JURNAL RISET INFORMATIKA **P-ISSN: 2656-1743 |E-ISSN: 2656-1735**
Vol. 5, No. 3. June 2023 DOI: https://doi.org/10.34288/jri.v5i3.558
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

PREDICTION OF RAINFALL AND WATER DISCHARGE IN THE JAGIR RIVER SURABAYA WITH LONG-SHORT-TERM MEMORY (LSTM)

Retzi Yosia Lewu-[1], Slamet-[2], Sri Wulandari-[3], Widdi Djatmiko-[4], Kusrini-[5], Mulia Sulistiyono-[6*)]

Magister of Informatics Engineering
Universitas Amikom Yogyakarta
Yogyakarta, Indonesia
[1]retzi.lewu@students.amikom.ac.id, [2]slametmieno@students.amikom.ac.id,
[3]sriwulandari@students.amikom.ac.id, [4]widdi.dj@students.amikom.ac.id,
[5]kusrini@amikom.ac.id, [6*)] muliasulistiyono@amikom.ac.id

(*) Corresponding Author

Abstract

Floods can occur at any time if the amount of river water discharge and rainfall intensity tends to be high, so preparations and ways of handling are needed to anticipate flooding quickly, precisely, and accurately for the Surabaya City Public Works Service. One of the steps to predict and analyze the status of the flood disaster alert level is to calculate predictions based on rainfall and the amount of river water discharge. This study uses the Long-Short Term Memory (LSTM) algorithm to predict using a time series dataset of rainfall and river water discharge in the Jagir River, Surabaya. This data is used to make predictions with the proportion of 70% training data and 30% testing data. Data normalization is performed in intervals of 0 and 1 using a min-max scaler and activated using ReLU (Rectified Linear Unit) and Adam Optimizer. The process continues by repeating the process to enter iterations, or epochs, until it reaches the specified epoch (n). The data is then normalized to their original values and visualized. The model was evaluated and produced acceptable performance evaluation results for the rainfall variable, namely at epoch (n) = 75 for training data, namely a score of 0.054 for MAE and 0.099 for RMSE. In contrast, data testing was given a score of 0.041 for MAE and 0.091 for RMSE. As for the water discharge variable, the performance evaluation shows the difference between the training and testing data. Results of training data MAE = 11.10 and RMSE=18RMSE =18.61.61 at epoch (n) = 150. Results of data testing MAE = 11.37 and RMSE = 21.08 at epoch (n) = 100. These results indicate an anomaly that needs to be discussed in further research.

Keywords: Rainfall; Water Discharge; Prediction; Flood; Long Short Term Memory (LSTM)

Abstrak

Banjir dapat terjadi sewaktu-waktu apabila faktor jumlah debit air sungai dan intensitas curah hujan cenderung tinggi, sehingga diperlukan persiapan dan cara penanganan untuk mengantisipasi banjir secara cepat, tepat, dan akurat bagi Dinas Pekerjaan Umum Kota Surabaya. Salah satu langkah untuk memprediksi dan menganalisis status tingkat siaga bencana banjir adalah dengan menghitung prediksi berdasarkan curah hujan dan jumlah debit air sungai. Penelitian ini menggunakan algoritma Long-Short Term Memory (LSTM) untuk memprediksi dengan menggunakan dataset time series curah hujan dan debit air sungai di Sungai Jagir Surabaya. Data ini digunakan untuk membuat prediksi dengan proporsi 70% data training dan 30% data testing. Normalisasi data dilakukan dalam interval 0 dan 1 menggunakan `minmax scaler` *dan diaktifkan menggunakan ReLU (Rectified Linear Unit) dan Adam Optimizer. Proses dilanjutkan dengan mengulang proses untuk memasukkan iterasi, atau epoch, hingga mencapai epoch (n) yang ditentukan. Data kemudian didenormalisasi ke nilai aslinya dan divisualisasikan. Model dievaluasi dan menghasilkan nilai hasil evaluasi kinerja yang dapat diterima untuk variabel curah hujan yaitu pada epoch (n) = 75 untuk data training yaitu skor 0,054 untuk MAE dan skor 0,099 untuk RMSE, seta data testing diberi skor 0,041 untuk MAE dan 0,091 untuk RMSE. Sedangkan untuk variabel debit air, evaluasi kinerja menunjukkan perbedaan antara data training dan data testing. Hasil data training MAE = 11.10 dan RMSE = 18.61 pada epoch (n) = 150. Hasil data testing MAE = 11.37 dan RMSE = 21.08 pada epoch (n) = 100. Hasil ini menunjukkan adanya anomali sehingga perlu dibahas pada penelitian selanjutnya.*

Kata kunci: Curah Hujan; Debit Air; Prediksi; Banjir; Long Short Term Memory (LSTM)

INTRODUCTION

As an archipelagic country close to the equator, Indonesia has an excellent opportunity to experience flooding. The monitoring results of the National Disaster Management Agency (BNPB) stated that since 2018 floods have become a disaster with the most significant impact, according to the available data (https://bnpb.go.id/infographics). The flood disaster occurred evenly in Indonesia, including in Surabaya.

There are several large rivers in Surabaya, one of them is the Jagir River, which needs to be examined for its flood alert status, considering that the river is an artificial river located in a densely populated area. Several factors, including rainfall and water discharge, can cause floods. These two factors can be used to determine flood alert status.
In hydrology, it is explained that river water discharge is a measure of the amount of water flowing out of a watershed (DAS) in volume units per second. The river water discharge unit is cubic meters per second (m3/second) (Asdak, 2023). Every river in Surabaya has an essential role in accommodating and storing water which will then flow into the major rivers in Surabaya and empty into the sea. An excessive river water discharge will result in a flood disaster that can damage or cause property loss and even claim lives.

Flood disasters can indeed occur when there is instability in the river's flow, and it comes relatively quickly. So preparation and handling methods are also needed to quickly, precisely, and accurately anticipate floods for the Dinas Pekerjaan Umum Pengairan Provinsi Jawa Timur UPT Pengelolaan Sumber Daya Air Surabaya. One of the steps to anticipate a flood disaster is calculating the predicted amount of river water discharge. The term prediction is similar to classification and estimation, in which prediction results lie in the future (Larose, 2005). Predictions can be made using several algorithms, including machine learning, Artificial Neural Networks (ANN), and LSTM (Long et al.). The LSTM algorithm was first introduced in 1997 by Hochreiter and Schmidhuber (Hochreiter & Schmidhuber, 1997). LSTM consists of several layers that can be repeated and has several basic variable calculation processes, including addition, multiplication, and other mathematical functions. So in this study, the prediction will be held by using the existing periodic time series data of the amount of river water discharge in recent years, and a predictive result of the river water discharge will be obtained for some time to come using LSTM as a method. Therefore, it will explain the use of LSTM for predicting rainfall and water discharge by analyzing data obtained from the past to obtain projections of future data.

Furthermore, to determine the performance of the LSTM algorithm model, a testing process will be carried out using MAE (Mean Absolute Error) (Bouktif, Fiaz, Ouni, & Serhani, 2018) and Mean Squared Error (MSE) (Shetty, Padmashree, Sagar, & Cauvery, 2021), in this case, RMSE (Root Mean Squared Error) (Elizabeth Michael, Mishra, Hasan, & Al-Durra, 2022; Kouadri, Pande, Panneerselvam, Moharir, & Elbeltagi., 2022), to test the prediction results on actual data. MAE is the absolute change between the original and prediction values (Wang & Lu, 2018) and the average for all the values. In contrast, It is explained as the square root of MSE (Mean Square Error), which is the square of change between the original and prediction values and the average for all the values (Navlan, Fandango, & Idris, 2021). Using the LSTM algorithm model, this study is expected to produce an acceptable score (near zero) for both MAE and RMSE. It is an understandable reason so that it can provide knowledge to increase information for UPT Pengelolaan Sumber Daya Air Surabaya in anticipating/managing floods in the Surabaya area, especially those caused by the Jagir River.

RESEARCH METHODS

Types of research

This study uses a quantitative approach. Using the method of literature study and observation is as follows:

Literature Study

Much research has been conducted on flood prediction using LSTM (Long Short Term Memory) and other methods.

Literature Study related to LSTM:

Rizki et al., in 2022, researched Rainfall Prediction for the City of Malang and found that the application successfully processed rainfall predictions for Malang with rainfall parameters (Rizki, Basuki, & Azhar, 2020). The number of hidden layer neurons with the most optimal results is 256 hidden layer neurons. This is because the 256 hidden layer neurons have the lowest error rate, 12,247 on the train data and 11,481 on the test data. The number of epochs with the most optimal results is 150 epochs. This is because the number of 150

JURNAL RISET INFORMATIKA **P-ISSN: 2656-1743 |E-ISSN: 2656-1735**
Vol. 5, No. 3. June 2023 DOI: https://doi.org/10.34288/jri.v5i3.558
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

epochs has the lowest error rate, namely on the train data of 12,079 and the test data of 11,288. The composition of Data Train and Data Test with the most optimal results is the composition of 50% train data and 50% test data. This is because the composition of 50% train data and 50% test data has the lowest error rate; namely, the train data is 12,079, and the test data is 11,288. This research is considered not too significant because it only uses one variable, namely rainfall;

Devi et al., in 2022, conducted a Dasarian Rainfall Prediction using the Vanilla RNN and LSTM Methods to Determine the Beginning of the Rainy and Dry Seasons. They obtained the best features: humidity, pressure, and visibility (Devi, Bayupati, & Wirdiani, 2022). Models with features that have been selected using the Backward Elimination method obtain more optimal performance compared to models that use all data features. Each model using the Vanilla RNN and LSTM methods obtained poor results at a learning rate of 0.0001. This study's learning rate of 0.0001 requires a more significant epoch to obtain optimal results. The best model is obtained by the vanilla RNN method with feature selection. The RMSE obtained was 28.4308, and R2 was 0.6139. The R2 value of 0.6139 is included in the strong category, where this model is suitable for predicting primary rainfall data. The information obtained from the results of the 2021 rainfall prediction is that June will enter the dry season in June, and 1 December will enter the rainy season.

Kardhana et al. 2022 improved the flood prediction method using the LSTM-RNN and Sadewa satellite data (Kardhana, Valerian, Rohmat, & Kusuma, 2022). The LSTM-RNN is used to predict the water level (Sudriani, Ridwansyah, & A Rustini, 2019) in the Katulampa Dam using Sadewa satellite data. The results show that the model can accurately predict the Katulampa Water Level and provides a potential for implementing and improving lead time for flood mitigation. Using the LSTM-RNN, the model can accurately predict the water level in Katulampa with repeated data t – 24 hours, with R2 above 0.82. The model can maintain R2 above 0.80 for the next 24 hours in the prediction.

Literature Study related flood prediction using other methods:

Supatmi et al. 2019 proposed a hybrid approach based on a neural network and a fuzzy inference system for flood vulnerability, namely the hybrid neuro-fuzzy inference system (HN-FIS). HN-FIS is a model that can automatically learn and obtain output that can present the essence of fuzzy logic 2 Computational Intelligence and Neuroscience (Supatmi, Hou, & Sumitra, 2019). The system is implemented in 31 districts in the city of Bandung. Flood prediction relies on several variable inputs: population density, area elevation, and rainfall in a time series from 2008 to 2012. The main contribution of this paper is to provide a hybrid prediction for flood susceptibility based on neural networks and a fuzzy inference system for accurate flood prediction. It used data variables that utilized the Bandung database for flood hazard prediction and developed a practical hybrid prediction approach for flood susceptibility with higher accuracy.

Noymanee & Theeramunkong conducted research in which machine learning techniques were developed to predict errors in rainfall simulations. A hybrid model based on MIKE11 and machine learning techniques will provide better predictive results than only one MIKE11 model (Noymanee & Theeramunkong, 2019).

Using the Variant Inflation Factor, Sampurno et al. conducted a statistical analysis to analyze the multicollinearity between the predictor variables (Sampurno, Vallaeys, Ardianto, & Hanert, 2022). The researcher tested four kernels, namely linear, polynomial, radial basis, and sigmoid, and found that the radial kernel had the best performance in the SVM algorithm.

B. Observation

Observations were made on the data available at the Dinas Pekerjaan Umum Pengairan Provinsi Jawa Timur UPT Pengelolaan Sumber Daya Air Surabaya. The Observation Results in the form of a dataset are then processed in the following manner:

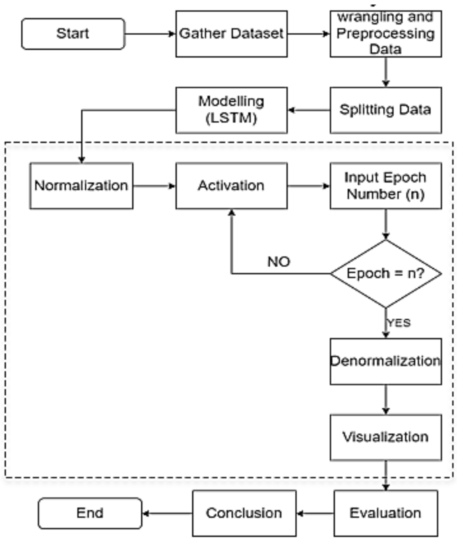

Figure 1. Research Flowchart

Time and Place of Research

The research was conducted from 29 May 2023 to 10 July 2023 with the details as shown in Figure 2 below:

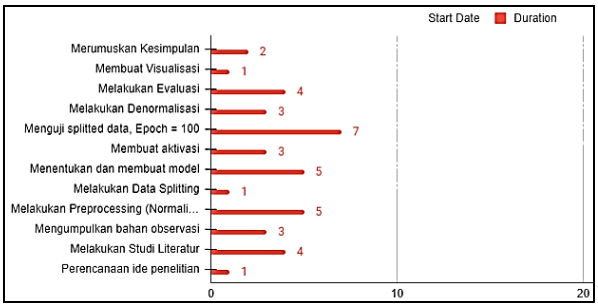

Figure 2. Research Schedule

Research took place at Dinas Pekerjaan Umum Pengairan Provinsi Jawa Timur UPT Pengelolaan Sumber Daya Air Surabaya.

Research target/Subject

The Subject is Dinas Pekerjaan Umum Pengairan Provinsi Jawa Timur UPT Pengelolaan Sumber Daya Air Surabaya, where we derive the population of data. The data population is a dataset of rainfall and water discharge, and the data samples are those captured from 2020 to 2022, with as many as 1096 rows. The data sample uses rainfall and water discharge as they are being used as the variables within the research.

Data, Instruments, and Data Collection Techniques

The data used in this study is a dataset from the Irrigation Public Works Office of East Java Province UPT Water Resources Management Surabaya captured the raw data using provided devices :

- Rainfall data is recorded based on the output of a device called the Automatic Rainfall Recorder (ARR) through the Wonokromo Station, and
- Water discharge data is recorded based on the output of a device called AWLR (Automatic Water Level Recorder) through the Jagir River floodgates in Surabaya.

These data will be used for future prediction calculations using the LSTM method, focusing on the following rainfall and water discharge as research variables.

Data Analysis Technique

The dataset is analyzed using some steps, as shown in Figure 1. They are:

1. Wrangling and Preprocessing, in which the attributes are checked whether each variable column has the potential to have anomalous attributes or columns with the potential to have no value (null).
2. Splitting the data into training and testing data with a composition of 70:30.
3. LSTM Modelling.

This is the primary process of the study. Python is being used to model the prediction. Each variable is analyzed using LSTM by processing into several layers during some iterations (named epoch) through these actions:

a) Normalization. Scaling is applied for the data in a specific interval of 0 and 1. So it is said that the value on the dataset is normalized into ≤1 using the min-max scaler.
b) Activation. This study uses ReLU (Rectified Linear Unit) to activate the output. The output of the activation function is expressed as 0 (zero) if the input is negative. However, if the input is positive, the output will equal the input value of the activation function (Szandala, 2021). Adam Optimizer is also used to iteratively update the weighted network based on training data in this step.
c) Input epoch. This is the part to input how many iterations through the codes. Epoch is defined from a certain number of iterations (n) during several basic variable calculation processes, including addition, multiplication, and other mathematical functions regards to LSTM until it is completed (reach the defined epoch).
d) Denormalization is when a scaler puts the result back into a normal form. inverse
e) Data Visualization is visualized into a plot diagram for each variable.

4. Evaluation is the next step, where the model is evaluated using some formula to measure the performance of each result. In this study, MAE (Mean Absolute Error) and RMSE (Root Mean Square Error) are used to show the performance of each variable.
5. The last step is to conclude the result and derive recommendations and suggestions for future works.

RESULTS AND DISCUSSION

The dataset which is collected from the Dinas Pekerjaan Umum Pengairan Provinsi Jawa Timur UPT Pengelolaan Sumber Daya Air Surabaya will be used for future prediction calculations using the LSTM method, focusing on the following data variables:

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v5i3.558
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

1. Rainfall data is recorded based on the Automatic Rainfall Recorder (ARR) output through the Wonokromo Station. There are guidelines for determining average level status, namely a rainfall value of less than 100 mm. In contrast, the alert status will apply if the rainfall value exceeds 100 mm. The data used is from January 2020 to December 2022. The sample rainfall dataset is listed in Table 1.

Table 1. Rainfall datasets.

Date	Rainfall
01/01/2020	85
02/01/2020	0
03/01/2020	0
...	...
30/12/2022	1
31/12/2022	1.4

2. Water discharge data is recorded based on the output of a device called AWLR (Automatic Water Level Recorder) through the Jagir River floodgates in Surabaya. The data used is from January 2020 to December 2022. There are guidelines for determining the status of the green level if the water debit value is more than or equal to 180 m3/second and the yellow level if the water debit value is more than or equal to 200 m3/second. The level was red if the water debit value was more than or equal to 220m3/second. The data used is from January 2020 to December 2022. The sample rainfall dataset is listed in Table 2.

Table 2. Water Discharge Dataset

Date	Water Discharge
01/01/2020	44.04
02/01/2020	26.18
03/01/2020	19.63
...	...
30/12/2022	117.8
31/12/2022	125.4

Preprocessing

The number of datasets collected from January 2020 to December 2022 is 1,096 data consisting of date, rainfall, and water discharge variables. The data will then go through an analysis process before making predictions by selecting data and checking the attributes of each variable column with the potential to have anomalous attributes and columns with the potential to have no value (null). To be useful for data mining, the databases must undergo preprocessing in the form of *data cleaning* and *data transformation*(Larose, 2005).

Data Splitting

Preprocessing will then be divided into two parts, with a ratio of 70% as training data and 30% as testing data. This split data process aims to train past data to predict future data. Based on the data sharing ratio above, out of 1074 data, 756 training data were obtained and 318 testing data. The data-sharing process in Python can be seen more clearly in Figure 3.

```
[20] seq_size = 10
     trainX, trainY = to_sequences(train, seq_size)
     testX, testY = to_sequences(test, seq_size)

[21] print("Shape of training set: {}".format(trainX.shape))
     print("Shape of test set: {}".format(testX.shape))

     Shape of training set: (756, 10)
     Shape of test set: (318, 10)
```

Figure 3. Splitting Data

LSTM Modelling

The Modelling process steps are:

1) **Normalization.** Scaling is applied for the data in a certain interval of 0 and 1. So it is said that the value on the dataset is normalized into ≤1 using the mimmaxScaler function as Figure 4, and 5 follows.

```
scaler = MinMaxScaler(feature_range=(0, 1))
dataset = scaler.fit_transform(dataset)
print(dataset)

[[0.7391304 ]
 [0.        ]
 [0.        ]
 ...
 [0.16956522]
 [0.00869565]
 [0.01217391]]
```

Figure 4. Normalization for rainfall variable

```
scaler = MinMaxScaler(feature_range=(0, 1))
dataset = scaler.fit_transform(dataset)
print(dataset)

[[0.13210684]
 [0.06463678]
 [0.03989271]
 ...
 [0.1512221 ]
 [0.4107514 ]
 [0.43946207]]
```

Figure 5. Normalization for water discharge variable

2) **Activation.** This study uses ReLU (Rectified Linear Unit) to activate the output. The output of

P-ISSN: 2656-1743 | E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v5i3.558

JURNAL RISET INFORMATIKA
Vol. 5, No. 3. June 2023

Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

the activation function is expressed as 0 (zero) if the input is negative. However, if the input is positive, the output will equal the input value of the activation function (Szandała, 2020). The activation process for each variable is shown by the codes below:

```
model = Sequential()
model.add(ConvLSTM2D(filters=64, kernel_size=(1,1), activation='relu', input_shape=(1, 1, 1,
seq_size)))
model.add(Flatten())
model.add(Dense(32))
model.add(Dense(1))
model.compile(optimizer='adam', loss='mean_squared_error')
model.summary()
```

Note that the Adam optimizer is also used to optimize, to update the weighted network based on training data iteratively. The codes yield:

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv_lstm2d (ConvLSTM2D)    (None, 1, 1, 64)          19200

 flatten (Flatten)           (None, 64)                0

 dense (Dense)               (None, 32)                2080

 dense_1 (Dense)             (None, 1)                 33

=================================================================
Total params: 21,313
Trainable params: 21,313
Non-trainable params: 0
```

Figure 6. Activation Result

Input epoch. Epoch is defined from a certain number of iterations (n) during several basic variable calculation processes, including addition, multiplication, and other mathematical functions regards to LSTM until it is completed (reach the defined epoch). Several epochs, namely 10, 50, 75, 100, and 150, were run in this study. The variations will also occur for each n (50, 75, 100 and 150) provided as input.

```
Epoch 1/10
24/24 - 0s - loss: 0.0076 - val_loss: 0.0092 - 156ms/epoch - 6ms/step
Epoch 2/10
24/24 - 0s - loss: 0.0073 - val_loss: 0.0092 - 116ms/epoch - 5ms/step
Epoch 3/10
24/24 - 0s - loss: 0.0075 - val_loss: 0.0092 - 110ms/epoch - 5ms/step
Epoch 4/10
24/24 - 0s - loss: 0.0075 - val_loss: 0.0095 - 120ms/epoch - 5ms/step
Epoch 5/10
24/24 - 0s - loss: 0.0076 - val_loss: 0.0099 - 116ms/epoch - 5ms/step
Epoch 6/10
24/24 - 0s - loss: 0.0074 - val_loss: 0.0091 - 112ms/epoch - 5ms/step
Epoch 7/10
24/24 - 0s - loss: 0.0072 - val_loss: 0.0094 - 123ms/epoch - 5ms/step
Epoch 8/10
24/24 - 0s - loss: 0.0072 - val_loss: 0.0092 - 123ms/epoch - 5ms/step
Epoch 9/10
24/24 - 0s - loss: 0.0073 - val_loss: 0.0091 - 144ms/epoch - 6ms/step
Epoch 10/10
24/24 - 0s - loss: 0.0073 - val_loss: 0.0094 - 116ms/epoch - 5ms/step
<keras.callbacks.History at 0x7d3305f86bf0>
```

Figure 6. The result is different for each ten iterations.

3) **Denormalization.** This is a step in which a scaler returns the result to a standard form.`inverse` function:
4) **Data Visualization** is the last step in which the model is visualized into a plot diagram in which each variable is presented. Note that the visualization may vary for each epoch and variable.

a. Rainfall Data Visualization

Calculations using the Adam optimization model on the Rainfall variable with a variation of 10 epoch values are presented in the graphical visualization in Figure 7.

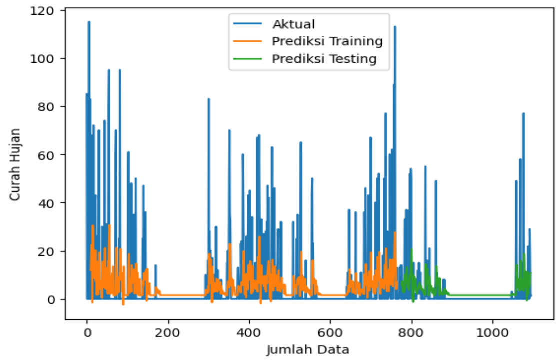

Figure 7. Adam Epoch Rainfall Graph Epoch 10

b. Water Discharge Data Visualization

Calculations using the Adam optimization model on the Water Discharge variable with a variation of 75 epoch values are presented in the graphical visualization in Figure 8.

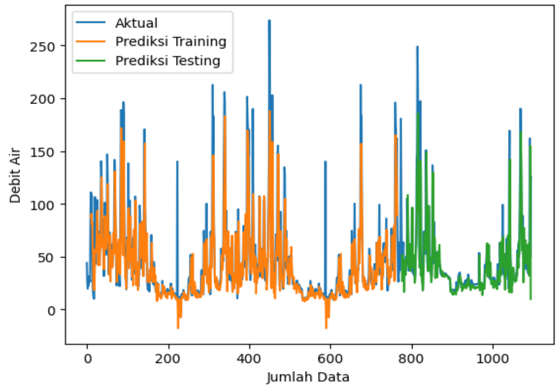

Figure 8. Adam Epoch Water Discharge Graph Epoch 75

Performance Evaluation

Based on the calculation of the epoch variations, an evaluation will be carried out using the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

The formula for each of them is as follows:

$$MAE(y, \hat{y}) = \frac{\sum_{i=0}^{N-1} |y_i - \hat{y}_i|}{N} \quad (1)$$

$$RMSE(y, \hat{y}) = \sqrt{\frac{\sum_{i=0}^{N-1} \left(y_i - \hat{y}_i\right)^2}{N}} \quad (2)$$

The documentation on both training data and testing data for the rainfall variable results of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) for each epoch is presented in Table 3 below:

Table 3. Rainfall evaluation results

No	Epoch	Training		Testing	
		MAE	RMSE	MAE	RMSE
1	10	5.73	9.40	5.95	11.12
2	50	7.31	11.72	6.05	10.62
3	**75**	**0.054**	**0.099**	**0.041**	**0.091**
4	100	7.73	10.77	7.65	11.10
5	150	6.43	9.56	7.00	12.11

Table 3 shows the acceptable value of performance evaluation results for the rainfall variable on epoch = 75 for both training, which scored 0.054 for MAE, 0.099 for RMSE, and testing data, 0.041 for MAE and 0.091 for RMSE.

In the same way as the previous variable, the following documentation on both training data and testing data for the water discharge variable results of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) for each epoch is presented:

Table 4. Water Discharge evaluation results

No	Epoch	Training		Testing	
		MAE	RMSE	MAE	RMSE
1	10	15.96	25.13	13.95	22.06
2	50	13,11	22.85	11.91	21.33
3	**75**	**13.93**	**21.53**	**12.87**	**21.08**
4	100	12.43	21.82	**11.37**	**21.08**
5	150	**11.10**	**18.61**	12.07	21.28

The result shows that the minor performance and water discharge evaluation scores differ for training and testing data. Training data results MAE = 11.10 and RMSE = 18.61 in epoch (n) = 150. Testing data results MAE = 11.37 and RMSE = 21.08 in epoch (n) = 100.

This result shows two anomalies:

a) For both training and testing data results, a high value of MAE and RMSE, which are far from 0 (zero);
b) The lowest score of both MAE and RMSE in training and testing data lies on different epochs. Training data is on epoch (n) = 150, while testing data is on epoch (n) = 100.

The dataset shows no zero value for the water discharge column (which means it is impossible to find the river dry).

CONCLUSIONS AND SUGGESTIONS

Conclusion

Implementing the LSTM method on the variables of rainfall and water discharge in certain epochs variations result in calculations of future data projections with certain conditions. Based on the research results, it can be concluded that the rainfall variable reached an acceptable accuracy on epoch 75 with a Mean Absolute Error (MAE) of 0.054 and the Root Mean Square Error (RMSE) of 0.099 for the training data. Also, it has acceptable accuracy on epoch 75 with a Mean Absolute Error (MAE) of 0.041 and the Root Mean Square Error (RMSE) of 0.091 for the testing data. The water discharge variable had anomalies, as the minor score was too far from the acceptable training and testing data score. Training data results MAE = 11.10 and RMSE = 18.61 in epoch (n) = 150, while testing data results MAE = 11.37 and RMSE = 21.08 in epoch (n) 100.

Future Work and recommendation

Since this study only compares two variables, namely rainfall and water discharge, it is recommended for further research to use more variables or other neural network methods (algorithms) and a comparative analysis process using several methods at once so that it can be seen that the performance results can be better than this study. The anomalies found in water discharge performance evaluation should be verified from another perspective as several reasons may cause the high score on MAE and RMSE. It is the modelling scheme that might not support non-zero datasets.

REFERENCES

Asdak, C. (2023). *Hidrologi dan Pengelolaan Daerah Aliran Sungai*. Yogjakarta: UGM PRESS.

Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. A. (2018). Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies*, *11*(7). https://doi.org/10.3390/en11071636

Devi, N. M. M. C., Bayupati, I. P. A., & Wirdiani, N. K.

A. (2022). Prediksi Curah Hujan Dasarian dengan Metode Vanilla RNN dan LSTM untuk Menentukan Awal Musim Hujan dan Kemarau. *JEPIN*, *8*(3), 405–411. Retrieved from https://jurnal.untan.ac.id/index.php/jepin/article/view/56606

Elizabeth Michael, N., Mishra, M., Hasan, S., & Al-Durra, A. (2022). Short-Term Solar Power Predicting Model Based on Multi-Step CNN Stacked LSTM Technique. *Energies*, *15*(6). https://doi.org/10.3390/en15062150

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Kardhana, H., Valerian, J. R., Rohmat, F. I. W., & Kusuma, M. S. B. (2022). Improving Jakarta's Katulampa Barrage Extreme Water Level Prediction Using Satellite-Based Long Short-Term Memory (LSTM) Neural Networks. *Water*, *14*(9), 1–17. https://doi.org/10.3390/w14091469

Kouadri, S., Pande, C. B., Panneerselvam, B., Moharir, K. N., & Elbeltagi., A. (2022). Prediction of irrigation groundwater quality parameters using ANN, LSTM, and MLR models. *Environmental Science and Pollution Research*, *29*, 21067–21091. https://doi.org/10.1007/s11356-021-17084-3

Larose, D. T. (2005). Discovering Knowledge in Data: An Introduction to Data Mining. *Discovering Knowledge in Data: An Introduction to Data Mining*, 2nd ed., pp. 1–222. New Jersey: John Willey & Sons Inc. https://doi.org/10.1002/0471687545

Navlan, A., Fandango, A., & Idris, I. (2021). *Python Data Analysis: Perform data collection, data processing, wrangling, visualization, and model building using Python*. Birmingham, United Kingdom: Packt Publishing Ltd.

Noymanee, J., & Theeramunkong, T. (2019). Flood Forecasting with Machine Learning Technique on Hydrological Modeling. *Procedia Computer Science*, *156*, 377–386. https://doi.org/10.1016/j.procs.2019.08.214

Rizki, M., Basuki, S., & Azhar, Y. (2020). Implementasi Deep Learning Menggunakan Arsitektur Long Short Term Memory(LSTM) Untuk Prediksi Curah Hujan Kota Malang. *Jurnal Repositor*, *2*(3), 331–338. https://doi.org/10.22219/repositor.v2i3.470

Sampurno, J., Vallaeys, V., Ardianto, R., & Hanert, E. (2022). Integrated hydrodynamic and machine learning models for compound flooding prediction in a data-scarce estuarine delta. *Nonlinear Processes in Geophysics*, *29*(3), 301–315. https://doi.org/10.5194/npg-29-301-2022

Shetty, S. A., Padmashree, T., Sagar, B. M., & Cauvery, N. K. (2021). Performance Analysis on Machine Learning Algorithms with Deep Learning Model for Crop Yield Prediction. *Data Intelligence and Cognitive Informatics*, 739–750. Springer, Singapore. https://doi.org/10.1007/978-981-15-8530-2_58

Sudriani, Y., Ridwansyah, I., & A Rustini, H. (2019). Long short term memory (LSTM) recurrent neural network (RNN) for discharge level prediction and forecast in Cimandiri river, Indonesia. *IOP Conference Series: Earth and Environmental Science*, *299*(1). https://doi.org/10.1088/1755-1315/299/1/012037

Supatmi, S., Hou, R., & Sumitra, I. D. (2019). Study of Hybrid Neurofuzzy Inference System for Forecasting Flood Event Vulnerability in Indonesia. *Computational Intelligence and Neuroscience*, *2019*, 1–12. https://doi.org/10.1155/2019/6203510

Szandała, T. (2020). Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks. In *Bio-inspired Neurocomputing* (pp. 203–224). Springer, Singapore. https://doi.org/10.1007/978-981-15-5495-7_11

Wang, W., & Lu, Y. (2018). Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model. *IOP Conference Series: Materials Science and Engineering*, *324*(1). https://doi.org/10.1088/1757-899X/324/1/012049