Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 6 No. 2 (2022) 234 - 239 ISSN Media Electronic: 2580-0760

Prediction of Water Levels on Peatland using Deep Learning

Namora[1], Jan Everhard Riwurohi[2]
[1]Program Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur
[2]Program Studi Sistem Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur
[1]namora16@gmail.com, [2]jan.everhard@budiluhur.ac.id

Abstract

The water level on peatlands is one of the causes of peatland fires, so water levels must be maintained at a safe standard value. Government Regulation No. 71/2014 stipulates water level standard value is 0.4 meters. The forest and land fires in 2015 caused huge losses of 220 trillion Rupiah. However, fires still occur frequently. BRGM (Peatland and Mangrove Restoration Agency) installed sensors measuring peatland water levels to obtain real-time water level data. These data can be used to predict water levels. Several previous studies used drought indices, regression models, and artificial neural networks to predict water levels. In this study, it is proposed to use deep learning Long Short-Term Memory (LSTM), and apply the CRISP-DM methodology. The dataset in this study contains water level data from 15 measurement stations in Central Kalimantan from 2018 through 2021. It was concluded that the LSTM model was able to predict water level well, as indicated by the average RMSE of 0.07 m, the average R^2 of 0.85, and the average MAE of 0.04 m. The optimal LSTM model parameters are 50 epochs, a 70%:30% ratio of training data to testing data, and 2 hidden layers.

Keywords: water level, peatlands, prediction, deep learning, LSTM, CRISP-DM

1. Introduction

The water levels of peatlands hold a significant role in determining greenhouse gas emissions and holding the global climate system. Water level management in peatlands is critical to preventing peatland fires and greenhouse gas emissions [1]. Peatlands in Indonesia cover more than 7% of the country's land, therefore the use of peatlands is unavoidable. Land clearing and construction of drainage networks can damage peat, resulting in a decrease in water level, subsidence of the peat surface, CO^2 emissions, land fires, and total drought (irreversible drying).

Major forest fires in 2015 burned 2,611,411.44 hectares of forest, including peatlands [2]. Handling forest fires requires a ton of money [3]. The National Disaster Management Agency (BNPB) budget in 2019 is mainly for handling forest fires, reaching 50% of the total budget of 6.7 trillion Rupiah [4].

The standard water levels value is 0.4 meters [5], and if it is more than that, the peatlands are declared vulnerable and prone to fire. Therefore, it is necessary to monitor the water level of peatlands, namely by predicting their value to estimate the water level condition for the next period.

Many studies have been carried out to predict hydrological phenomena, including water levels. Among these are studies that use the drought index to predict the water levels of peatlands [1] and the use of Artificial Neural Network (ANN) and LSTM to construct rain runoff models [6]. Research that uses LSTM to predict rainfall has also been carried out [7], [8], [9]. Another study was conducted to predict water depth in agricultural areas using LSTM and Fast Forward Neural Network (FFNN) [10]. The use of linear regression estimation models has also been done to predict the water levels on tropical peatlands [11]. In addition, research [12] concluded the use of Autoregressive (AR) and LSTM resulted in a viable model for predicting the hydrological time series. Furthermore, the literature review [13] shows that research related to hydrology and water resources suits the use of deep learning methods.

This research complements previous research in terms of utilizing deep learning LSTM. This study uses the data on water levels in peatlands which sets it apart from previous research. The resulting model is expected to become supporting information in peatland monitoring efforts and in determining policies to reduce the potential for peatland fires.

Accepted: 05-03-2022 | Received in revised: 12-04-2022 | Published: 20-04-2022

2. Research Methods

The research methodology used in this study is the CRISP-DM (Cross Industry Standard Process for Data Mining) methodology. CRISP-DM is a standard process for data mining introduced in 1996 by a consortium of companies established as a standard process in data mining by the European Commission. CRISP-DM can be applied in various industrial sectors [14].

CRISP-DM applies a life cycle process for a data mining project consisting of six stages shown in Figure 1. The sequence is not rigid and can move back and forth between stages.

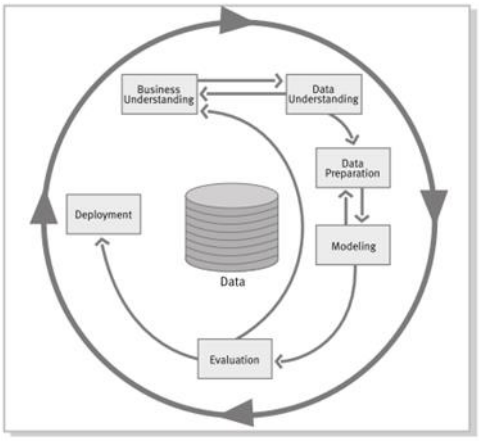

Figure 1. Stages in CRISP-DM [14]

These stages are business understanding, data understanding, data preparation, modeling, evaluation, and deployment.

Stage 1, Business Understanding is the stage of understanding the goals and needs from a business point of view, then translating this knowledge into problem definition. Afterward, plans and strategies are determined to achieve these goals. This stage builds an understanding of the water level and the problems that need to be solved. That started with an analysis of the importance of obtaining alternative information about the condition of the water levels of peatlands in the form of water levels prediction using existing historical data.

Stage 2, Data Understanding begins with data collection and continues with understanding the data, identifying data quality problems, or checking for interesting parts of the data to develop hypotheses based on the concealed information. In this case, the hypothesis is a prediction and temporary conclusion regarding the connection between variables or phenomena in peatland water levels.

Stage 3, Data Preparation includes all activities to build the final dataset (data to be processed at the modeling stage) from raw data. This stage also includes the selection of tables, records, and data attributes, including the process of cleaning and transforming data to be used as input in the modeling stage.

Stage 4, Modeling is the selection and application of modeling techniques with the parameters adjusted to get the optimal value.

Stage 5, Evaluation, evaluates the effectiveness and quality of the model and determines whether the model can achieve the goals set at Stage 1 (Business Understanding).

Stage 6, Deployment, at this stage, the gathered knowledge or information will be compiled and presented in a special form so that the users can use it. This stage often involves applying live models in the organization's decision-making process, like using real-time personalization of web pages.

In this study, the process in stage 2-6 has been simplified as shown in the flowchart in Figure 2, which is the steps taken to predict water levels on peatlands.

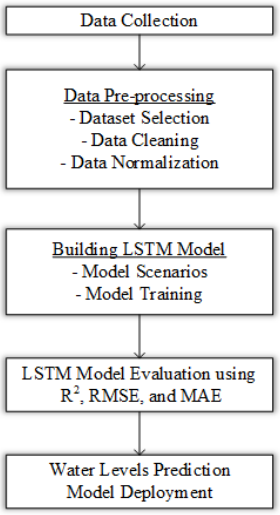

Figure 2. Water Levels Prediction Flowchart

The stages start from the collection of water level data, followed by data pre-processing so the data can be used in model development. The built model is then evaluated using the R^2, RMSE and MAE metrics to ensure that it has good performance. In the final stage, the model is deployed on a web-based application, so that it can be easier to use in predicting water levels.

2.1 Data Set

The data containing the peatland water levels was obtained from the internal data (on-premise) server of the Peat and Mangrove Restoration Agency (BRGM), managed by the Agency for the Assessment and Application of Technology (BPPT). The data is collected from the recording of sensors (water logger telemetry) installed by BPPT on peatlands. Data aggregation (average) is carried out on this data so that the previous hour period is converted into the daily

DOI: https://doi.org/10.29207/resti.v6i2.3919

period. Water level data is obtained in Comma-separated values (CSV) file format with the data period December 2018 to November 2021.

Tabel 1. Example of Water Level Data

datetime	tma
12/2/2018	-78,3
12/3/2018	-28,3
12/4/2018	-6,2
12/5/2018	0,8
…	…
11/12/2021	61,3
11/13/2021	79,1

Table 1 is an example of data in one of the files consisting of date (**datetime** column) and water level in cm (**tma** column). The negative value in the **tma** column indicates that the water is below the peatland surface.

2.2 LSTM Model

LSTM or Long Short-Term Memory is a special type of Recurrent Neural Network (RNN), which was created to avoid the problem of a long-term dependency on RNN so that LSTM can remember long-term information in the deep learning process. Figure 3 shows the iteration of the module in the RNN, which only uses one simple layer, the tanh layer [15].

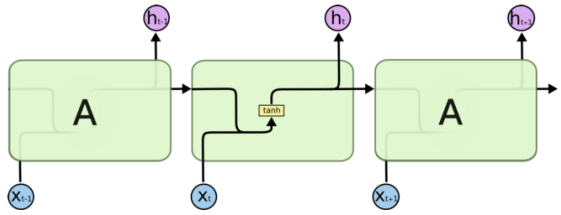

Figure 3. Repeating Module in RNN with One Layer [15]

Meanwhile, with LSTM, the module is repeated for several layers, as shown in Figure 4.

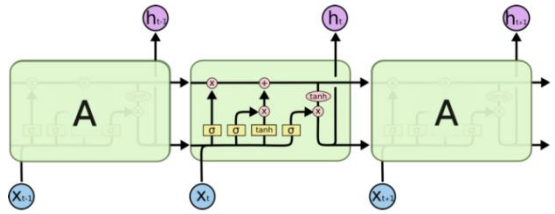

Figure 4. Repeating Module in RNN with Four Layers [15]

In this research, the Python programming language and its supporting libraries are used to build the LSTM model.

2.3. Model Evaluation and Deployment

In this study, three parameters were tested in building and training the LTSM model, namely the number of epochs; ratio of training data and testing data; and the number of hidden layers. The goal is to select the optimal parameters for the entire dataset.

The dataset for the parameter trial was chosen at random, and the Tanjung Sangalang measurement station dataset was selected from the entire dataset. Figure 5 shows the steps taken in selecting the optimal parameters.

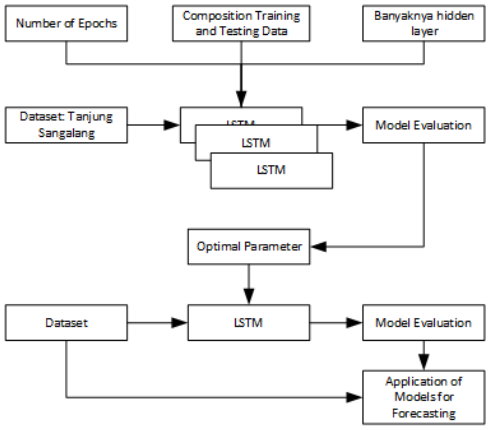

Figure 5. LSTM with Optimal Parameters

In Figure 5, three LSTM parameters are used, namely the number of epochs, the ratio of the training and testing data, and the number of hidden layers. Then, LSTM models were formed using the dataset of one of the measurement stations (Tanjung Sangalang). The optimal parameter values are determined based on the model's evaluation results.

Furthermore, the optimal parameter values are determined in the LSTM model using the entire dataset. After being evaluated and having good performance, the model is then implemented in a web-based application for forecasting purposes.

Three measurement metrics are used to evaluate the performance of the model prediction results, namely R^2 (Coefficient Determination) and RMSE (Root Mean Square Error), and MAE (Mean Absolute Error). R^2 measures the degree to which the results are replicated by the model. The value ranges between $[-\infty, 1]$ where for optimal model prediction, the R^2 score is close to 1. R^2 is expressed by equation (1).

$$R^2 = \frac{\sum_{i=1}^{N}(y_i-\bar{y})^2-\sum_{i=1}^{N}(y_i-\bar{y}_i)^2}{\sum_{i=1}^{N}(y_i-\bar{y})^2} \quad (1)$$

Where y_i is the measured value at the time i, $\bar{y}$ is the average value of y_i, at i = 1, …, N, while $\hat{y}$ is the predicted value at the time i.

RMSE is used to measure the average value of the error value of the model prediction results. The RMSE formula is shown in equation 2, where the measured value at the time i is expressed by y_i , while $\hat{y}_i$ is the predicted value at the time i.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \quad (2)$$

DOI: https://doi.org/10.29207/resti.v6i2.3919

Whereas MAE measures the difference between the observed and modeled results. MAE is the mean of absolute error as shown in equation (3), where the measured value at the time i is expressed by y_i, while $\hat{y}_i$ is the predicted value at the time i.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (3)$$

In this study, the model is deployed in the operational environment in the form of a web application. In addition to the model itself, an interface, library, and supporting infrastructure are also prepared so that users can predict the water levels generated from the model.

3. Results and Discussions

3.1 Data Pre-Processing

Data cleaning is done to clean the dataset from unexpected data, such as outliers. Figure 6 indicates an example of an outlier, which is a water level data in 2000, while the data period should begin from 2018.

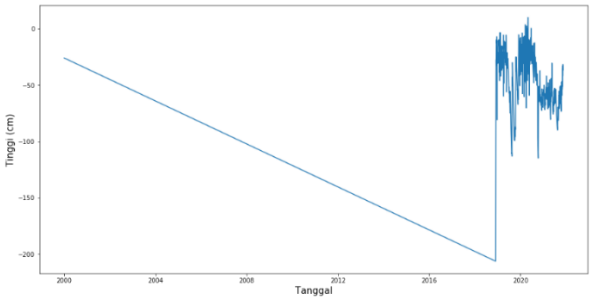

Figure 6. Example of an Outlier

The data is cleaned, or removed from the dataset. Figure 7 is the dataset plot after cleaning.

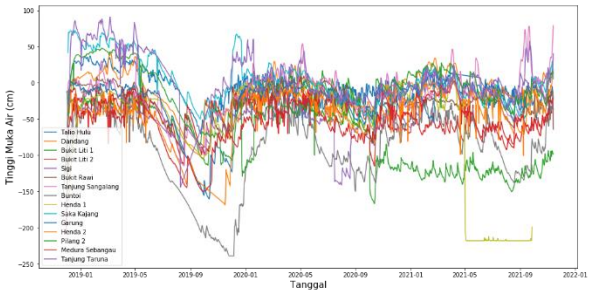

Figure 7. Complete Dataset Plot

To support data understanding, the description of the dataset statistics used is served in Table 2.

The next process is to normalize or rescale the data from its original range so that all values are in the range of 0 and 1. The method used is Minmax as the equation (4) so that water level data that is too high or too low does not affect the modeling process.

$$y = \frac{x - min}{max - min} \quad (4)$$

In equation (4), y is the normalized value, x is the original value (input), while min and max are the minimum and maximum values of the entire data, respectively.

Table 2. Statistic Descriptive of Water Levels

Stasiun	Average (cm)	Standard deviation (cm)	Min (cm)	Max (cm)
Talio Hulu	-8,5	22,0	-70,5	37,2
Dandang	-13,1	26,8	-113,2	40,7
Bukit Liti 1	-65,4	49,2	-166,7	21,5
Bukit Liti 2	-44,7	25,9	-206,4	10,1
Sigi	-41,4	32,7	-168,7	19,4
Bukit Rawi	-27,0	24,6	-151,9	16,0
Tanjung Sangalang	-18,1	31,0	-114,7	79,1
Buntoi	-90,9	54,6	-239,0	-8,1
Henda 1	-56,9	70,6	-218,0	11,3
Saka Kajang	1,6	30,9	-59,9	72,5
Garung	-27,4	36,1	-160,4	21,5
Henda 2	-39,7	29,9	-168,7	15,6
Pilang 2	-20,3	40,2	-157,4	47,8
Medura Sebangau	-54,9	32,4	-150,1	2,3
Tanjung Taruna	7,8	31,2	-144,2	90,7

3.2 Modeling and Analysis

The LSTM model is applied to the Tanjung Sangalang dataset with the initial parameter, which is 50 epoch, training and testing data split into a 70:30 ratio, and 2 hidden layers. The epoch parameter has been tested in the LSTM model to determine the optimal quantity of epoch, with the amount of 1, 10, 30, 50, 100, and 200 epochs.

The model is trained and validated using several combinations of dataset ratios, namely (90:10), (80:20), (70:30), and (60:40) each for data training and data testing.

Table 3 is the result of RMSE, R^2, and MAE from the application of the model parameters. It shows that with various variations in the ratio of data on training and testing, the best prediction results can be obtained by using 50 epochs. This is indicated by the small value of RMSE (0.05 and 0.06 meters), and the R^2 value is close to 1, and MAE is no more than 0.04.

Table 3. RMSE, R^2 , dan MAE with various numbers of epochs

Rasio	Epoch:	1	10	30	50	100	200
90:10	RMSE	0,25	0,10	0,09	0,07	0,06	0,08
	R^2	0,56	0,93	0,94	0,97	0,98	0,96
	MAE	0,18	0,06	0,06	0,04	0,04	0,05
80:20	RMSE	0,21	0,09	0,06	0,06	0,06	0,06
	R^2	0,50	0,91	0,96	0,96	0,96	0,96
	MAE	0,15	0,07	0,05	0,03	0,04	0,03
70:30	RMSE	0,18	0,10	0,06	0,05	0,05	0,05
	R^2	0,53	0,86	0,95	0,97	0,96	0,96
	MAE	0,12	0,06	0,05	0,04	0,04	0,04
60:40	RMSE	0,18	0,10	0,06	0,05	0,05	0,05
	R^2	0,57	0,86	0,96	0,97	0,96	0,96
	MAE	0,12	0,07	0,03	0,03	0,03	0,04

Figure 8 and 9 are examples of actual data plots and water level predictions at Tanjung Sangalang measurement stations with 30, and 50 epochs. The blue line shows the actual data (or prediction target), while the red line shows the prediction results.

DOI: https://doi.org/10.29207/resti.v6i2.3919

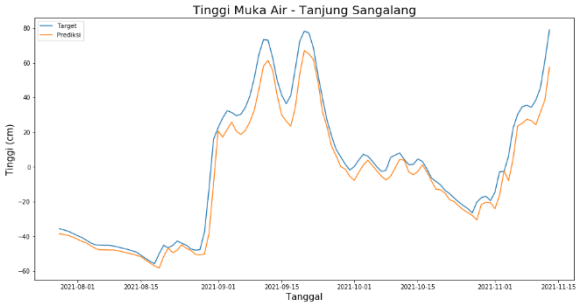

Figure 8. Plot Prediction with 30 Epoch

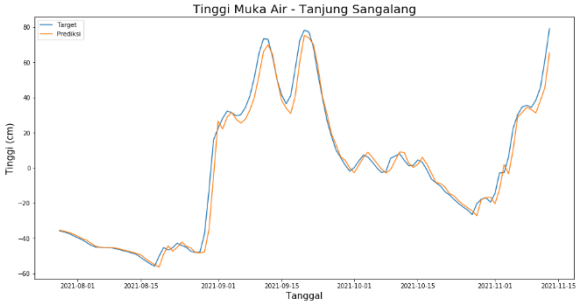

Figure 9. Plot Prediction with 50 Epoch

Once the epoch value is set to 50, more trials are applied to various combinations of ratios of testing data and training data. The goal is to obtain an optimal ratio of testing data and training data. The LSTM program code was run four times for each data ratio, and the RMSE, R^2, and MAE values were calculated, as shown in Table 4.

Table 4 shows that the composition of training data and testing data that provides a satisfactory average predictive result with 50 epochs is the ratio (70:30) and (60:40). Both generated RMSE, R^2, and MAE values of 0.05 m; 0.96; and 0.03 m respectively. In this study, the ratio (70:30) was chosen as the best parameter, to use more testing data than the ratio (60:40).

Table 4. RMSE, R^2, and MAE with Various of Data Ratio

Ratio	Trial:	0	1	2	3	4	Average
90:10	RMSE	0,07	0,06	0,06	0,05	0,06	0,06
	R^2	0,97	0,98	0,97	0,98	0,98	0,98
	MAE	0,04	0,04	0,04	0,04	0,04	0,04
80:20	RMSE	0,06	0,05	0,06	0,08	0,05	0,06
	R^2	0,96	0,97	0,96	0,93	0,97	0,96
	MAE	0,03	0,04	0,05	0,06	0,04	0,04
70:30	RMSE	0,05	0,05	0,06	0,05	0,06	0,05
	R^2	0,97	0,97	0,96	0,97	0,96	0,96
	MAE	0,04	0,03	0,03	0,03	0,04	0,03
60:40	RMSE	0,05	0,05	0,05	0,05	0,05	0,05
	R^2	0,97	0,96	0,97	0,96	0,96	0,96
	MAE	0,03	0,04	0,03	0,03	0,03	0,03

The LSTM model has been tested with a variety of hidden layers to find the optimal number of hidden layers while maintaining good model performance. Table 5 shows the results of the RMSE, R^2, and MAE metrics from these tests.

Tabel 5. RMSE, R^2 , and MAE with Various Hidden Layer

	Training Data : Testing Data Ratio (70 : 30)		
Hidden Layer	RMSE	R^2	MAE
2	0,05	0,97	0,04
3	0,05	0,96	0,04
4	0,05	0,96	0,04
5	0,05	0,95	0,04

Table 5 shows that the addition of a hidden layer to the LSTM model in this study did not provide a significant increase in performance. In this case the model with 2 hidden layers provides better performance than the 3, 4, or 5 hidden layers.

3.3 Evaluation and Deployment

Based on the analysis of the results of the LSTM parameter tests, it is concluded that the optimal LSTM parameters in this study are as shown in Table 6.

Table 6. Optimal Parameter

Parameter	Value
Number of epoch	50
Ratio of training data and testing data	70% : 30%
Number of hidden layer	2

The model is evaluated by applying the optimal parameters from Table 6 to the entire dataset, which includes water level data from 15 measurement stations. After completing the training process, each model's average value of RMSE, R^2, and MAE were calculated, and the results were presented in Table 7.

Table 7. Summary of RMSE, R^2, dan MAE

No	Station	RMSE	R^2	MAE
1	Talio Hulu	0,09	0,62	0,07
2	Dandang	0,10	0,68	0,05
3	Bukit Litih 1	0,05	0,74	0,05
4	Bukit Litih 2	0,05	0,75	0,04
5	Sigi	0,06	0,91	0,04
6	Bukit Rawi	0,04	0,83	0,02
7	Tanjung Sangalang	0,05	0,96	0,04
8	Buntoi	0,07	0,96	0,04
9	Henda 1	0,13	0,98	0,06
10	Saka Kajang	0,03	0,97	0,02
11	Garung	0,06	0,87	0,03
12	Henda 2	0,09	0,80	0,05
13	Pilang 2	0,07	0,89	0,03
14	Medura Sebangau	0,06	0,95	0,04
15	Tanjung Taruna	0,05	0,91	0,04
	Rata-rata:	0,07	0,85	0,04

The model is implemented as a web-based application so that it can be used live and in real-time. Table 8 lists the specifications and supporting data for the web applications that were successfully implemented.

Figure 10 shows a web application that has been created as a result of model implementation.

The user can predict the water level for the next 60-day period (after the last date in the dataset) by selecting the name of the measurement station, and pressing the ‘Tampilkan’ button shown in Figure 10.

DOI: https://doi.org/10.29207/resti.v6i2.3919

Table 8. Web Deployment Specifications

Specification	Value
Server	Heroku.com
Web address (URL)	https://prediksi-tma.herokuapp.com/
Programming Language / Framework	Python / Flask, Javascript, HTML
Tools	Anaconda, pip, Heroku CLI, git
Python runtime (version)	3.7.12
Requirements	Flask==2.0.2, gunicorn==20.1.0 joblib==1.1.0, numpy==1.21.5 pandas==1.3.5, pickleshare==0.7.5 plotly==5.5.0, requests==2.26.0 scikit-learn==0.21.3, tensorflow-cpu==2.7.0

Figure 10. Web Application Display

4. Conclusion

From a series of model experiments and model training, the optimal modeling parameters for LSTM in this study were 50 epochs, 70% training data and 30% testing data ratios, and 2 hidden layers. The LSTM model is proven to be able to predict the water level on peatlands. This is indicated by the average value of the RMSE, R^2, and MAE metrics of 0.07, 0.85, and 0.04, respectively. This means that on average, the difference between the predicted results and the actual water level is 0.07 m or 7 cm, while the average error is 0.04 m, with a prediction accuracy of 85%.

These results support the results of previous studies in the use of LSTM, for example the RMSE value in study [7] was around 0.11 to 0.12, while in this study it was 0.07. While the value of R^2 in this study is in accordance with the results of research [9], which is above 0.8. This indicates that the LSTM model generated from this study is suitable for predicting water level on peatlands.

For future research, it is recommended to build an LSTM model by optimizing other LSTM parameters and using datasets from other regions or provinces.

Acknowledgment

The author would like to thank various parties who supported this research, especially BRGM and BPPT who assisted in providing the research data.

Reference

[1] N. Febrianti, K. Murtilaksono, and B. Barus, "Analisis Model Estimasi Tinggi Muka Air Tanah Menggunakan Indek Kekeringan," *J. Penginderaan Jauh dan Pengolah. Data Citra Digit.*, vol. 15, no. 1, pp. 25–36, 2018, doi: 10.30536/j.pjpdcd.2018.v15.a2867.

[2] BNPB, "Rekapitulasi Luas Kebakaran Hutan dan Lahan (Ha) Per Provinsi Di Indonesia Tahun 2014-2019," *Karhutla Monitoring Sistem*, 2019. http://sipongi.menlhk.go.id/hotspot/luas_kebakaran (accessed Aug. 01, 2021).

[3] WorldBank, "Krisis Kebakaran dan Asap Indonesia," *Worldbank.Org*, 2015. https://www.worldbank.org/en/news/feature/2015/12/01/indonesias-fire-and-haze-crisis (accessed Jul. 01, 2021).

[4] M. Bernie, "BNPB Habiskan Rp6 , 7 Triliun untuk Tangani Bencana Selama 2019," 2019. https://tirto.id/bnpb-habiskan-rp67-triliun-untuk-tangani-bencana-selama-2019-epRS (accessed Jul. 01, 2021).

[5] PP-71, "Peraturan Pemerintah Republik Indonesia Nomor 71 Tahun 2014 Tentang Perlindungan Dan Pengelolaan Ekosistem Gambut," p. 38, 2014.

[6] S. Poornima and M. Pushpalatha, "Prediction of rainfall using intensified LSTM based recurrent Neural Network with Weighted Linear Units," *Atmosphere (Basel).*, vol. 10, no. 11, 2019, doi: 10.3390/atmos10110668.

[7] M. Rizki, S. Basuki, and Y. Azhar, "Implementasi Deep Learning Menggunakan Arsitektur Long Short Term Memory(LSTM) Untuk Prediksi Curah Hujan Kota Malang," *J. Repos.*, vol. 2, no. 3, p. 331, 2020, doi: 10.22219/repositor.v2i3.470.

[8] C. J. Zhang, J. Zeng, H. Y. Wang, L. M. Ma, and H. Chu, "Correction model for rainfall forecasts using the LSTM with multiple meteorological factors," *Meteorol. Appl.*, vol. 27, no. 1, pp. 1–15, 2020, doi: 10.1002/met.1852.

[9] C. Hu, Q. Wu, H. Li, S. Jian, N. Li, and Z. Lou, "Deep learning with a long short-term memory networks approach for rainfall-runoff simulation," *Water (Switzerland)*, vol. 10, no. 11, pp. 1–16, 2018, doi: 10.3390/w10111543.

[10] J. Zhang, Y. Zhu, X. Zhang, M. Ye, and J. Yang, "Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas," *J. Hydrol.*, vol. 561, no. April, pp. 918–929, 2018, doi: 10.1016/j.jhydrol.2018.04.065.

[11] N. E. Putra, S. Sutikno, and M. Fauzi, "Model Prediksi Kedalaman Muka Air Tanah pada Lahan Gambut Tropis," *Apl. Teknol.*, vol. 11, no. 2, 2019.

[12] J. Qin, J. Liang, T. Chen, X. Lei, and A. Kang, "Simulating and predicting of hydrological time series based on tensorflow deep learning," *Polish J. Environ. Stud.*, vol. 28, no. 2, pp. 795–802, 2019, doi: 10.15244/pjoes/81557.

[13] M. Sit, B. Z. Demiray, Z. Xiang, G. J. Ewing, Y. Sermet, and I. Demir, "A comprehensive review of deep learning applications in hydrology and water resources," *Water Sci. Technol.*, 2020, doi: 10.2166/wst.2020.369.

[14] C. Pete *et al.*, "Crisp-Dm 1.0, Step-by-step data mining guide," *Cris. Consort.*, p. 76, 2000.

[15] C. Olah, "Understanding LSTM Networks [Blog]," *Web Page*, 2015. http://colah.github.io/posts/2015-08-Understanding-LSTMs/ (accessed Jul. 01, 2021).

DOI: https://doi.org/10.29207/resti.v6i2.3919