Comparison of Monthly Rainfall Prediction using Long Short Term Memory and Multi Layer Perceptron Methods in South Tangerang City

GA Monang Lumban Gaol[1], Mohammad Syafrullah[2], Supardi[3]
Fakultas Teknologi Informasi, Program Studi Magister Ilmu Komputer[1], [2], [3]
Universitas Budi Luhur[1], [2], Institut Sains dan Bisnis Atma Luhur[3]
Kota Jakarta Selatan[1], [2] , Kota Pangkal Pinang[3], Indonesia
E-mail: 2111600371@student.budiluhur.ac.id[1], mohammad.syafrullah@budiluhur.ac.id[2], supardi@atmaluhur.ac.id[3]

***Abstract*— Rainfall is one of the meteorological and climatological parameters whose information must be disseminated to the public and related stakeholders. Rainfall information has an important role in the sectors of people's lives. In agriculture, the amount of rainfall has an important role in determining the planting season, so that this can prevent potential crop failure. On Disaster, South Tangerang City during the 2016-2021 period experienced floods, landslides, and droughts. Therefore, the importance of rainfall prediction information can improve meteorological and climatological information services in various sectors. Nevertheless, it is still difficult for the community and stakeholders to get monthly rainfall predictions with high accuracy in the long term. In this research, monthly rainfall prediction is designed using MLP (Multi Layer Perceptron) and LSTM (Long Short Term Memory). The data used is the monthly rainfall data of Climate Hazards Group InfraRed Precipitations (CHIRPS) for 42 years (period 1981-2022) with coordinate boundaries according to the research location, namely South Tangerang City, which is located between 106.625 ° - 106.825 ° East and 6.4 ° - 6.2 ° LS as many as 16 grids with a resolution of 0.05 ° each grid. Monthly rainfall prediction using MLP produces an RMSE value of 90.19, and a MAPE of 40.55, while the LSTM method produces an RMSE value of 88.12 and a MAPE of 40.49. Monthly rainfall prediction results using the LSTM method are better than the MLP method; this can be seen from the RMSE value of the LSTM method is smaller than MLP.**

Keywords*— *rainfall, prediction, LSTM, MLP

I. INTRODUCTION

The more comprehensive society's request for rainfall prediction information, especially for related agencies and decision-making stakeholders in various sectors, such as agriculture, infrastructure development, and disaster, has increased. In agriculture, rainfall has an essential role in determining the planting season [1], so this can prevent potential crop failure. Floods and landslides from high rainfall can cause losses [2]. Rainfall prediction information with a more extended period in the future, more accurate, and faster dissemination is a new challenge in meeting the increasing need for information. The design of a deep learning-based rainfall prediction system is expected to be one of the supporting forces in improving services related to meteorological and climatological information.

Monthly rainfall forecast information over a long period has become a significant requirement for institutions and communities in various sectors that generally require monthly rainfall forecast/prediction information for more than six months, such as one year or five years, as information to support strategic planning. Monthly rainfall prediction information is also required to have a reasonably high accuracy, where currently the accuracy of BMKG rainfall forecast/prediction results ranges from 61-87% (BMKG empirical data), so there is still room for improvement in accuracy with a range of better prediction accuracy levels.

The difficulty of society and stakeholders in getting high-accuracy monthly rainfall predictions in the long term to plan future activities motivated the author to conduct this research. The output of this research is expected to provide a choice of solutions to these difficulties.

II. RESEARCH METHOD

A. Study Area

The research is located in South Tangerang City, Banten Province. South Tangerang City is located in the eastern part of Banten Province, an expansion of Tangerang Regency. Administratively, South Tangerang City consists of 7 (seven) sub-districts, including Ciputat, East Ciputat, Pamulang, Pondok Aren, Serpong, North Serpong, and Setu, and 54 urban villages. While geographically, it is located at the coordinates 106º 38' - 106º 47' East Longitude and 06º 13'30 "- 06º 22'00" South latitude, with the following boundaries: to the north, it borders Tangerang City, to the east it borders DKI Jakarta Province, to the south it borders Bogor Regency and Depok City of West Java Province, and to the west it borders Tangerang Regency.

B. Data

This study uses secondary data in the form of Climate Hazards Group InfraRed Precipitations (CHIRPS) monthly

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2149, Copyright ©2024
Submitted : April 25, 2024, Revised : June 13, 2024, Accepted : June 20, 2024, Published : June 30, 2024

rainfall data for 42 years (1981-2022) with coordinate boundaries according to the research location, namely South Tangerang City, between 106.625º - 106.825º East and 6.4º - 6.2º LS as many as 16 grids as shown in Figure 1. The data was downloaded through the IRI Data Library, with the website address : http://iridl.ldeo.columbia.edu/SOURCES/.UCSB/.CHIRPS/.v2p0/.monthly/.global/.precipitation/.CHIRPS monthly rainfall data is reanalysis data, a combination of observation station rainfall data and rainfall data derived from satellite estimation results covering almost all land on earth (50º LU—50º LS) and available on a daily, five-day / pentad and monthly time scale from 1981 to the present [3]. CHIRPS data has a relatively high spatial resolution of 0.05º or approximately 5 x 5km.

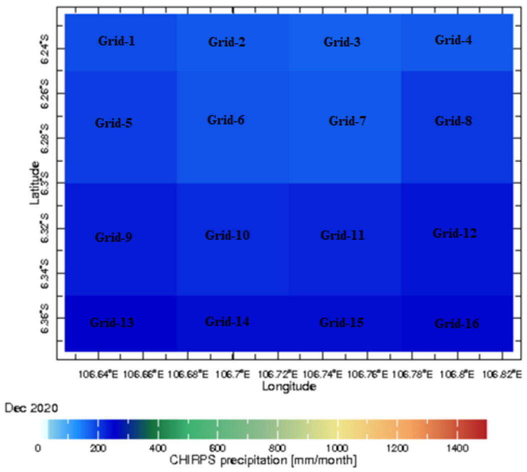


Figure 1. Rainfall Data in December 2020 with South Tangerang City Coordinates in Grid Form (Source: IRIDL)

C. *Flowchart*

Flowchart of this study show technical diagram of data analysis in Figure 2, where there are 2 (two) deep learning architectures that will be built to predict monthly rainfall.

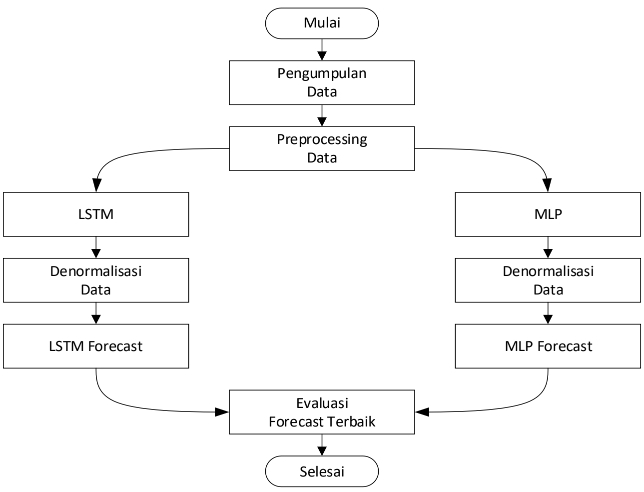


Figure 2. Flowchart of this study

Data preprocessing will be carried out on the collected rainfall data, which includes the normalization stage with the activation function, followed by the stage of dividing the composition of training data (training data) by 80% and testing data (testing data) by 20% [4], [5].

III. Result And Discussion

A. *Result of Data Preprocessing*

The data used in this study is spatially obtained rainfall data. The data is reanalysis data monthly, resulting in monthly rainfall data. The total amount of monthly rainfall data available is 504 data. Table 1 shows the time series of monthly rainfall data on each grid, which is 16 grids.

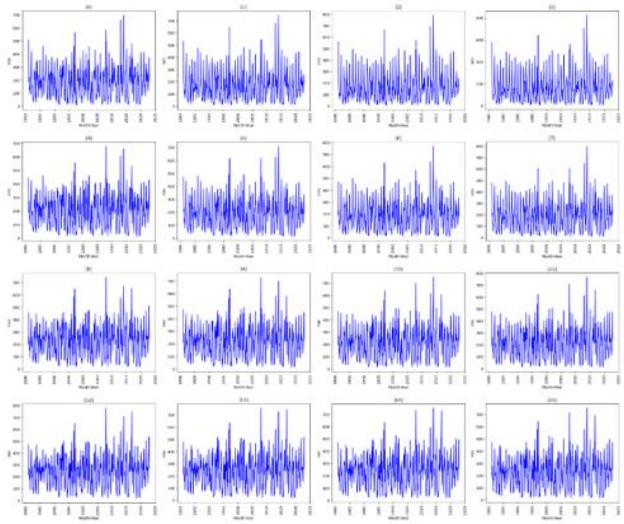


Figure 3. Monthly Rainfall data in 1981 -2022

Figure 3 presents a significant finding, demonstrating the similarity of monthly rainfall data patterns across 16 grids from 1981 to 2022. Despite this consistency, the amount of rainfall in each grid varies, a phenomenon attributed to extreme weather, uneven rain, and geographical factors [6].

Bulk data patterns can be used to forecast monthly rainfall in the next period [7]. The method used in this research uses Multiple-Layer Perceptron (MLP) and Long-Short-Term Memory (LSTM). Before modeling, it is necessary to compose training data and test data. The composition of the data used in this study is 403 data (80%) as training data and 101 data (20%) as test data.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2149, Copyright ©2024
Submitted : April 25, 2024, Revised : June 13, 2024, Accepted : June 20, 2024, Published : June 30, 2024

TABLE i. Descriptive Statistics of Monthly Rainfall Time Series Data January 1981-December 2022

No	Grid	Rainfall		
		Average	*Maximum*	*Minimum*
1	Grid 1	182.9	696.8	7.0
2	Grid 2	166.3	744.0	5.8
3	Grid 3	159.4	791.0	4.8
4	Grid 4	160.9	823.0	4.4
5	Grid 5	202.1	678.5	13.6
6	Grid 6	189.1	710.1	12.8
7	Grid 7	185.9	796.2	12.6
8	Grid 8	185.2	799.1	10.2
9	Grid 9	224.6	744.9	14.8
10	Grid 10	216.7	729.9	14.3
11	Grid 11	213.7	747.8	13.7
12	Grid 12	209.7	767.0	12.0
13	Grid 13	241.8	779.2	17.9
14	Grid 14	237.9	756.7	17.2
15	Grid 15	234.4	750.4	17.4
16	Grid 16	224.2	752.2	14.5

Based on the processing and statistical analysis using time series data with average, maximum, and minimum rainfall parameters in Table 1, it is found that there are data variations from each grid, then plotting the data in the graph for each grid in Figure 3.

B. Result of Multiple Layer Perceptron (MLP) Modeling

MLP consists of an input layer, hidden layer, and output layer. MLP processes inputs through a network of neurons with specific adaptive weighting. The MLP output is a combination of activation function, input weighting, and bias, which is mathematically expressed as follows [8]:

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{1}$$

Where y is the output, x is the input, w is the weight, and b is the bias. The function is a general form of activation function. This research uses non-linear activation functions, namely sigmoid, tangential hyperbolic (tanh), and rectified linear unit functions [9]. The sigmoid, tanh, and relu functions are respectively expressed as follows [10]:

$$f_s = f(x) = \frac{1}{1+e^{-x}} \tag{2}$$

$$f_t = tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{3}$$

$$f_r = f(x) = max(0, x) \tag{4}$$

The fs function is a sigmoid function equation, the ft function is a hyperbolic tangential function equation, and the fr function is a rectified linear unit function. The initial stage of MLP modeling is to determine simple hyperparameter values. The level of hyperparameter complexity determines the algorithm's computation time. Table 2 shows the hyperparameter composition of the MLP model [11].

TABLE ii. Composition of MLP Hyperparameter

Number of Layer	100
Number of Layer	50
Batch Size	6
Epoch	150

This study uses the relay activation function, a batch size of 150, and six batch sizes. The data composition used is 403 data (80%) as training data and 101 data (20%) as testing data on 16 grids of monthly bulk data. Table 3 shows the RMSE values of training data and testing data, as well as the MAPE value of testing data.

TABLE iii. Result Test of MLP Performance

No	Grid	RMSE	MAPE
1	Grid 1	77.03	0.80
2	Grid 2	65.22	0.79
3	Grid 3	64.52	0.79
4	Grid 4	63.91	0.80
5	Grid 5	86.58	0.82
6	Grid 6	72.68	0.84
7	Grid 7	74.72	0.81
8	Grid 8	82.09	0.80
9	Grid 9	97.57	0.76
10	Grid 10	88.60	0.81
11	Grid 11	91.29	0.78
12	Grid 12	97.04	0.79
13	Grid 13	108.63	0.75
14	Grid 14	108.05	0.80
15	Grid 15	106.68	0.75
16	Grid 16	100.10	0.74
	Averagae	**86.54**	**0.79**
	Maximum	**108.63**	**0.84**
	Minimum	**63.91**	**0.74**

Table 3 shows the results of the MLP model performance test in predicting monthly rainfall on 16 data grids in South Tangerang City. The average RMSE value is 86.54, and the average MAPE value is 0.79. The highest RMSE value occurred in Grid 1 of, 108.63. The highest MAPE value was 0.84 on Grid 6. The lowest RMSE value was 63.91 on Grid 4, and the lowest MAPE was 0.74 on Grid 16.

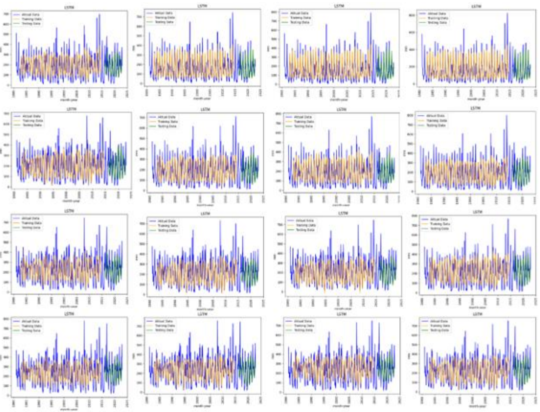


Figure 4. MLP Performance

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2149, Copyright ©2024
Submitted : April 25, 2024, Revised : June 13, 2024, Accepted : June 20, 2024, Published : June 30, 2024

Figure 4 plots performance data test results using the MLP method. The blue line pattern results from accurate monthly rainfall data, the orange line results from training data, and the green color results from testing data. Overall, the processed results on the grid show the same pattern. However, the results of training and testing data have a range of data that is quite far from the actual data due to extreme rain events in specific periods with a high enough accumulation, resulting in a high error value and the number of datasets that are considered low enough to produce a high RMSE value.

C. Result of Long Short Term Memory (LSTM) Modeling

Predicting monthly rainfall is based on the LSTM algorithm, which uses a data composition of 403 data (80%) as training data and 101 data (20%) as testing data on 16 data grids. This study uses the number of hidden layer neurons of as many as 256, an epoch of as many as 150, a batch size of 6, and the Adam optimizer function. Table 4 shows the results of the LSTM algorithm processing used to determine accuracy and performance using RMSE and MAPE indicators.

Tabel 4. Result Test of LSTM Performance

No	Grid	RMSE	MAPE
1	Grid 1	82.45	0.86
2	Grid 2	65.07	0.73
3	Grid 3	61.41	0.77
4	Grid 4	64.64	0.78
5	Grid 5	91.03	0.80
6	Grid 6	77.98	0.85
7	Grid 7	76.28	0.83
8	Grid 8	78.53	0.83
9	Grid 9	100.04	0.81
10	Grid 10	90.13	0.78
11	Grid 11	93.30	0.75
12	Grid 12	92.53	0.82
13	Grid 13	106.20	0.76
14	Grid 14	102.26	0.74
15	Grid 15	102.73	0.82
16	Grid 16	96.89	0.74
Averagae		**86.34**	**0.79**
Maximum		**106.20**	**0.86**
Minimum		**61.41**	**0.73**

Table 4 shows the results of the LSTM model performance test in predicting monthly rainfall on 16 data grids in South Tangerang City. The average RMSE value is 86.34, and the average MAPE value is 0.79. The highest RMSE value occurs on Grid 13 of 106.20. The highest MAPE value was 0.86 on Grid 1. The lowest RMSE value was 61.41 on Grid 3, and the lowest MAPE on Grid 2 was 0.73.

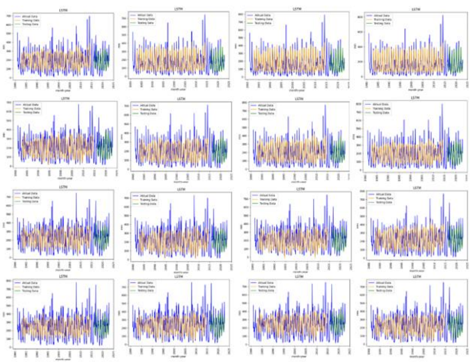
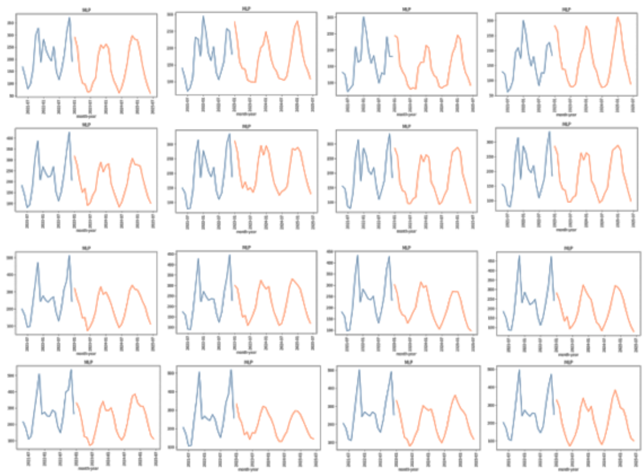


Figure 5. MLP Performance

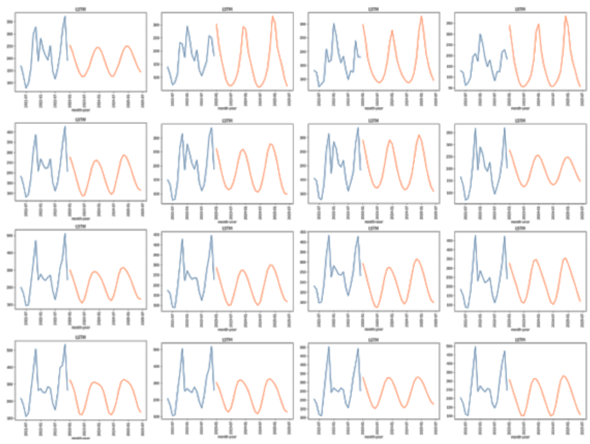
Figure 5 plots the performance data of the LSTM test results. The blue line pattern results from accurate monthly rainfall data, the orange line results from training data, and the green color results from testing data. The data processing results using LSTM modeling produce the same results as MLP modeling due to extreme rain events in a certain period with a high enough accumulation, resulting in a high enough error value and the number of datasets considered low enough to produce a high enough RMSE value.

D. Result of Monthly Prediction Rainfall using LSTM and MLP

The monthly rainfall prediction model has been designed; the next step is to show the results of monthly rainfall predictions in South Tangerang City in graphical form. The prediction results are monthly rainfall predictions from January 2023 to June 2025, then combined with actual data plots from January 1981 to December 2022. Figure 6 shows the results of the MLP and LSTM prediction models.



(a)



(b)

Figure 6. Result of Monthly Rainfall Prediction Using Method (a) MLP ; (b) LSTM

Figure 6 shows the plot results of the MLP and LSTM model data for monthly rainfall from 1981 to 2022 and the next prediction model from January 2023 to June 2025. The blue

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2149, Copyright ©2024
Submitted : April 25, 2024, Revised : June 13, 2024, Accepted : June 20, 2024, Published : June 30, 2024

graph results from the actual monthly rainfall data plot, and the orange graph results from the predicted rainfall data plot using MLP and LSTM. The prediction model designed for monthly rainfall was chosen because it considers the need for monthly rainfall data in extreme climate mitigation.

Tabel 5. Result Test of LSTM Performance

No	Grid	MLP		LSTM	
		RMSE	**MAPE**	**RMSE**	**MAPE**
1	Grid 1	87.87	57.10	79.76	73.78
2	Grid 2	99.29	73.71	104.79	66.36
3	Grid 3	106.74	56.93	114.68	62.96
4	Grid 4	95.34	34.99	117.52	50.40
5	Grid 5	87.94	50.50	80.21	42.38
6	Grid 6	76.84	42.69	76.46	36.91
7	Grid 7	90.17	39.98	89.88	41.40
8	Grid 8	79.45	31.76	82.26	31.38
9	Grid 9	89.03	29.48	77.65	31.17
10	Grid 10	76.73	25.54	89.88	33.21
11	Grid 11	80.00	29.91	100.16	35.59
12	Grid 12	83.47	33.98	79.08	30.62
13	Grid 13	89.57	35.42	68.50	23.95
14	Grid 14	76.48	24.44	76.18	26.39
15	Grid 15	110.88	39.47	79.37	26.93
16	Grid 16	113.19	42.95	93.51	34.39
	Averagae	**90.19**	**40.55**	**88.12**	**40.49**
	Maximum	**113.19**	**73.71**	**117.52**	**73.78**
	Minimum	**76.48**	**24.44**	**68.50**	**23.95**

The following test of the designed prediction system is to compare monthly rainfall data measured by rainfall measuring instruments with predictions of monthly rainfall designed for January to June 2023. The evaluation results of monthly rainfall predictions are in the form of RMSE and MAPE values. The average values of RMSE and MAPE with the MLP method are 90.19 and 40.55. The average value of RMSE and MAPE with the LSTM method shows the values of 88.12 and 40.49. The overall RMSE and MAPE values can be seen in Table 5.

IV. CONCLUSION

Performance test of monthly rainfall data using MLP modeling on 16 data grids in South Tangerang City, resulting in an average RMSE value of 86.34 and an average MAPE value of 0.79. The highest RMSE value of 106.20 occurred on grid 13. A monthly rainfall data performance test using LSTM modeling on 16 data grids in South Tangerang City resulted in an average RMSE value of 86.34 and an average MAPE value of 0.79. The highest RMSE value is 106.20, which occurs on grid 13. Monthly rainfall prediction using the MLP method produces an RMSE value of 90.19 and a MAPE value of 40.55 while using the LSTM method produces an RMSE value of 88.12 and a MAPE value of 40.49. Monthly rainfall prediction results using the LSTM method are better than the MLP method's. This can be seen from the RMSE value of the LSTM method, which is smaller than that of MLP.

REFERENCE

[1] Herlina, N., & Prasetyorini, A. (2020). Pengaruh Perubahan Iklim pada Musim Tanam dan Produktivitas Jagung (Zea mays L.) di Kabupaten Malang (Effect of Climate Change on Planting Season and Productivity of Maize (Zea mays L.) in Malang Regency). Jurnal Ilmu Pertanian Indonesia (JIPI), Januari, 25(1), 118–128. https://doi.org/10.18343/jipi.25.1.118

[2] Sainstek, J., & Pekanbaru, S. (2021). SAINSTEK (e-Journal) Analisis Kondisi Atmosfer Saat Banjir dan Tanah Longsor (Studi Kasus : Nganjuk, 14 Februari 2021) INFORMASI ARTIKEL ABSTRACT.

[3] Funk, C.C., Peterson, P.J., Landsfeld, M.F., Pedreros, D.H., Verdin, J.P., Rowland, J.D., Romero, B.E., Husak, G.J., Michaelsen, J.C. and Verdin, A.P., 2014. A quasi-global precipitation time series for drought monitoring. US Geological Survey data series, 832(4), pp.1-12.

[4] Huang, X., Gao, L., Crosbie, R. S., Zhang, N., Fu, G., & Doble, R. (2019). Groundwater recharge prediction using linear regression, multi-layer perception network, and deep learning. Water (Switzerland), 11(9). https://doi.org/10.3390/w11091879.

[5] Wibawa AP, Utama AB, Elmunsyah H, Pujianto U, Dwiyanto FA, Hernandez L. Time-series analysis with smoothed Convolutional Neural Network. Journal of big Data. 2022 Apr 26;9(1):44.

[6] Prasetyo B, Pusparini N. Pemanfaatan SATAID Untuk Analisa Atmosfer di Wilayah Perairan. Jurnal Fisika dan Aplikasinya. 2018 Jun 1;14(2):37-44.

[7] Kafara Z, Rumlawang FY, Sinay LJ. Peramalan Curah Hujan Dengan Pendekatan Seasonal Autoregressive Integrated Moving Average (Sarima). Barekeng: Jurnal Ilmu Matematika dan Terapan. 2017 Mar 1;11(1):63-74.

[8] Bilgili M, Sahin B. Comparative analysis of regression and artificial neural network models for wind speed prediction. Meteorology and atmospheric physics. 2010 Nov;109:61-72.

[9] Lin Z, Feng J, Lu Z, Li Y, Jin D. Deepstn+: Context-aware spatial-temporal neural network for crowd flow prediction in metropolis. InProceedings of the AAAI conference on artificial intelligence 2019 Jul 17 (Vol. 33, No. 01, pp. 1020-1027).

[10] Szandala, Tomasz. (2021). Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks. In Bio-inspired Neurocomputing, 203-224. Springer, Singapore

[11] Sahu RK, Müller J, Park J, Varadharajan C, Arora B, Faybishenko B, Agarwal D. Impact of input feature selection on groundwater level prediction from a multi-layer perceptron neural network. Frontiers in Water. 2020 Nov 19;2:573034.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2149, Copyright ©2024
Submitted : April 25, 2024, Revised : June 13, 2024, Accepted : June 20, 2024, Published : June 30, 2024