Modeling Monthly Rainfall in Malang using Long Short Term Memory

Eni Sumarminingsih[1*], Suci Astutik[2], Nur Silviyah Rahmi[3], Aqsa Yudhistira Redi[4], Natasha Aulia[5]

[1,2,3,4,5] *Universitas Brawijaya*
[*] *Corresponding author. Email: eni_stat@ub.ac.id*

ABSTRACT
Malang Regency has the potential to experience flooding. The flood was caused by the high rainfall that occurred in the area. The purpose of this study is to model monthly rainfall at three stations in Malang, namely Abd. Saleh, Karangkates station and Karangploso station use Long Short Term Memory (LSTM). The results of this study (1) The best model for monthly rainfall at Station Abd. Saleh uses Station And rainfall input. Saleh lag 6, (2) The best model for monthly rainfall at Karangploso Station using rainfall input at Karangploso Station lag 6 and (3) The best model for monthly rainfall at Karangkates Station using rainfall input at Karangkates Station lag 1, lag 4 and lag 6, rainfall at Abd. Saleh Air Base Station lag 1, lag 4 and lag 6 as well as rainfall at Karangploso Station lag 1, lag 4 and lag 6

Keywords: *Modeling, Rainfall, Long Short Term Memory*

1. INTRODUCTION

One area in Indonesia that has the potential to experience flooding is Malang Regency. The flood was caused by the high rainfall that occurred in the area. Floods can cause various problems including damage to agricultural land. Therefore, forecasting rainfall is necessary to anticipate the impact of flooding.

Precipitation forecasting is quite difficult because it has a space-time dependence. Several monthly rainfall models that have been carried out are supported by Charaniya, N. et al.[1] and Suhartono et al.[2] . Several researchers, namely Diani, K. et al [3] ; Safitri, B. et al [4] ; Simamora, R. et al[5] ; Sulistyono, A. et al [6] ; Sumarminingsih, E. et al [7] ; Sumaarminingsih, E. et al [8] have conducted rainfall modeling in Malang Regency. The methods used in this research include ARIMA(Autoregressive Integrated Moving Average), VAR(Vector Autoregressive), GSTAR (Generalized Spatio-Temporal Autoregressive) multilayer perceptron neural network, extreme learning machine, Feedforward Neural Network, Hybrid ARIMA-NN, VAR -NN and SpVAR-NN and ensemble method based on ANFIS-ARIMA.

In this research, monthly rainfall modeling will be carried out in Malang district, namely the Abd. Saleh Air Base Station, Karangploso Station and Karangkates Station. The method used is Long Short Term Memory (LSTM), which is a method with a machine learning approach that is suitable for time series data. Several studies that have used LSTM in rainfall modeling such as Chen, S et al.[9] and Liu, Y. et al. [10] show that LSTM has good performance. According to Suhartono et al. [11] selecting the right input will determine forecasting accuracy. Based on several studies on rainfall that rainfall has a spatio-temporal relationship, the input used in this study is the rainfall lag in the same location and the rainfall lag in neighboring locations.

2. DATA

Y. A. Yusran et al. (eds.), *Proceedings of the 2023 Brawijaya International Conference (BIC 2023)*, Advances in Economics, Business and Management Research 294,
https://doi.org/10.2991/978-94-6463-525-6_18

The research data, namely monthly rainfall (January 2015 – December 2021) for the Abd. Saleh air base station, Karangploso station and Karangkates station were obtained from the site https://malangkab.bps.go.id

3. METHOD

3.1. Long Short Term Memory (LSTM)

Long Short Term Memory (LSTM) is a machine learning approach designed for time series data. The LSTM consists of memory cells, forget gates, input gates and output gates. Memory cells are useful for storing long-term information. Input Gate and output gate are used to control the flow of information to and from memory cells. The first step is to calculate the output value of the previous time and the value of the current time. These two values are used as input for the forget gate and input gate. The second step is to calculate the forget gate and input gate. Calculations on the forget gate using the equation

$$f_t = \sigma\left(W_f.\left[l_{t-1}, y_t\right] + b_f\right)\frac{-b\pm\sqrt{b^2-4ac}}{2a} \quad (1)$$

$f_t = \sigma\left(W_f.\left[l_{t-1}, y_t\right] + b_f\right)$Calculations at the input gate using the equation

$$\frac{n!}{r!(n-r)!}i_t = \sigma\left(W_i.\left[l_{t-1}, y_t\right] + b_i\right) \quad (2)$$

$$C_t^{'} = \tanh tanh\left(W_c.\left[l_{t-1}, y_t\right] + b_c\right) \quad (3)$$

Where

f_t: forget gate at time t, the value range of f_t is (0,1)

i_t: input gate at time t, the value range of i_t is (0,1)

$C_t^{'}$: candidate cell at input gate at time t

W_f: weight of forget gate

W_i: weight of input gate

W_c: the weight of the candidate cell at the input gate

l_{t-1}: the output value of processing at a previous time

y_t: input value at time t

b_f: the bias value applied to the forget gate

b_i: bias value applied to the input gate

b_c: the bias value applied to the input gate candidate cells

Tanh : Hyperbolic tan activation function with a function using the equation (4)

$$f(x) = \frac{\exp exp\ (2y)-1}{\exp exp\ (2y)+1} \quad (4)$$

σ : sigmoid activation function with a function using the equation (5)

$$f(x) = \frac{1}{1+exp(-y)} \quad (5)$$

The next step is to adjust the cell values or model parameters. The formula for the calculation in this step is

$$C_t = f_t \times C_{t-1} + i_t \times C_t^{'} \quad (6)$$

The output value is obtained from

$$O_t = \sigma\left(W_o\left[l_{t-1}, y_t\right] + b_o\right) \quad (7)$$

Where

O_t: output gate at time t, value range of O_t is (0,1)

W_o: weight of output gate

b_o: bias value applied to output gate

The final output value of the output gate is obtained based on the formula

$$l_t = O_t \times \tanh tanh\left(C_t\right) \quad (8)$$

Where

l_t: the final output value of the LSTM at the time t by the output gate.

After obtaining several LSTM models from several inputs, the next step is to select the best model using the smallest MAPE (Mean Absolute Percentage Error) criteria. The MAPE formula is

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100\% \quad (9)$$

Input determination is important in LSTM modeling. In LSTM, one of the inputs that can be used is the lag of the variable. To determine the lag variable used as input, Partial Autocorrelation Function (PACF) is used. Meanwhile, to determine other variables as input, correlation is used.

3. 2. Partial Autocorrelation (PACF)

PACF denoted by ϕ_{kk} is the correlation between Y_t and Y_{t+k} after the linear relationship between Y_t $dengan$ $Y_{t+1}, Y_{t+2}, ..., Y_{t+k-1}$ is omitted. According to (Wei, 2006) PACF can be calculated recursively starting from $\hat{\phi}_{11} = \hat{\rho}_1$ and $\hat{\phi}_{kk}$ can be calculated from the following formula

$$\hat{\phi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^{k}\hat{\phi}_{kj}\hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^{k}\hat{\phi}_{kj}\hat{\rho}_j} \quad (10)$$

and

$$\hat{\phi}_{k+1,j} = \hat{\phi}_{kj} - \hat{\phi}_{k+1,k+1}\hat{\phi}_{k,k+1-j}, \; j = 1, 2, ..., k \quad (11)$$

Where

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k}\left(Y_t - \bar{Y}\right)\left(Y_{t+k} - \bar{Y}\right)}{\sum_{t=1}^{n}\left(Y_t - \bar{Y}\right)^2}, \; k = 0, 1, 2, ... \quad (12)$$

3.3. Correlation

Correlation is a statistical measure to determine the closeness of the relationship between two variables. The correlation coefficient is between -1 to 1. The closer to -1 means the stronger the negative relationship. The closer to 1 means the stronger the positive relationship. The correlation coefficient is close to zero indicating that there is no relationship between variables. The correlation coefficient can be calculated from the following equation

$$r_{XY} = \frac{n\sum_{i=1}^{n}X_iY_i - \sum_{i=1}^{n}X_i\sum_{i=1}^{n}Y_i}{\sqrt{n\sum_{i=1}^{n}X_i^2 - \left(\sum_{i=1}^{n}X_i\right)^2}\sqrt{n\sum_{i=1}^{n}Y_i^2 - \left(\sum_{i=1}^{n}Y_i\right)^2}} \quad (13)$$

4. Results and Discussion

The first step in analyzing time series data is plotting the data. This is useful to know the characteristics or patterns that exist in the data. The rainfall plots at the Abd. Saleh Air Base Station, at the Karangkates Station and at the Karangploso Station are presented in Figure 1, Figure 2, and Figure 3 respectively. The three figures show that there is a seasonal pattern in the rainfall data at the three stations.

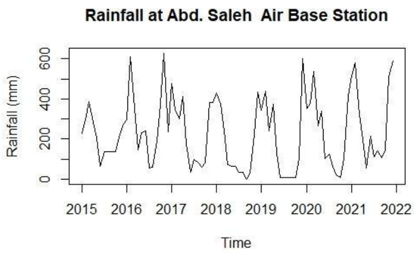

Figure 1 Plot of Rainfall at Abd. Saleh Air Base Station

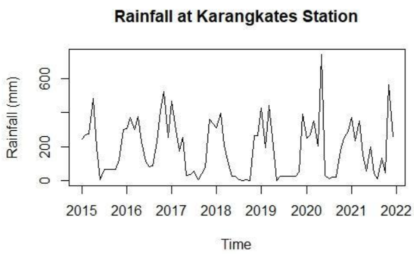

Figure 2 Plot of Rainfall at Karangkates Station

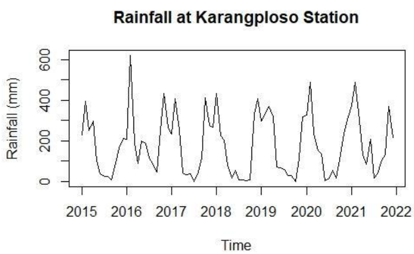

Figure 3 Plot of Rainfall at Karangploso Station

PACF is used to determine the input of lag variables for rainfall. Figure 4, Figure 5, and Figure 6 present the PACF plots of the rainfall at Abd. Saleh Air Base Station, Karangkates Station and Karangploso Station respectively. Figure 4 and Figure 5 show that the significant lags are lag 1, lag 4 and lag 6. Meanwhile Figure 6 shows that the significant lags are lag 1, lag 4, lag 5, lag 6 and lag 7. Based on Figure 4, Figure 5, and Figure 6, the input used is lag 1, lag 4 and lag 6.

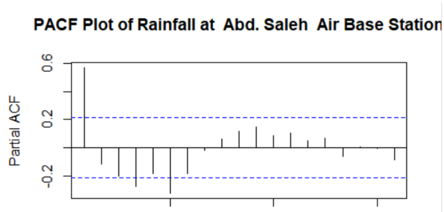

Figure 4 PACF Plot of Rainfall at Abd. Saleh Air Base Station

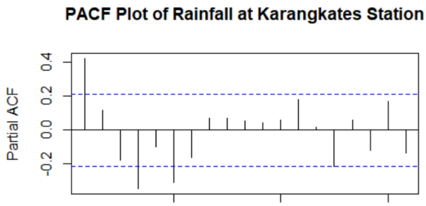

Figure 5 PACF Plot of Rainfall at Karangploso Station

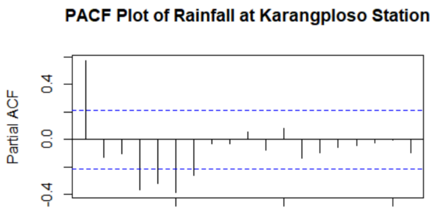

Figure 6 PACF Plot of Rainfall at Karangploso Station

Rainfall has a spatio-temporal relationship. Therefore, other station rainfall is also used as input. The plot of the relationship between rainfall at one station and another is presented in Figure 7, Figure 8 and Figure 9. All three figures show that there is a relationship between rainfall between stations.

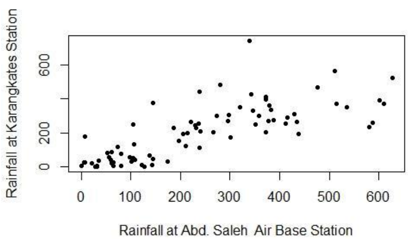

Figure 7 Plot of Rainfall at Abd. Saleh Air base Station vs at Karangkates Station

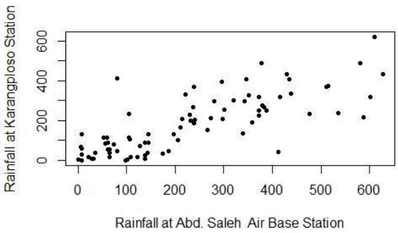

Figure 8 Plot of Rainfall at Abd. Saleh Air base Station vs at Karangploso Station

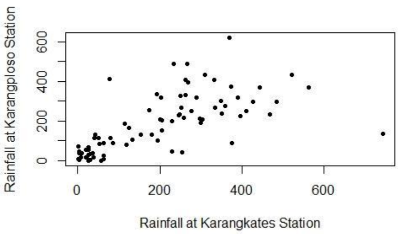

Figure 9 Plot of Rainfall at Karangkates Station vs at Karangploso Station

The relationship of rainfall between stations can also be seen from the correlation of rainfall between stations. The correlation coefficient of rainfall between locations is presented in Table 1 and it can be seen that the correlation of rainfall between locations is high.

Table 1. Correlation of Rainfall between Station

Correlation Coefficient	Rainfall of Abd. Saleh Air Base Satation	Rainfall of Karangkates Station	Rainfall of Karangploso Station
Rainfall of Abd. Saleh Air Base Satation	1	0.777	0.784
Rainfall of Karangkates Station	0.777	1	0.686
Rainfall of Karangploso Station	0.784	0.686	1

Based on the discussion about PACF and the correlation coefficient, the inputs used are lag 1, lag 4 and lag 6 rainfall of the station itself and lag 1, lag 4 and lag 6 rainfall at other (neighboring) stations. A comparison of the MAP of several models with different inputs for the rainfall at the Abd. Saleh Airbase Station is presented in Table 2.

Table 2. MAP of Rainfall at Abd. Saleh Air Base Station Model

Input	MAPE
Rainfall at Abd. Saleh Air Base Station lag 1	1,4318
Rainfall at Abd. Saleh Air Base Station lag 4	0,7797
Rainfall at Abd. Saleh Air Base Station lag 6	0,7156
Rainfall at Abd. Saleh Air Base Station lag 1, lag 4 and lag 6	1,4668
Rainfall at Abd. Saleh Air Base Station lag 1, rainfall at Karangploso Station lag 1, rainfall Karangkates Station lag 1	1,2426
Rainfall at Abd. Saleh Air Base Station lag 4, rainfall at Karangploso Station lag 4, rainfall at Karangkates Station lag 4	1,7215
Rainfall at Abd. Saleh Air Base Station lag 6, rainfall at Karangploso Station lag 6, rainfall at Karangkates Station lag 6	1,4911
Rainfall at Abd. Saleh Air Base Station lag 1, lag 4 and lag 6, rainfall at Karangploso Station lag 1, lag 4 and lag 6, rainfall at Karangkates Station lag 1, lag 4 and lag 6	0,8864

Based on Table 2, it can be seen that the best model for the rainfall at the Abd. Saleh Airbase Station is the model with the input of Rainfall at the Abd. Saleh Air Base Station Lag 6 with a MAP of 0.7156.

Table 3. MAP of Rainfall at Karangkates Station Model

Input	MAPE
Rainfall at Karangkates Station lag 1	1,6522
Rainfall at Karangkates Station lag 4	1,7710
Rainfall at Karangkates Station lag 6	1,8147
Rainfall at Karangkates Station lag 1, lag 4 and lag 6	2,4739
Rainfall at Abd. Saleh Air Base Station lag 1, rainfall at Karangploso Station lag lag 1, rainfall at Karangkates Station lag 1	1,5855

Rainfall at Abd. Saleh Air Base Station lag 4, rainfall at Karangploso Station lag 4, rainfall at Karangkates Station lag 4	1,7541
Rainfall at Abd. Saleh Air Base Station lag 6, rainfall at Karangploso Station lag 6, rainfall at Karangkates Station lag 6	1,0661
Rainfall at Abd. Saleh Air Base Station lag 1, lag 4 and lag 6, rainfall at Karangploso Station lag 1, lag 4 and lag 6, rainfall at Karangkates Station lag 1, lag 4 and lag 6	0,4593

Based on Table 3, it can be seen that the best model for rainfall at Karangkates Station is a model with input rainfall at Karangkates Station lag 1, lag 4 and lag 6, rainfall at Abd. Saleh Air Base Station lag 1, lag 4 and lag 6 as well as rainfall at Karangploso Station lag 1, lag 4 and lag 6

Table 4. MAP of Rainfall at Karangploso Station Model

Input	MAPE
Rainfall at Karangploso Station lag 1	1,2038
Rainfall at Karangploso Station lag 4	1,2405
Rainfall at Karangploso Station lag 6	0,4571
Rainfall at Karangploso Station lag 1, lag 4 and lag 6	1,5603
Rainfall at Abd. Saleh Air Base Station lag 1, rainfall at Karangploso Station lag 1, rainfall at Karangkates Station lag 1	1,2537
Rainfall at Abd. Saleh Air Base Station lag 4, rainfall at Karangploso Station lag 4, rainfall at Karangkates Station lag 4	1,9120
Rainfall at Abd. Saleh Air Base Station lag 6, rainfall at Karangploso Station lag 6, rainfall at Karangkates Station lag 6	1,2799
Rainfall at Abd. Saleh Air Base Station lag 1, lag 4 dan lag 6, rainfall at Karangploso Station lag 1, lag 4 dan lag 6, rainfall at Karangkates Station lag 1, lag 4 dan lag 6	0,4991

Based on Table 4, the best model for rainfall at Karangploso Station is a model with an input rainfall Karangploso Station lag 6. The best model for rainfall at Karangploso Station is similar to the best model for rainfall at Abd. Saleh Air Base Station, namely with an input lag 6 from rainfall at the station itself. While the best model for rainfall at Karangkates Station is a model with quite complex inputs, namely rainfall at Karangkates Station and rainfall at other stations, namely Abd. Saleh Air Base Station and Karangploso Station lag 1, lag 4 and lag 6. The rainfall at the Karangploso Station and the Abd. Saleh Air Base Station has similar characteristics because the distance between the two stations is closer than the Karangkates Station. Actual data and predictive data based on the best model for rainfall at Abd. Saleh Air Base Station, Karangkates Station and Karangploso Station are presented in Figure 10, Figure 11 and Figure 12 respectively.

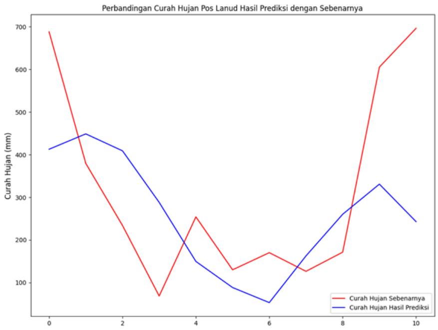

Figure 10 Plot of predicted vs actual of Rainfall at Abd. Saleh Air Base Station

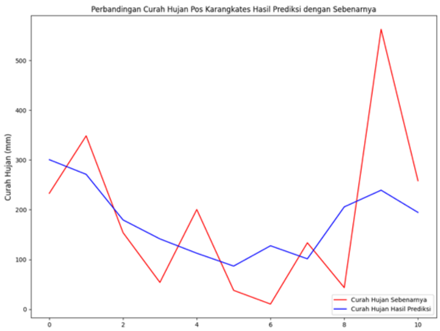

Figure 11 Plot of predicted vs actual of Rainfall at Karangkates Station

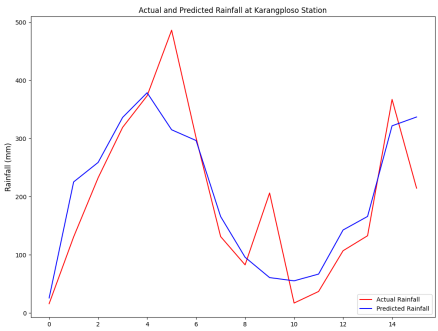

Figure 12 Plot of predicted vs actual of Rainfall at Karangploso Station

Figure 10, Figure 11 and Figure 12 show that the predicted data and actual data are quite close. This shows that the resulting model is quite good to use.

5. CONCLUSION

The best monthly rainfall models at the Abd. Saleh Air Base Station and Karangploso Station have the same input, namely the rainfall lag 6 at the station itself. While the best monthly rainfall model for the Karangkates Station uses input at the Abd. Saleh Air Base Station lag 1, lag 4 and lag 6, the rainfall at the Karangploso Station lag 1, lag 4 and lag 6, and the rainfall at the Karangkates Station lag 1, lag 4 and lag 6. In the rainfall at the Karangkates Station, there is a spatio-temporal relationship, while the rainfall of the Abd. Saleh Air Base Station and Karangploso Station only have a temporal relationship. However, all the three exhibit seasonal characteristics.

ACKNOWLEDGMENTS

My deepest gratitude to the Faculty of Mathematics and Natural Science, Universitas Brawijaya, which has funded this research through the Hibah Doktor Lektor Kepala program.

REFERENCES

[1] Charaniya, N. A., & Dudul, S. V. (2011). Committee of artificial neural networks for monthly rainfall prediction using wavelet transform. *ICBEIA 2011 - 2011 International Conference on Business, Engineering and Industrial Applications*, 125–129. https://doi.org/10.1109/ICBEIA.2011.5994225

[2] Suhartono, Faulina, R., Lusia, D. A., Otok, B. W., Sutikno, & Kuswanto, H. (2012). Ensemble method based on ANFIS-ARIMA for rainfall prediction. *ICSBE 2012 - Proceedings, 2012 International Conference on Statistics in Science, Business and Engineering: "Empowering Decision Making with Statistical Sciences,"* 240–243. https://doi.org/10.1109/ICSSBE.2012.6396564

[3] Diani, K. A. N., Setiawan, & Suhartono. (2013). Pemodelan VAR-NN dan GSTAR-NN untuk Peramalan Curah Hujan di Kabupaten Malang. *Jurnal Sains Dan Seni POMITS*, *2*(1), D31-D-36.

[4] Safitri, B. A., Iriany, A., & Wardhani, N. W. S. (2021). Perbandingan Akurasi Peramalan Curah Hujan dengan menggunakan ARIMA, Hybrid ARIMA-NN, dan FFNN di Kabupaten Malang (Comparison of Rainfall Forecasting Accuracy using ARIMA, Hybrid ARIMA-NN, and FFNN in Malang District). *Seminar Nasional Official Statistics*, 245–253.

[5] Simamora, R. J. D., Tibyani, & Sutrisno. (2019). Peramalan Curah Hujan Menggunakan Metode Extreme Learning Machine. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *3*(10), 9670–9676. http://j-ptiik.ub.ac.id

[6] Sulistyono, A. D., Iriany, A., Iriany, A., & Suryawardhani, N. W. (2018). Perbandingan Bobot Normalisasi Cross Correlation dengan Cross Covariance dalam Model Peramalan Curah Hujan. *Prosiding Seminar Nasional VI Hayati* .

[7] Sumarminingsih, E., Lusia, D. A., Wahyuni, D., & Wibawa, T. S. (2022). Spatial Vector Autoregressive-Neural Network (SpVAR-NN) Model to Predict Rainfall and other Weather Variables in Malang and Karangkates. *Indonesia. International Journal of Agricultural and Statistical Sciences. Int. J. Agricult. Stat. Sci*, *18*(2), 557–568.

[8] Sumarminingsih, E., Solimun, Wahyuni, D., & Wibawa, T. S. (2022). Vector Autoregressive Model with Seasonal Indicator and Feed Forward Neural

[9] Chen, S., Adjei, C. O., Tian, W., Onzo, B.-M., Adu, E., Kedjanyi, G., & Darteh, O. F. (2021). Rainfall Forecasting in Sub-Sahara Africa-Ghana using LSTM Deep Learning Approach. *International Journal of Engineering Research & Technology*, *10*(3), 464–470. www.ijert.org

[10] Liu, Y., Zhang, W., Yan, Y., Li, Z., Xia, Y., & Song, S. (2022). An Effective Rainfall–Ponding Multi-Step Prediction Model Based on LSTM for Urban Waterlogging Points. *Applied Sciences (Switzerland)*, *12*(23). https://doi.org/10.3390/app122312334

[11] Suhartono, Suhermi, N., & Prastyo, D. D. (2018). Design of Experiment to Optimize the Architecture of Deep Learning for Nonlinear Time Series Forecasting. *Procedia Computer Science*, *144*, 269–276. https://doi.org/10.1016/j.procs.2018.10.528