |
|   | 223 | 7.785 | 7.785 |
| 16. Banten | 18.981 | 13.740 | 16.375 |
|   | 17.426 | 84 | 84 |
|   | 0.766 | 0.422 | 0.422 |
| 17. Bali | 82 | 18.981 | 148.491 |
|   | 84 | 172.890 | 172.890 |
|   | 4 | 10.637 | 10.637 |
|   | 16 | 148.491 | 148.491 |
|   | 223 | 205.996 | 205.996 |
|   | 16.812 | 123.354 | 123.354 |
|   | 10.637 | 52 | 52 |
|   | 16 | 123.354 | 123.354 |
|   | 2.213 | 49.279 | 49.279 |
|   | 1.717 | 6.907 | 6.907 |
|   | 2.773 | 44.192 | 44.192 |
|   | 14 | 44.192 | 44.192 |
|     | Nusa Tenggara Barat | Count | 229 | 223 | 223 |
|-----|---------------------|-------|-----|-----|-----|
|     | Mean                | 17.249| 14.565| 0.987 |
|     | Standard Deviation  | 12.069| 15.198| 1.320 |
|     | Max                 | 54    | 110 | 6 |
| 19  | Nusa Tenggara Timur| Count | 229 | 223 | 223 |
|     | Mean                | 2.947 | 2.076 | 0.031 |
|     | Standard Deviation  | 6.173 | 5.708 | 0.175 |
|     | Max                 | 47    | 51  | 1 |
| 20  | Kalimantan Barat    | Count | 229 | 223 | 223 |
|     | Mean                | 7.131 | 5.955 | 0.081 |
|     | Standard Deviation  | 11.584| 10.680| 0.383 |
|     | Max                 | 81    | 66  | 4 |
| 21  | Kalimantan Tengah   | Count | 229 | 223 | 223 |
|     | Mean                | 18.899| 17.143| 0.673 |
|     | Standard Deviation  | 19.150| 23.888| 1.250 |
|     | Max                 | 92    | 120 | 9 |
| 22  | Kalimantan Selatan  | Count | 229 | 223 | 223 |
|     | Mean                | 51.367| 47.695| 2.161 |
|     | Standard Deviation  | 40.613| 54.865| 2.650 |
|     | Max                 | 181   | 277 | 27 |
| 23  | Kalimantan Timur    | Count | 229 | 223 | 223 |
|     | Mean                | 60.135| 47.260| 2.108 |
|     | Standard Deviation  | 79.031| 67.194| 3.819 |
|     | Max                 | 392   | 275 | 27 |
| 24  | Kalimantan Utara    | Count | 229 | 223 | 223 |
|     | Mean                | 3.546 | 3.063  | 0.031 |
|     | Standard Deviation  | 5.797 | 5.799  | 0.175 |
|     | Max                 | 48    | 50   | 1 |
| 25  | Sulawesi Utara      | Count | 229 | 223 | 223 |
|     | Mean                | 23.668| 20.094| 0.901 |
|     | Standard Deviation  | 24.336| 23.925| 1.255 |
|     | Max                 | 134   | 134  | 7 |
| 26  | Sulawesi Tengah     | Count | 229 | 223 | 223 |
|     | Mean                | 3.738 | 2.798  | 0.152 |
|     | Standard Deviation  | 7.047 | 5.534  | 0.439 |
|     | Max                 | 51    | 39   | 3 |
| 27  | Sulawesi Selatan    | Count | 229 | 223 | 223 |
|     | Mean                | 79.926| 73.099| 2.076 |
|     | Standard Deviation  | 60.605| 83.902| 3.172 |
|     | Max                 | 218   | 519  | 24 |
| 28  | Sulawesi Tenggara   | Count | 229 | 223 | 223 |
|     | Mean                | 21.415| 16.475| 0.363 |
|     | Standard Deviation  | 32.239| 29.540| 0.663 |
|     | Max                 | 226   | 238  | 3 |
| 29  | Gorontalo           | Count | 229 | 223 | 223 |
|     | Mean                | 13.135| 12.709| 0.381 |
|     | Standard Deviation  | 24.311| 21.699| 0.779 |
|     | Max                 | 142   | 152  | 5 |
| 30  | Sulawesi Barat      | Count | 229 | 223 | 223 |
|     | Mean                | 4.318 | 3.664  | 0.063 |
|     | Standard Deviation  | 11.740| 10.809| 0.293 |
|     | Max                 | 156   | 148  | 2 |
| 31  | Maluku              | Count | 229 | 223 | 223 |
Table 2 above shows that the most significant mean of COVID-infected cases was DKI Jakarta, amounting to 455.873, but the mean number of recovery cases was also the largest. Meanwhile, the largest mean for patients who died was in Jawa Timur province, amounting to 16.812. The smallest mean number of infected cases was in the Kepulauan Bangka Belitung of 2.524, while the minor mean of points of patients who had recovered was in Nusa Tenggara Timur at 2.076. Still, Nusa Tenggara Timur had a mean of cases of death of COVID-19 sufferers of 0.031. The table also shows that the most significant standard deviation of infected data was DKI Jakarta, amounting to 424.471. However, DKI Jakarta had the most significant standard deviation of recovered patients, namely 446.625. Meanwhile, the highest standard deviation for death cases from COVID-19 sufferers was in Jawa Timur province, amounting to 10.637.

3.2. Long Short-Term Memory
Long Short-Term Memory (LSTM) is another type of processing module for Recurrent Neural Network (RNN). This method consists of modules with repeated processing. The LSTM key is the path that connects the previous cell $\xi_{t-1}$ to the new cell $\xi_t$ at the top of the LSTM module, as shown in Figure 3.

![Figure 3. LSTM key factor.](image)

The working principle of the LSTM is to decide which input will be removed from the candidate values $\xi_{t-1}$ of the cell state using a sigmoid gate called the forget gate $\phi$. This gate reads the values of $\xi_{t-1}$ and $x_t$, which will produce values ‘0’ to stop the input element and ‘1’ to forward the input element for each component in $\xi_{t-1}$, as depicted in figure 4.
Figure 4. Information selection in LSTM.

The equation for Figure 4 is written as equation 1.

\[ y_t = \sigma(\omega_{y'} \cdot [\xi_{t-1}, x_t] + \beta_{y'}) \]  \hspace{1cm} (1)

The symbol \([\xi_{t-1}, x_t]\) is a concatenation operation, \(\omega\) is a weight of input, and \(\beta\) is a bias of the gate input. The next step is to decide which new information to use in \(\xi_t\). In this process, there are two parts: the sigmoid gate, known as the input gate \(\omega_i\), which determines the information to be updated, and a gate without generating a new vector candidate \(\omega_c\). The two processes combine to make up-to-date information, as shown in Figure 5.

Figure 5. New information selection in LSTM.

Figure 5 contains two functions that could be written in the equations, as shown in Equations 2 and 3, respectively.

\[ \alpha_t = \sigma(\omega_i \cdot [\xi_{t-1}, x_t] + \beta_i) \]  \hspace{1cm} (2)

\[ \xi_t = \tanh(\omega_c \cdot [\xi_{t-1}, x_t] + \beta_c) \]  \hspace{1cm} (3)

This process adds new information to the cell to replace long-forgotten elements. Updating the old cell \(\xi_{t-1}\) to the new cell \(\xi_t\) will need to multiply that old cell by \(y_t\) to forget things it decided to ignore. This process also multiplies the new candidate cell by \(\xi\) with \(\alpha_t\) to determine how many new candidate cells to include, then adding both of them, as shown in Figure 6.
The equation formed from Figure 6 is written in equation 4.

\[ \xi_t = \psi_t \cdot \xi_{t-1} + \alpha_t \cdot \zeta_t \]  

(4)

In this case, dumping old information and adding new information is as decided in the previous step. This output will be based on the value in the cell passed to a filter. The process is by running a sigmoid gate called the output gate \( \psi_t \) to decide which pieces of information will be generated. Then give this information through \( \tanh \) to make the value between \(-1\) and \(1\), multiply that by the output of the sigmoid gate so that only the part is decided upon. It can be depicted as shown in Figure 7.

Equations 5 and 6 show the illustration of Figure 7.

\[ \phi_t = \sigma(\omega_{\phi} [\xi_{t-1}, x_t] + \beta_{\phi}) \]  

(5)

\[ \xi_t = \phi_t \tanh \xi_t \]  

(6)

The LSTM modeling applied in this paper uses five hidden layers, 1000 epochs, and a learning rate of 0.001. The optimizer used in this paper has employed an Adam optimizer while to evaluate the prediction model using the root mean square error (RMSE).

4. Result and Discussion

This section discusses the results obtained from the predictive model of LSTM. The model evaluation results were obtained using the RMSE for each infected case, patients who recovered from COVID-19, and patients who died for each province in Indonesia, as shown in Table 3.
Table 3. LSTM RMSE of each province in Indonesia has five hidden neurons, 1000 epochs, and a learning rate of 0.001.

| No. | Provinces               | RMSE  | Cases   | Recovered | Death |
|-----|-------------------------|-------|---------|-----------|-------|
| 1.  | Aceh                    | 29.795| 54.830  | 1.672     |       |
| 2.  | Sumatera Utara         | 35.585| 27.650  | 1.941     |       |
| 3.  | Sumatera Barat         | 50.575| 43.512  | 1.747     |       |
| 4.  | Riau                    | 21.625| 27.765  | 1.649     |       |
| 5.  | Jambi                   | 8.394 | 5.973   | 0.414     |       |
| 6.  | Sumatera Selatan       | 19.843| 22.460  | 1.930     |       |
| 7.  | Bengkulu                | 7.162 | 6.903   | 0.512     |       |
| 8.  | Kepulauan Bangka Belitung | 3.426 | 4.096   | 0.189     |       |
| 9.  | Lampung                 | 8.744 | 8.349   | 0.740     |       |
| 10. | Kepulauan Riau          | 30.747| 27.508  | 1.134     |       |
| 11. | DKI Jakarta             | 90.044| 172.095 | 6.891     |       |
| 12. | Jawa Barat              | 129.014| 125.691| 4.735     |       |
| 13. | Jawa Tengah             | 61.740| 157.893 | 7.658     |       |
| 14. | Daerah Istimewa Yogyakarta | 10.945 | 11.430 | 0.741     |       |
| 15. | Jawa Timur              | 76.360| 62.233  | 6.160     |       |
| 16. | Banten                  | 30.936| 45.390  | 2.373     |       |
| 17. | Bali                    | 18.141| 19.550  | 1.475     |       |
| 18. | Nusa Tenggara Barat    | 9.709 | 13.588  | 1.246     |       |
| 19. | Nusa Tenggara Timur    | 5.742 | 5.893   | 0.176     |       |
| 20. | Kalimantan Barat       | 10.346| 9.920   | 0.385     |       |
| 21. | Kalimantan Tengah      | 17.068| 21.022  | 1.252     |       |
| 22. | Kalimantan Selatan     | 27.988| 38.038  | 2.544     |       |
| 23. | Kalimantan Timur       | 44.082| 37.032  | 3.294     |       |
| 24. | Kalimantan Utara       | 5.696 | 5.829   | 0.177     |       |
| 25. | Sulawesi Utara         | 19.207| 15.925  | 1.186     |       |
| 26. | Sulawesi Tengah        | 7.049 | 5.133   | 0.449     |       |
| 27. | Sulawesi Selatan       | 19.843| 57.544  | 3.084     |       |
| 28. | Sulawesi Tenggara      | 25.562| 25.127  | 0.623     |       |
| 29. | Gorontalo               | 20.989| 17.100  | 0.778     |       |
| 30. | Sulawesi Barat         | 11.665| 10.839  | 0.290     |       |
| 31. | Maluku                  | 19.738| 23.327  | 0.522     |       |
| 32. | Maluku Utara           | 14.978| 38.196  | 0.757     |       |
| 33. | Papua                   | 43.662| 45.890  | 1.308     |       |
| 34. | Papua Barat            | 28.199| 58.576  | 0.743     |       |

Table 3 shows that the smallest RMSE obtained in the LSTM model for cases of COVID-19 infection in the province of Kepulauan Bangka Belitung is 3.426, as well as the smallest RMSE for the LSTM model in cases of patients recovering from COVID-19 infection in this province of 4.096. However, the smallest RMSE of cases of dying patients who have been infected with COVID-19 is in the province of Nusa Tenggara Timur is 0.176. The RMSE of the LSTM model for the most prominent infected cases is in the modelling for Jawa Barat province of 129.014. The LSTM prediction model RMSE for patients who have recovered from COVID-19 is in the prediction model for DKI Jakarta province of 172.095. Meanwhile, the largest RMSE for the prediction model for COVID-19 patients who died was in Jawa Tengah province with a value of 7.658.

5. Conclusion
This paper has modeled predictions from cases of COVID-19 infection, patients who have recovered from COVID-19, and cases of COVID-19 patients who have died for each province in Indonesia. The
The smallest RMSE value of the LSTM model that uses five hidden neurons, 1000 epochs, a learning rate of 0.001, and Adam's optimizer for cases of infection, recovery, and death from COVID-19 patients are found in the provinces of Jawa Barat, DKI Jakarta, and Jawa Tengah, respectively. There is a tendency for a relationship between the mean and standard deviation of the dataset and the RMSE results obtained from LSTM modeling. However, this needs to be tested further so that to get a rough idea of the results from the RMSE model of this LSTM, it can be seen from the mean, the standard deviation of the entered data set, and RMSE of the LSTM modelling for infected, recovered, and died cases, as shown in Figure 8, 9 and 10, respectively.

**Figure 8.** Data mean, data standard deviation, and RMSE from LSTM modeling for cases of patients infected with COVID-19.

**Figure 9.** Data mean, data standard deviation, and RMSE from LSTM modeling for cases of patients recovering from COVID-19.
Figure 10. Data mean, data standard deviation, and RMSE from LSTM modeling for cases of COVID-19 patients who have died.

6. References

[1] Sukmana M and Yuniarti F A 2020 The pathogenesis characteristics and symptom of Covid-19 in the context of establishing a nursing diagnosis Jurnal Kesehatan Pasak Bumi Kalimantan vol 3 no 1

[2] Clerkin K J, Fried J A, Raikhelkar J, Sayer G, Griffin J M, Masoumi A, Jain S S, Burkhoff D, Kumariah D, Rabbani L, Schwartz A, and Uriel N 2020 COVID-19 and cardiovascular disease Circulation 141 p 1648-1655 doi: 10.1161/CIRCULATIONAHA.120.046941

[3] Wihayati and Wibowo F W 2021 Prediction of air quality in Jakarta during the COVID-19 outbreak using long short-term memory machine learning IOP Conf. Ser. Earth Environ. Sci.

[4] Liu F, Wang J, Liu J, Li Y, Liu D, Tong J, Li Z, Yu D, Fan Y, Bi X, Zhang X, and Mo S 2020 Predicting and analyzing the COVID-19 epidemic in China: Based on SEIRD, LSTM and GWR models PLoS ONE 15(8): e0238280 doi: 10.1371/journal.pone.0238280

[5] Shahid F, Zameer A, and Muneeb M 2020 Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM Chaos, Solitons & Fractals vol 140 110212 doi: 10.1016/j.chaos.2020.110212

[6] Devaraj J, Elavarasan R M, Pugazhendhi R, Shaﬁullah G M, Ganesan S, Jeysree A K, Khan I A, and Hossain E 2021 Forecasting of COVID-19 cases using deep learning models: Is it reliable and practically signiﬁcant? Results in Physics vol 21 103817 doi: 10.1016/j.rinp.2021.103817

[7] Shastri S, Singh K, Kumar S, Kour P, and Mansotra V 2020 Time Series Forecasting of Covid-19 using Deep Learning Models: India-USA Comparative Case Study Chaos, Solitons & Fractals 110227 doi:10.1016/j.chaos.2020.110227

[8] Wibowo F W and Wihayati 2021 Prediction modelling of COVID-19 outbreak in Indonesia using a logistic regression model J. Phys.: Conf. Ser. 1803 012015

[9] OCHA services 2020 Indonesia: coronavirus (COVID-19) subnational cases https://data.humdata.org/dataset/indonesia-covid-19-cases-recoveries-and-deaths-per-province (accessed on 1 November 2020)

Acknowledgments
The authors would like to thank Hemispheres for providing the opportunity to contribute to this research paper’s publication.