

SIGMA: JURNAL PENDIDIKAN MATEMATIKA DAN SAINS

Volume 4, Issue 2, 27-34, August 2026

e-ISSN: 2987-2669

<https://ejournal.unibo.ac.id/index.php/sigma>

MODEL TIME SERIES DENGAN EXOGENOUS VARIABLE MENGUNAKAN LONG SHORT TERM MEMORY (LSTM) UNTUK PREDIKSI CURAH HUJAN DI KABUPATEN SIDOARJO

Muhammad Ikrom^{1*}, Leady Dione Alfa Giovanni², Elfa Khorifa Izlia Aqnuri³

¹Pendidikan Matematika, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Bondowoso, Jawa Timur, Indonesia

²Pendidikan Matematika, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Bondowoso, Jawa Timur, Indonesia

³Pendidikan Bimbingan Konseling, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Bondowoso, Jawa Timur, Indonesia

[1*muhammadikrom0405@gmail.com](mailto:muhammadikrom0405@gmail.com), [2leadydione17@gmail.com](mailto:leadydione17@gmail.com), [3elfaizlia@gmail.com](mailto:elfaizlia@gmail.com)

Received: August 10, 2026 | Revised: August 17, 2026 | Accepted: August 21, 2026 | Published: August 31, 2026

*Corresponding author

Abstract:

Sidoarjo Regency is one of the areas that has about 90% of technically irrigated rice fields with low-lying conditions, so that when rainfall is high, some rice fields experience flooding and can result in decreased crop productivity. Therefore, prediction of rainfall is needed to overcome this. In addition, Sidoarjo district has a weather station close to the airport, making it possible to obtain more complete data as an exogenous variable in predicting rainfall. Besides that, there are several difficulties when including exogenous variables, namely having to ensure that exogenous variables are relevant to the forecasted time series data and possibly missing values or experiencing measurement errors in the data. The method used to predict rainfall is using Long Short Term Memory (LSTM) with the advantage of having a memory block that is able to find the output value most relevant to the given input, and using the backward method to overcome missing values. The results of the study using daily period rainfall data obtained an RMSE value of 13.68 when using yesterday's timestep data ($t-1$) with the proportion of training and testing data of 90%:10% and the best parameters were batch size 32, epoch 100, and neuron 10, and overcome the missing value by using the backward method. While the results of the study using monthly rainfall data obtained an RMSE value of 2.11 when using yesterday's timestep data ($t-1$) with the proportion of training and testing data of 90%:10% and the best parameters were batch size 32, epoch 100, and 30 neurons.

Keywords: Prediction, Rainfall, Multiattribute, Missing Value, LSTM

How to Cite: Ikrom, A., Geovanni, B., & Aqnuri, C. (2026). Model Time Series Dengan Exogenous Variable Menggunakan Long Short Term Memory (Lstm) Untuk Prediksi Curah Hujan Di Kabupaten Sidoarjo. SIGMA : Jurnal Pendidikan Matematika dan Sains, 4(2), 17-24. <https://doi.org/10.35316/sigma.2026-vvix.xx-xx>



Content from this work may be used under the terms of the [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/) that allows others to share the work with an acknowledgment of the work's authorship and initial publication in this journal.

Pendahuluan

Secara geografis, Indonesia terletak pada garis khatulistiwa yang berarti masuk ke dalam negara beriklim tropis yang memiliki dua musim yaitu musim hujan dan musim kemarau. Kedua musim tersebut biasanya terjadi secara bergantian setiap enam bulan sekali. Namun, terdapat beberapa daerah di Indonesia memiliki musim hujan lebih panjang daripada musim kemarau yang biasanya disebabkan oleh fenomena cuaca yang tidak menentu, salah satunya yaitu di Kabupaten Sidoarjo.

Kabupaten Sidoarjo merupakan salah satu kota di Provinsi Jawa Timur yang terletak pada $7^{\circ}03'-7^{\circ}05'$ LS dan $112^{\circ}05'-112^{\circ}09'$ BT dengan luas daerah sebesar 634,39 km². Kabupaten Sidoarjo memiliki lahan sawah sekitar 90% berpengairan teknis dengan kondisi daerah dataran rendah, sehingga ketika curah hujan tinggi, sebagian area persawahan mengalami banjir dan menyebabkan produktivitas tanaman menurun [1]. Oleh karena itu, prediksi curah hujan sangat dibutuhkan untuk mengatasi hal tersebut. Selain itu Kabupaten Sidoarjo mempunyai stasiun cuaca yang dekat dengan bandara dibandingkan dengan stasiun cuaca di daerah lain, sehingga data yang diperoleh untuk melakukan prediksi curah hujan lebih lengkap. Dimana prediksi curah hujan dapat digunakan untuk mengetahui potensi curah hujan yang akan terjadi.

Pada saat melakukan prediksi curah hujan sering kali terdapat beberapa kesulitan atau tantangan ketika memasukkan variabel eksogen (eksternal) dalam model time series. Salah satu kesulitannya adalah memastikan bahwa variabel eksogen relevan dan bermakna untuk data time series yang diramalkan. Dalam beberapa kasus, hubungan antara variabel eksogen dan deret waktu mungkin tidak dipahami dengan baik, sehingga sulit untuk menggabungkan variabel eksogen secara efektif ke dalam model peramalan. Tantangan lain adalah memastikan bahwa variabel eksogen selaras dengan data deret waktu, hal ini sangat penting ketika variabel eksogen dikumpulkan pada frekuensi yang berbeda atau dengan lag yang berbeda dibandingkan dengan data deret waktu.

Di samping itu, terdapat masalah kualitas dan ketersediaan data. Variabel eksogen mungkin memiliki missing value atau mengalami kesalahan pengukuran sehingga dapat berdampak negatif terhadap keakuratan ramalan jika tidak ditangani dengan benar [2]. Oleh karena itu, meskipun memasukkan variabel eksogen dalam peramalan deret waktu dapat menjadi cara yang efektif untuk meningkatkan akurasi peramalan, namun sangat penting untuk mempertimbangkan dengan hati-hati tantangan dan kesulitan yang ada pada saat melakukannya. Selain itu, penggunaan metode dan analisis yang tepat perlu diperhatikan dalam melakukan suatu peramalan.

Machine Learning telah menjadi alat yang semakin penting dalam pengembangan model untuk prakiraan cuaca dan siaran sekarang. Penggunaan algoritma pembelajaran mesin telah memungkinkan ahli meteorologi untuk membuat prediksi yang lebih akurat dengan menggabungkan data yang lebih luas dan meningkatkan representasi sistem cuaca yang kompleks. Salah satu metode yang dapat digunakan untuk memprediksi curah hujan yaitu Long Short Term Memory (LSTM). Metode LSTM merupakan pengembangan dari neural network untuk menganalisis data time series dalam memprediksi suatu nilai dengan rentan waktu yang ditentukan. Kelebihan dari metode LSTM yaitu mempunyai blok memori yang mampu mencari nilai output paling relevan dengan input yang diberikan. Metode LSTM juga memiliki cell yang menyimpan suatu nilai (cell state) baik dalam periode jangka pendek maupun jangka panjang [3].

Pada tahun 2019, Atika et al., melakukan penelitian terhadap peramalan curah hujan di kota Surabaya menggunakan LSTM dan Moving Average tunggal (MA). Hasil yang diperoleh pada penelitian ini memiliki nilai akurasi RMSE berkisar antara 13.766

sampai dengan 15.719 [4]. Zhang et al., (2021) melakukan penelitian terkait peramalan penyakit tangan, kaki, dan mulut (HMFD) di Ningbo, China menggunakan perbandingan LSTM dan Autoregressive Integrated Moving Average (ARIMA) dengan melibatkan variabel eksogen. Penelitian ini dilakukan dengan empat skenario pengujian yaitu skenario pertama menggunakan LSTM dengan variabel eksogen, skenario kedua menggunakan LSTM tanpa variabel eksogen, skenario ketiga menggunakan ARIMA dengan variabel eksogen, skenario keempat menggunakan ARIMA tanpa variabel eksogen. Hasil yang diperoleh pada penelitian ini yaitu kedua skenario LSTM lebih baik daripada kedua skenario ARIMA dalam peramalan HMFD di Ningbo, China. Dimana skenario LSTM dengan variabel eksogen memiliki kinerja yang lebih optimal dibandingkan tanpa variabel eksogen [5]. Pada tahun 2023, Carnegie et al., melakukan penelitian terhadap perbandingan LSTM dan GRU untuk memprediksi curah hujan. Penelitian ini membagi dataset training dan testing menjadi tiga bagian yaitu 7:3, 8:2, 9:1 dimana dataset 7:1 memiliki hasil terbaik dalam memprediksi curah hujan. Nilai akurasi terbaik yang diperoleh pada penelitian ini yaitu RMSE sebesar 16.81 [6].

Berdasarkan sumber literatur yang telah dijelaskan, peneliti tertarik untuk memprediksi curah hujan menggunakan metode LSTM. Penelitian ini akan dilakukan dengan 9 variabel dari Kabupaten Sidoarjo. Variabel tersebut yaitu temperatur minimum, temperatur maksimum, temperatur rata-rata, kelembaban rata-rata, curah hujan, lamanya penyinaran matahari. Kecepatan angin maksimum, arah angin saat kecepatan maksimum, dan kecepatan angin rata-rata. Penelitian ini diharapkan dapat memberikan informasi kepada masyarakat di Kabupaten Sidoarjo terkait prediksi curah hujan.

Metode Penelitian

Penelitian ini dilakukan secara sistematis dengan langkah-langkah sebagai berikut:

2.1 Data Penelitian

Data yang digunakan adalah data curah hujan di Kabupaten Sidoarjo yang diperoleh dari *website* resmi Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Jumlah data yang didapatkan sebanyak 853 data dengan periode harian dari tanggal 1 Oktober 2020 sampai dengan tanggal 31 Januari 2023. Dimana inputan data yang digunakan untuk melakukan prediksi curah hujan hari ini (t) akan berdasarkan data kemarin ($t - 1$) dan data kemarin lusa ($t - 2$). Variabel eksogen yang digunakan meliputi temperatur minimum, temperatur maksimum, temperatur rata-rata, kelembaban rata-rata, lama penyinaran matahari, kecepatan angin maksimum, arah angin, dan kecepatan angin rata-rata.

2.2 Pre-processing data

Pada tahap *pre-processing* data terdapat 2 aktivitas yang dilakukan yaitu mengatasi *missing value* menggunakan cara sederhana dan normalisasi data. Pada saat mengatasi *missing value* metode yang digunakan yaitu *backward*, *forward* serta gabungan antara metode *backward* dan *forward*. Metode *backward* akan mengisi nilai yang *missing* pada data menggunakan isian mundur sedangkan metode *forward* akan mengisi nilai yang *missing* pada data menggunakan isian maju. Setelah *missing value* teratasi aktivitas selanjutnya yaitu melakukan normalisasi data menggunakan teknik *min - max scaling* menggunakan persamaan berikut ini [7]:

$$X' = \frac{x - \min_x}{(\max_x - \min_x)}$$

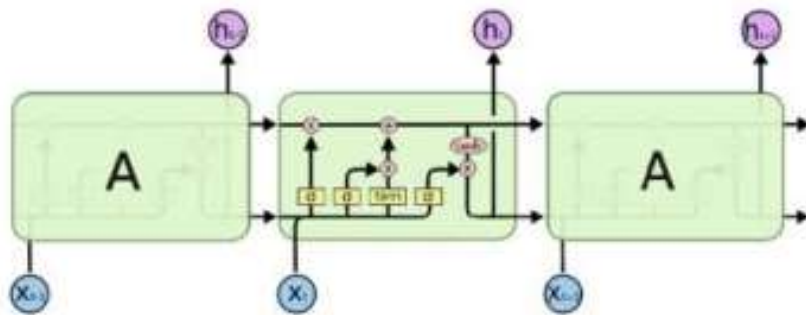
2.3 Pembagian Data *Training* dan Data *Testing*

Data curah hujan dari tanggal 1 Oktober 2020 sampai 31 Januari 2023 akan dibagi menjadi data *training* dan data *testing*, dimana akan dilakukan dengan tiga kali percobaan dengan proporsi data yaitu 90%: 10%, 80%: 20%, dan 70%: 30% . Data *training* digunakan untuk melatih algoritma pembentukan sebuah model, sedangkan data *testing* digunakan untuk mengukur sejauh mana tingkat akurasi dan performa yang diperoleh dari data *training*.

Proses *training* dan *testing* data dalam pembentukan model menggunakan LSTM diperlukan parameter untuk memperoleh model terbaik. Penelitian ini menggunakan hiperparameter untuk mendapatkan model terbaik. Parameter yang digunakan yaitu dengan mencoba nilai *batch size* 8,16,32, *epoch* 50,100, dan *neurons* 10,20,30. Dimana hasil akhir akan memunculkan *grid search* terbaik dari model yang didapatkan.

2.4 Long Short Term Memory (LSTM)

Long Short Term Memory (LSTM) merupakan pengembangan dari algoritma RNN. Struktur LSTM memiliki persamaan dengan RNN yaitu terdiri dari *input layer*, *hidden layer*, dan *output layer*. Perbedaan LSTM dan RNN terletak pada *hidden layer*, dimana *hidden layer* pada LSTM memiliki fungsi aktivasi tanh dan sigmoid, sedangkan *hidden layer* pada RNN memiliki fungsi aktivasi tanh [8]. LSTM memiliki komponen *gates* atau gerbang yang berfungsi untuk mengontrol atau menyeleksi aliran informasi yang masuk *kedalam recurrent cell* [9].



Gambar 1. Arsitektur Long Short Term Memory (LSTM)

Arsitektur LSTM memiliki tiga *gates* yaitu *forget gate* (f_t), *input gate* (i_t), dan *output gate* (o_t). *Forget gate* (f_t) digunakan untuk menghapus informasi dari *cell state*, *input gate* (i_t) digunakan untuk memutuskan informasi baru yang masuk *kedalam cell state*, dan *output gate* (o_t) digunakan untuk memilah informasi yang berguna dari arus *cell state* dan menampilkannya sebagai *output* [10]. Persamaan yang digunakan untuk proses LSTM sebagai berikut:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3)$$

$$C_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (4)$$

$$C_t = f_t \times C_{t-1} + i_t \times C_t \quad (5)$$

$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = o_t \times \tanh(C_t) \quad (7)$$

2.5 Denormalisasi

Denormalisasi merupakan tahap yang dilakukan sebelum menghitung hasil akurasi hasil prediksi [3]. Denormalisasi digunakan untuk mengubah data menjadi real. Persamaan dari denormalisasi sebagai berikut,

$$d = d'(max - min) + min \quad (8)$$

2.6 Evaluasi RMSE

Root Mean Square Error atau RMSE dapat digunakan untuk mengukur keakuratan suatu model prediksi. Model dikatakan akurat apabila nilai RMSE mendekati angka 0. Semakin kecil nilai RMSE maka tingkat keakuratannya semakin besar, begitu pula sebaliknya. RMSE dilandasi oleh nilai error antara data aktual dengan data prediksi [11]. RMSE memiliki rumus persamaan

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (9)$$

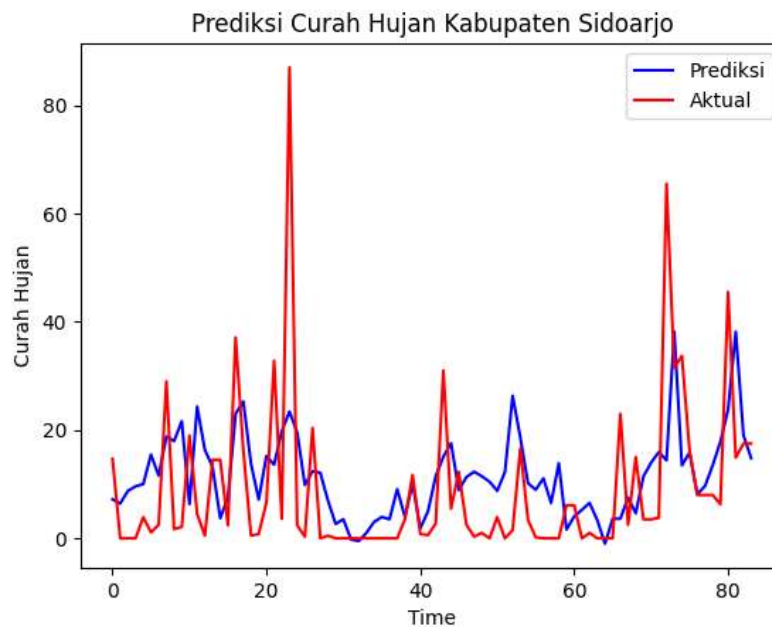
Hasil dan Pembahasan

Hasil pada penelitian ini yang menggunakan algoritma LSTM dalam melakukan prediksi curah hujan di kabupaten sidoarjo dapat dilihat pada Tabel 1. Dimana pengujian model dilakukan berkali-kali menggunakan *hyperparameter* untuk mendapatkan model yang terbaik.

Tabel 1. Hasil pengujian model

<i>Timestep</i>	<i>Handling Missing Value</i>	<i>Splitting data</i>	<i>Batch Size</i>	<i>Epoch</i>	<i>Neuron Hidden</i>	<i>RMSE Training</i>	<i>RMSE Testing</i>
<i>t - 1</i>	<i>Backward</i>	90%:10%	32	100	10	15,13	13,68
		80%:20%	8	100	10	14,36	17,13
		70%:30%	32	100	10	14,28	15,86
	<i>Forward</i>	90%:10%	32	100	20	15,15	13,82
		80%:20%	8	50	20	14,51	17,14
		70%:30%	32	100	20	14,43	15,90
	<i>Backward & Forward</i>	90%:10%	32	100	30	15,08	13,89
		80%:20%	8	50	30	15,51	17,23
		70%:30%	16	100	10	14,37	15,87
	<i>Backward</i>	90%:10%	16	100	20	13,93	16,40
		80%:20%	16	50	30	13,80	18,20
		70%:30%	32	100	30	13,78	16,34
<i>t - 2</i>	<i>Forward</i>	90%:10%	8	50	20	14,61	15,14
		80%:20%	8	50	20	13,84	17,91
		70%:30%	16	100	20	13,68	16,43
	<i>Backward & Forward</i>	90%:10%	16	100	20	14,26	15,82
		80%:20%	8	50	10	14,05	18,06
		70%:30%	16	50	30	13,80	16,36

Hasil prediksi curah hujan di Kabupaten Sidoarjo berdasarkan Tabel 1 menggunakan data model *time series* dengan variabel eksogen memperoleh model terbaik pada saat menggunakan *timestep* untuk prediksi curah hujan hari ini (*t*) yaitu berdasarkan data kemarin (*t - 1*). *Missing value* pada dataset diatasi dengan menggunakan metode *backward*, yaitu dengan cara mengisi nilai yang kosong pada data berdasarkan nilai setelahnya (isian mundur). Nilai dari *batch size*, *epoch* dan *neuron* berdasarkan *hyperparameter* terbaik yang didapatkan berturut-turut yaitu 32, 100, dan 10, serta berhasil memperoleh nilai RMSE *testing* sebesar 13,68. Dimana plot perbandingan antara data aktual dan data prediksi berdasarkan model data *testing* terbaik dapat dilihat pada Gambar 2.



Gambar 2. Plot perbandingan data aktual dan hasil prediksi pada data *testing*

Gambar 2 merupakan hasil prediksi curah hujan yang melibatkan variabel eksogen dengan nilai RMSE terkecil pada data *testing* berdasarkan hasil *hyperparameter* terbaik. Dimana hasil data prediksi ditunjukkan dengan garis berwarna biru dan data aktual curah hujan ditunjukkan dengan garis warna merah.

Nilai RMSE data *testing* terbaik pada Tabel 1 akan dilakukan perbandingan berdasarkan hasil yang akan diperoleh terhadap nilai RMSE prediksi curah hujan menggunakan data periode bulanan. Proses perbandingan pada penelitian ini dilakukan untuk melihat performa model dalam mendapatkan model terbaik pada saat melakukan prediksi curah hujan menggunakan data periode harian dan data periode bulanan. Pada penelitian ini prediksi curah hujan menggunakan data periode bulanan dimulai sejak awal bulan Oktober 2020 sampai dengan bulan Januari 2023.

Prediksi curah hujan berdasarkan data periode bulanan dilakukan menggunakan metode LSTM yang melibatkan variabel eksogen dengan membangun model *hyperparameter* untuk mendapatkan model terbaik, serta melakukan beberapa *splitting* data *training* dan data *testing* dengan porsi yang berbeda. Hasil prediksi curah hujan yang menggunakan data periode bulanan dapat dilihat pada Tabel 2 berikut ini.

Tabel 2. Prediksi curah hujan periode bulanan

<i>Timestep</i>	<i>Splitting data</i>	<i>Batch Size</i>	<i>Epoch</i>	<i>Neuron Hidden</i>	<i>RMSE Training</i>	<i>RMSE Testing</i>
t – 1	90%: 10%	32	100	30	5,25	2,11
	80%: 20%	32	100	30	5,37	2,37
	70%: 30%	16	50	20	6,10	3,40
t – 2	90%: 10%	16	100	30	3,36	7,52
	80%: 20%	8	50	10	5,72	2,20
	70%: 30%	16	50	10	6,25	3,46

Berdasarkan Tabel 2 dapat diketahui bahwa hasil prediksi curah hujan menggunakan data periode bulanan memperoleh model LSTM terbaik pada saat menggunakan *timestep* data curah hujan kemarin ($t - 1$), dengan proporsi data *training* dan *testing* yaitu 90%: 10%. Parameter terbaiknya yaitu *batch size* bernilai 32, *epoch* bernilai 100 dan *neuron* bernilai 30. Nilai RMSE terkecil yang didapatkan pada proses *testing* sebesar 2,11 dan nilai RMSE pada saat proses *training* sebesar 5,25.

Berdasarkan nilai RMSE data *testing* yang didapatkan pada Tabel 1 dan Tabel 2, prediksi curah hujan terbaik yang melibatkan variabel eksogen menggunakan metode LSTM diperoleh pada saat menggunakan data periode bulanan. Dimana nilai RMSE data *testing* periode bulanan sebesar 2,11, sedangkan nilai RMSE data *testing* periode harian sebesar 13,68. Hal ini dikarenakan pada dataset periode harian terdapat fluktuasi yang dianggap sebagai akibat pencilaan pada data, sedangkan pada dataset periode bulanan relatif memiliki nilai yang konstan.

Kesimpulan dan Saran

Berdasarkan hasil analisis dan pengujian yang dilakukan dapat diperoleh kesimpulan sebagai berikut:

- a. Cara sederhana dalam mengatasi *missing value* dalam data *time series* pada model LSTM yaitu menggunakan metode *backward*.
- b. Model LSTM terbaik menggunakan data periode harian yang melibatkan variabel eksogen diperoleh pada saat menggunakan *timestep* data kemarin ($t - 1$) dengan proporsi data *training* dan *testing* yaitu 90%: 10% dan parameter terbaiknya yaitu *batch size* bernilai 32, *epoch* bernilai 100, dan *neuron* bernilai 10. Sedangkan Model LSTM terbaik menggunakan data periode bulanan diperoleh pada saat menggunakan *timestep* data kemarin ($t - 1$) dengan proporsi data *training* dan *testing* yaitu 90%: 10% dan parameter terbaiknya yaitu *batch size* bernilai 32, *epoch* bernilai 100, dan *neuron* bernilai 30.

Hasil prediksi dari model terbaik menggunakan data periode harian yang melibatkan variabel eksogen mampu menghasilkan nilai RMSE terkecil sebesar 13.68 dengan kemampuan peramalan selama 28 hari yang memperoleh nilai RMSE sebesar 24.6. Sedangkan hasil prediksi model terbaik dengan menggunakan data periode bulanan menghasilkan nilai RMSE terkecil sebesar 2.11.

Daftar Pustaka

- [1] Badan Pusat Statistik Jawa Timur. 2018. Statistik Daerah Kabupaten Sidoarjo. Sidoarjo: Badan Pusat Statistik
- [2] Chatfield, C. 2000. The Analysis of Time Series. New York: An Introduction (Vol. 6). CRC Press.
- [3] Wiranda, L., dan M. Sadikin. 2019. Penerapan Long Short Term Memory Pada Data Time Series untuk Memprediksi Penjualan Produk PT. Metiska Farma. Jurnal Nasional Pendidikan Teknik Informatika. 8(3): 184-196.
- [4] Atika, R., A. Fariza, dan A. R. Barakbah. 2019. Forecast Rainfall Data Time Series Using MultiAttribute Long Short Time Memory. Surabaya: IEEE.
- [5] Zhang, R., Z. Guo, Y. Meng, S. Wang, S. Li, R. Niu, Y. Wang, Q. Guo, dan Y. Li. 2021. Comprasion of ARIMA and LSTM the Incidence of HMFD Combined and Uncombined with Exogenous Meteorological Variabels in Ningbo, China. Int J Environ Res Public Health. 18(11): 6174

-
- [6] Carnegie, A. D. M., Chairani. 2023. Perbandingan Long Short Term Memory (LSTM). *Jurnal Media Informatika Budidarma*. 7(3): 1022 -1032.
 - [7] Aldi, M. W. P., Jondri, dan A. Aditsania. 2018. Analisis dan Implementasi Long Short Term Memory Neural Network Untuk Prediksi Harga Bitcoin. *e -Proceeding of Engineering*. 5(2): 3548.
 - [8] Salahinejad, H., S. Sankar, J. Barfett, E. Colak, dan S. Valaee. 2018. Recent Advances in Recurrent Neural Network. *CoRR*, abs/1801.01078.
 - [9] Maulidani, S. S., dan N. Ihsan, Sulistiawaty. 2015. Analisis Pola dan Intensitas Curah Hujan Berdasarkan Data Observasi dan Satelit Tropical Rainfall Measuring Missions (TRMM) 3B42 V7 di Makassar. *Jurnal Sains dan Pendidikan Fisika*. 11(1): 98 -103.
 - [10] Lazzeri, F. 2021. *Machine Learning for Time Series Forecasting with Python*. Canada: Wiley.
 - [11] Wei, W. W. S. 2006. *Time Series Analysis Univariate and Multivariate Method*. Second edition. New York: Addison Wesley Publishing Company, Inc.