Characteristic Parameters of Epoch Deep Learning to Predict Covid-19 Data in Indonesia

Widi Hastomo¹, Adhitio Satyo Bayangkari Karno², Nawang Kalbuana³, Andri Meiriki⁴, Sutarno⁵

¹ Information Systems Department, Institut Teknologi dan Bisnis Ahmad Dahlan, Indonesia
² Information Systems Department, Gunadarma University, Indonesia
³ Politeknik Penerbangan Indonesia Curug, Indonesia
⁴ Managemen Department, Institut Teknologi dan Bisnis Ahmad Dahlan, Indonesia
⁵ Information Systems Department, STMIK Jakarta STI&K, Indonesia

*Widie.has@gmail.com

Abstract. This study aims to predict Covid-19 data in Indonesia using LSTM machines learning and GRU using python. As a comparison, two datasets from other countries which have strong correlation were used. The dataset is of the ourworldindata.org page. The results of the LSTM model with epoch 15, RMSE 68,417 require rapid processing time and better accuracy than GRU with epoch 400, RMSE 90,173. The results from Covid-19 data processing in Indonesia have a robust correlation with Covid-19 data in Azerbaijan, Bangladesh, Bhutan, Cape Verde, Curacao, Slovenia, South Africa, and Thailand. The epoch characteristics of LSTM and GRU are a challenge since the amount of Covid-19 data is relatively minor.

1. Introduction

Since March 2020, the Indonesian government first announced that residents were infected with Covid-19, this has had a profound impact not only in the health sector, but also in the economy [1]–[5], education, social. People's purchasing power has decreased by 0.57% in the first quarter of this year, the cessation of the business world has made the investment climate experience uncertainty. Efforts that have been made by the government are making social distancing policies, large-scale social restrictions to reduce the spread of Covid-19.

The characteristics of the data on the spread of Covid-19 are time series data, this is possible to predict using machine learning [6],[7] The results of the predictions are expected to be a 2nd opinion in making policies, so that investors can invest safely and be able to manage risks that are will arise at a later date. Prediction modeling using machine learning about Covid-19 has been done quite a lot by researchers [8]–[10] [5] but for the publication of predictions in Indonesia it is still limited.

In this decade, the development of science has developed quite rapidly, this is evidenced by the use of technology based on artificial intelligence (AI) which has been used in various aspects of human life [7], [11]–[14]. Based on the data provided, this technology makes machines capable of learning on their own just like humans [15]–[17]. With the combination of layers and the addition of nodes in machine learning (machine learning), it forms a separate classification called Deep Learning (DL). Research related to DL is still being developed to get the best results. Two learning machines are classified as deep learning that will be used in this study, namely GRU (Gated Recurrent Unit) and LSTM (Long Short Term Memory) [18]–[20].
Python programming language is used because it has many libraries that make it easy to create programs that involve a lot of vector and matrix manipulation (hard libraries) tensorflow [21]–[24],[25] numpy, pandas, as well as visual displays of various graphics that are attractive and easy to read library scikit learn, matplotlib, as well as heatmaps (library seaborn) to show correlation in the form of color and numerical maps.

In this study, we will see how the effect of changes in the increase in the number of epochs on the RMSE (Root Mean Square Error) value of the two LSTM and GRU learning machine models and uses a relatively minimal amount of data, namely the Covid-19 time series data from all countries. To strengthen the results of this study from Covid-19 data in the world, 2 datasets will be selected that have a strong relationship with Covid-19 data in Indonesia. So data processing will use 3 Covid-19 datasets. The process of determining the strong relationship between all countries and the Covid-19 data in Indonesia, used the correlation method with results that are easily visible in the form of visual levels of color (heat map matrix) and also numerically.

Visualization of the test results will show the high similarity of the prediction graph to the real data graph, quantitatively the level of similarity in the graph can be shown by the smaller RMSE value. The number of epochs used to obtain the minimum RMSE value will certainly affect the running time of the program. A larger number of epochs will take more time, and a smaller number of epochs will speed up the running time of the program. Of course, what is expected is how to obtain a small number of epochs with a minimal RMSE value.

In the process of predicting data using machine learning, the challenge is to determine the parameter of the number of epochs that will be used. The right number of epochs will produce a high accuracy value (small RMSE) and determine the slow speed of the program running. This is a challenge in itself in using LSTM and GRU machine learning

2. Methodology

This study uses an optimized LSTM and GRU deep learning algorithm approach. The process started with the determination of the RMSE epoch, followed by training, testing optimization of intermediate predictions.

![Figure 1. Research Scenarios](image)

The process begins with the source of the Covid-19 dataset for all countries, taken from “ourworldindata.org” starting from 31 December 2019 to 29 May 2020. This dataset consists of 5 column data features, namely location, new_cases, new_deaths, total_cases, total_deaths, totaling 20,235 lines. Meanwhile, Indonesia's Covid-19 data contains 144 rows of data. From the five feature data columns, one feature is selected which will be used as further data to be predicted, namely the new_cases feature.

![Figure 2. New_cases Covid-19 world](image)
Dataset selection, so that research results can be compared based on different but similar data (strong correlation), two other countries will be selected with data that have a strong correlation value with Covid-19 data in Indonesia. The results of the correlation process are shown in the heat map matrix.

The heatmap matrix, it can be seen that the correlation value of each of the 2 countries in the world, the strong correlation \( r=1 \) is shown in bright red. Countries that have a strong correlation value with Covid-19 data in Indonesia are 8 countries, Azerbaijan, Bangladesh, Bhutan, Cape Verde, Curacao, Slovenia, South Africa, Thailand. Of the 8 countries, 2 countries (Bangladesh and South Africa) were chosen to carry out the prediction process to compare the final results.

The new_cases data for the three countries (Indonesia, Bangladesh and South Africa) starts from March 2020. Then the data that will be used for training starts from March 1, 2020 to May 15, 2020 (69 data), and for testing from May 16, 2020 to May 29 2020 (14 data)
The total data that will be used to predict LSTM and GRU in 3 countries are 83 data respectively. The LSTM architecture used consists of 1 input layer, 3 hidden layers, each layer consists of 50 nodes and 1 dense layer with 1 node. Dropout value = 0.2. The GRU architecture consists of 4 layers (50 nodes per layer) and 1 dense layer with 1 node. Dropout value = 0.2.

Epoch is the number of repetition of the learning process carried out by machine learning. The greater the amount of epoch, the more time it will take in processing. This does not mean that a larger number of epochs will result in a higher accuracy value. Of course what is expected is good results with a small Epoch and accuracy (small RMSE value).

Therefore it is very difficult to determine the right number of Epochs to be used in machine learning to get good accuracy results. By creating a programming iteration with the output in the form of a graphic the RMSE value for each Epoch increment from the data of the three countries. From the graph, it can be seen that LSTM machine learning to achieve high accuracy requires a smaller Epoch (faster amount of processing time) compared to GRU.

3. Result and Discussion

Epoch versus RMSE charts for 3 countries, the epoch values to be used for LSTM and GRU for Indonesia are 15 and 400, Bangladesh is 20 and 300, and South Africa is 20 and 340. The output graph with epoch values from 3 countries can be seen in table below.

| Table 1. Three country output chart |
|-----------------------------------|
| **Epoch** | **LSTM** | **GRU** |
| 15 | ![](LSTM_15.png) | ![](GRU_15.png) |
| Indonesia | 400 | ![](LSTM_400.png) | ![](GRU_400.png) |
| 20 | ![](LSTM_20.png) | ![](GRU_20.png) |
| Bangladesh | 300 | ![](LSTM_300.png) | ![](GRU_300.png) |
| 20 | ![](LSTM_20.png) | ![](GRU_20.png) |
| South Africa | 340 | ![](LSTM_340.png) | ![](GRU_340.png) |
The results of the study using Covid-19 data, obtained that the epoch size with high accuracy in Indonesia is 15 for LSTM and 400 for GRU. Bangladesh is 20 for LSTM and 300 for GRU. The South African countries are 20 for LSTM and 340 for GRU.

4. Conclusion

Based on the results of the discussion of the epoch values of the three countries of Indonesia, Bangladesh and South Africa, it can be concluded as follows;

a) Time series data prediction uses a minimal amount of data, it turns out that LSTM machine learning will require faster processing time compared to GRU.

b) The results of this study are contrary to Hung, Junyoung et al, 2014, which states that GRU is able to show better capabilities, especially for small datasets.

c) It is necessary to conduct research with a more varied amount of data or change the parameters of machine learning architecture (layers and nodes) to test the capabilities of LSTM and GRU, especially in terms of processing time series data.

d) By knowing the characteristics of the epoch changes to this accuracy, it will be easier to estimate the number of epochs needed for certain types of data, so that the expected results can be better.

e) Covid-19 data Indonesia has strong data correlation \( r = 1 \) with Azerbaijan, Bangladesh, Bhutan, Cape Verde, Curacao, Slovenia, South Africa, and Thailand.

f) The correlation between Indonesia's Covid-19 data and other countries can be seen from many factors and can be used as material for extensive discussion by looking at various points of view (geographic, social, economic, political and so on).

References

[1] R. Luwihono, A., Suherman, B., Sembiring, D., Rasyid, S., Kalbuana, N., Saputro, R., Prasetyo, B., Taryana, T., Suprihartini, Y., Asih, P., Mahfud, Z & Rusdiyanto, “Macroeconomic effect on stock price: Evidence from Indonesia,” *Accounting*, vol. 7, no. 5, pp. 1189–1202, 2021.

[2] Z. Susanto, H., Prasetyo, I., Indrawati, T., Aliyyah, N., Rusdiyanto, R., Tjaraka, H., Kalbuana, N., Rochman, A., Gazali, G & Zainurrafiqi, “The impacts of earnings volatility, net income and comprehensive income on share price: evidence from Indonesia stock exchange,” *Accounting*, vol. 7, no. 5, pp. 1009–1016, 2021.

[3] M. Jannah et al., “Effect of ISO 9001, ISO 45001 and ISO 14000 toward financial performance of Indonesian manufacturing.” *Syst. Rev. Pharm.*, vol. 11, no. 10, 2020.

[4] A. Kunaifi, “Quantitative Easing In Overcoming the Crisis: Criticism of QE in the Monetary System Of Capitalism,” *Turkish J. Comput. Math. ...,* vol. 12, no. 03, pp. 3703–3714, 2021.

[5] I. Prasetyo, N. Aliyyah, Rusdiyanto, H. Tjaraka, N. Kalbuana, and A. S. Rochman, “Vocational training has an influence on employee career development: a case study indonesia,” *Acad. Strateg. Manag. J.*, vol. 20, no. 2, pp. 1–14, 2021.

[6] W. Hastomo and A. Satyo, “Long Short Term Memory Machine Learning Untuk Memprediksi Akurasi Nilai Tukar IDR Terhadap USD,” vol. 3, 2019.

[7] Adhito Satyo Bayangkari Karno, “Prediksi Data Time Series Saham Bank BRI Dengan Mesin Belajar LSTM ( Long ShortTerm Memory ),” vol. 1, no. 1, pp. 1–8, 2020.

[8] F. Rustam et al., “COVID-19 Future Forecasting Using Supervised Machine Learning Models,” *IEEE Access*, vol. 8, pp. 101489–101499, 2020.

[9] N. Zheng et al., “Predicting COVID-19 in China Using Hybrid AI Model,” *EEE Trans. Cybern.*, vol. 50, no. 7, pp. 2891–2904, 2020.

[10] H. Panwar and R. Morales-menendez, “Application of Deep Learning for Fast Detection of COVID-19 in X-Rays using nCOVnet,” *Chaos, Solitons Fractals Interdiscip. J. Nonlinear Sci. Nonequilibrium Complex Phenom.,* p. 109944, 2020.

[11] W. Hastomo, R. Dalam, K. Baru, T. Selatan, and U. Gunadarma, “KEMAMPUAN LONG SHORT TERM MEMORY MACHINE,” vol. 4, no. September, pp. 229–236, 2020.

[12] W. Hastomo, “Gesture Recognition For Pencak Silat Tapak Suci Real-Time Animation,” vol. 2, pp. 91–102, 2020.

[13] W. Karno, A. S. B., & Hastomo, “Optimalisasi Data Terbatas Prediksi Jangka Panjang Covid-
19 Dengan Kombinasi Lstm Dan GRU,” Pros. SeNTIK, vol. 4, no. September, pp. 181–191, 2020.

[14] A. S. B. Karno, W. Hastomo, and I. S. K. Wardhana, “Prediksi Jangka Panjang Covid-19 Indonesia Menggunakan Deep Learning Long-Term,” Semin. Nas. Teknol. Inf. dan Komun., pp. 483–490, 2020.

[15] V. Ayumi, L. M. R. Rere, M. I. Fanany, and A. M. Arymurthy, “Optimization of convolutional neural network using microcanonical annealing algorithm,” 2016 Int. Conf. Adv. Comput. Sci. Inf. Syst. ICACSiS 2016, pp. 506–511, 2017.

[16] L. M. Rasdi Rere, M. I. Fanany, and A. M. Arymurthy, “Metaheuristic Algorithms for Convolutional Neural Network,” Comput. Intell. Neurosci., vol. 2016, 2016.

[17] L. M. R. Rere, M. I. Fanany, and A. M. Arymurthy, “Simulated Annealing Algorithm for Deep Learning,” Procedia Comput. Sci., vol. 72, pp. 137–144, 2015.

[18] W. Li, H. Wu, N. Zhu, Y. Jiang, J. Tan, and Y. Guo, “Prediction of dissolved oxygen in a fishery pond based on gated recurrent unit (GRU),” Inf. Process. Agric., no. xxxx, pp. 1–9, 2020.

[19] Francois Fleuret, “EE-559 – Deep learning,” idiap Res., 2020.

[20] D. Čeponis and N. Goranin, “Investigation of dual-flow deep learning models LSTM-FCN and GRU-FCN efficiency against single-flow CNN models for the host-based intrusion and malware detection task on univariate times series data,” Appl. Sci., vol. 10, no. 7, 2020.

[21] A. Novikov, P. Izmailov, V. Khrulkov, M. Figurnov, and I. Oseledets, “Tensor train decomposition on tensorflow (T3F),” J. Mach. Learn. Res., vol. 21, pp. 1–7, 2020.

[22] S. Lagouvardos, J. Dolby, and N. Grech, “Static Analysis of Shape in TensorFlow Programs,” 34th Eur. Conf. Object-Oriented Program. (ECOO2020), no. 15, pp. 1–15, 2020.

[23] H. Andrade-Loarca and G. Kutyniok, “tfShearlab: The TensorFlow Digital Shearlet Transform for Deep Learning,” pp. 1–25, 2020.

[24] P. Alonso-Jimenez, D. Bogdanov, J. Pons, and X. Serra, “Tensorflow Audio Models in Essentia,” no. i, pp. 266–270, 2020.

[25] A. Bayu Aji and I. Surjandari, “Hybrid vector autoregression-recurrent neural networks to forecast multivariate time series jet fuel transaction price,” IOP Conf. Ser. Mater. Sci. Eng., vol. 909, no. 1, 2020.