

Wind Speed Forecasting Using Long Short-Term Memory (LSTM) and Meteorological Variables at Tanjung Perak, Surabaya, Indonesia

Yesiska M Siallagan, Tugiyono Aminoto*, Husnul Hamdi

Physics Study Program, Faculty of Science and Technology, University of Jambi, Jambi 36361, Indonesia

(Received October 30, 2025; revised June 05, 2026; accepted June 22, 2026; published online October 24, 2026)

This study evaluated Long Short-Term Memory (LSTM) models for wind-speed prediction using meteorological observations from Tanjung Perak, Surabaya, for 2015–2024. Air temperature, relative humidity, atmospheric pressure, and rainfall were used as predictors. Data preprocessing, alternative training–test partitions (70:30, 80:20, and 90:10), and model configurations were evaluated using root mean square error (RMSE) and mean absolute error (MAE). Among the configurations reported for the 80:20 partition, 50 LSTM units and 50 epochs yielded the lowest test errors (RMSE = 1.2965 knots; MAE = 1.0175 knots). For January 2025, reported forecast errors were RMSE = 1.4253 and MAE = 1.2261 knots for 7 days; RMSE = 1.3178 and MAE = 1.0943 knots for 14 days; and RMSE = 1.2117 and MAE = 0.8886 knots for 31 days. These results refer to different verification periods and do not establish superior performance at longer forecast horizons. Multiple linear regression identified significant associations with temperature, humidity, and pressure ($p < 0.001$), while rainfall was not significant ($p = 0.164$). Further chronological and operational validation is needed.

Keywords: *Wind speed forecasting; Long Short-Term Memory (LSTM); Multivariate time series; Meteorological predictors; Maritime meteorology.*



This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-sa/4.0/) license.
Copyright © 2026 by Author. Published by Physical Society of Indonesia

1. INTRODUCTION

Accurate and timely weather information is essential for activities that depend on atmospheric conditions (Prayogo et al., 2021). Accordingly, the Indonesian Agency for Meteorology, Climatology and Geophysics (BMKG) is expected to provide reliable weather information. Advances in forecasting methods offer opportunities to improve the accuracy and timeliness of these services (Anwar, 2017).

Surabaya is a coastal city where meteorological conditions influence maritime transport and other economic activities. Variations in temperature, humidity, atmospheric pressure, rainfall, and wind speed are therefore important for local weather monitoring and forecasting (Sunarmi et al., 2022; Prasita et al., 2023).

Wind speed is an important meteorological parameter for maritime activities and weather forecasting. Wind is primarily driven by atmospheric pressure gradients and modified by surface friction and other atmospheric processes (Novrita et al., 2021). Wind speed is commonly measured with an anemometer (Putri et al., 2023).

Long Short-Term Memory (LSTM), introduced by Hochreiter and Schmidhuber (1997), is a recurrent neural network architecture designed to learn temporal dependencies. Its gating mechanism allows information to be retained or discarded over time, making LSTM suitable for time-series applications. In this study, historical meteorological observations, including air temperature, rainfall, air pressure, and relative humidity, are used as predictors of wind speed (Tanjung et al., 2024).

Air temperature is a measure that indicates the level of heat or cold in an environment, including the air itself. The unit used to measure this temperature is degrees Celsius ($^{\circ}\text{C}$) (Caturisari, 2024). Rainfall refers to the height of water collected in a flat area without evaporation, absorption, or flow. In addition, rainfall intensity indicates the amount of rainfall that occurs in a certain period. The unit used

*Contact Author: tugiyono@unja.ac.id

to measure rainfall is millimeters (mm) (Muflih et al., 2019). Humidity is a measure of the wetness of the air, which is caused by the presence of water vapor. It is the amount of water vapor in the air (Friadi & Junadhi, 2019). Air pressure can be referred to as the atmospheric pressure of the earth, as fluids change with altitude and weather. Air pressure is a very important parameter studied in meteorology. Air pressure data is essential for several fields, such as aviation services (vital data), isobar analysis, and so on. The units of air pressure are millibars (mb) or hectopascals (hPa) (Putri, 2024).

This study has three objectives: (1) to develop an LSTM-based wind-speed prediction model using air temperature, rainfall, atmospheric pressure, and relative humidity; (2) to evaluate the model using RMSE and MAE for the tested data partitions and forecast horizons; and (3) to investigate statistical associations between wind speed and the meteorological variables through multiple linear regression. Previous applications of recurrent models to meteorological prediction provide the context for this study (Luthfiarta et al., 2020; Supriyadi, 2021; Devi et al., 2022; Akbar et al., 2023).

2. METHOD

Daily meteorological observations for 2015–2024 were obtained from the Tanjung Perak Surabaya Maritime Meteorological Station of BMKG. The variables comprise air temperature (°C), relative humidity (%), atmospheric pressure (hPa), rainfall (mm), and wind speed. Wind speed is reported in knots in the results tables; any conversion from the source units must be documented and applied consistently.

Python was used for LSTM modelling, while SPSS was used for multiple linear regression. The workflow comprised checking and preprocessing observations, handling missing values by linear interpolation, applying min–max scaling, dividing observations into training and test subsets (70:30, 80:20, and 90:10), training alternative LSTM configurations, and evaluating predictions using RMSE and MAE. For a time-series forecast, the data must be split chronologically and scaling parameters fitted to training data only; the manuscript should confirm whether these safeguards were implemented.

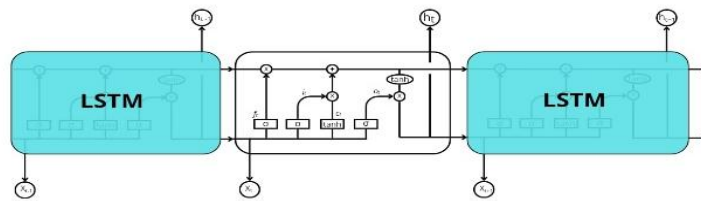


Figure 1. Schematic architecture of the Long Short-Term Memory (LSTM) network, illustrating the flow of information between successive LSTM cells and the mechanisms for retaining, updating, and transmitting temporal information through the forget, input, and output gates.

2.1 Input Gate (i_t)

The Input Gate plays an important role in the LSTM network, namely to receive output from the previous step and new input, then process it through the sigmoid activation function layer. This gate produces values between 0 and 1, which indicate how much information will be retained or discarded.

$$i_t = \sigma(W_i S_{t-1} + W_i X_t)$$

Where I_t is the input gate, σ is the sigmoid activation function, W_i is the weight of the input gate, S_{t-1} is the previous state or state at time t-1, X_t is the input at time t. The input gate value is multiplied by the output of the candidate layer ($\tilde{c}$). Where ($\tilde{c}$) can be formulated as follows:

$$\begin{aligned} (\tilde{c}) &= \tanh(W_c S_{t-1} + W_c X_t) \\ c_t &= (i_t * \tilde{c}_t + f_t * c_{t-1}) \end{aligned}$$

Where c_t is the intermediate cell state, W_c is the weight of the cell state, S_{t-1} is the previous state or state at time t-1, X_t is the input at time t, f_t is the forget gate.

2.2 Forget Gate (f_t)

The forget gate is a component that uses a sigmoid activation function to process the output at time t-1 and the input at time t, then combines them. Where, if $(f_t) = 0$ this indicates that the previous state will be ignored or forgotten. Conversely, if $(f_t)=1$, it means that the previous state is retained without change.

$$f_t = \sigma(W_f S_{t-1} + W_f X_t)$$

Where f_t is the forget gate, σ is the sigmoid activation function, W_f is the weight of the forget gate, S_{t-1} is the previous state or state at time t-1, X_t is the input at time t.

2.3 Output Gate (o_t)

The output gate functions to regulate how much information from the cell state will be passed on as output, and its mechanism is similar to the other gates in LSTM. This gate produces a new value from the cell state (h_t) which is used in the next step.

$$o_t = \sigma(W_o S_{t-1} + W_o X_t)$$

$$h_t = o_t * \tanh(c_t)$$

Where O_t is the output gate, σ is the sigmoid activation function, W_o is the weight of the output gate, S_{t-1} is the previous state or state at time t-1, X_t is the input at time t.

2.4 Model Evaluation

Model evaluation is carried out using several error metrics, namely RMSE (Root Mean Square Error) and MAE (Mean Absolute Error). According to Wiranda and Sadikin (2019):

2.4.1 RMSE (Root Mean Square Error)

is the average prediction error, but it is calculated by squaring the difference first, then taking the square root. Where, if the result gives a smaller error value, then the result has high accuracy.

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

Where $\hat{y}_i$ is the predicted value, y_i is the actual value, n is the number of data points.

2.4.2 MAE (Mean Absolute Error)

is the average difference between the predicted result and the actual data, calculated using absolute values (without squaring).

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

Where $\hat{y}_i$ is the predicted value, y_i is the actual value, n is the number of data points.

2.5 Multiple Linear Correlation Test

Regression testing is a statistical method used to determine the relationship between one variable and one or more other variables. In linear regression testing, there are two variables, namely the independent variable and the dependent variable. The independent variable is referred to as the independent variable (x), which is the variable that influences other variables. Meanwhile, the dependent variable is also referred to as the dependent variable (y), which is the variable that is influenced by the independent variable (Yuliara, 2016).

Regression testing is a statistical method to determine the relationship between one variable and one or more other variables. In analyzing data using SPSS, researchers use a significance level of alpha 5% or 0.05. This means that there is a 5% risk of rejecting the null hypothesis when the null hypothesis is true (Luthfiarta et al., 2020). The concept of significance level (α) was first introduced by statistician

Sir Ronald A. Fisher in the early 20th century. Fisher recommended that 0.05 was a “sufficient” threshold for declaring statistical test results significant.

The simple linear regression model can be written as:

$$Y = a + B_1X_1 + B_2X_2 + B_3X_3 + B_4X_4 + e$$

Where Y is the dependent variable, a is the constant, B_1, B_2, B_3, B_4 are the regression coefficients, X_1, X_2, X_3, X_4 are the independent variables, e is the residual or error.

Table 1. Classification of wind speed according to the Beaufort wind force scale, including wind speed ranges in knots, corresponding wind categories, and their observable effects.

Scale	Wind Speed (knots)	Name of Wind	Notes
0	0-1	Calm	Smoke rises straight up
1	1-3	Light breeze	Smoke rises at an angle
2	4-6	Gentle breeze	Leaves move
3	7-10	Moderate breeze	Twigs move
4	11-16	Tense wind	Branches move
5	17-21	Strong wind	Tree trunks move
6	22-27	Very strong wind	Large tree trunks move
7	28-33	Stormy wind	Branches break
8	34-40	Violent stormy wind	Small trees break
9	41-47	Hurricane	Large trees break
10	48-55	Severe hurricane	Houses collapse
11	56-63	Typhoon	Heavy objects fly around
12	64-lebih	Severe typhoon	Severe damage

2.6 Research Stages

The research was conducted through several systematic stages, as illustrated in Figure 2. First, meteorological data comprising air temperature, rainfall, atmospheric pressure, relative humidity, and wind speed were collected from the Tanjung Perak Surabaya Maritime Meteorological Station for the period 2015–2024. The collected data underwent preprocessing to identify missing values, inconsistencies, and potential errors that could affect model performance. Missing observations were addressed using linear interpolation to maintain the continuity of the time series. Subsequently, Min-Max normalization was applied to transform all variables into a standardized range of [0, 1], ensuring comparable numerical scales and facilitating stable neural network training. The normalized dataset was then divided into training and testing subsets using three different proportions: 70:30, 80:20, and 90:10. These configurations were examined to evaluate the influence of data allocation on predictive performance. The Long Short-Term Memory (LSTM) model was developed and implemented in Python using the Keras and TensorFlow libraries. Model optimization involved evaluating different neural network configurations, including variations in the number of neurons (25, 50, and 75) and training epochs (50 and 100). Each configuration was assessed to identify the most suitable combination for predicting wind speed from meteorological variables. Model performance was evaluated using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), which quantify the differences between predicted and observed wind speeds. The configuration producing the lowest prediction errors was selected for subsequent forecasting. Finally, the resulting predictions were analyzed to assess model performance and identify potential applications in maritime meteorological forecasting.

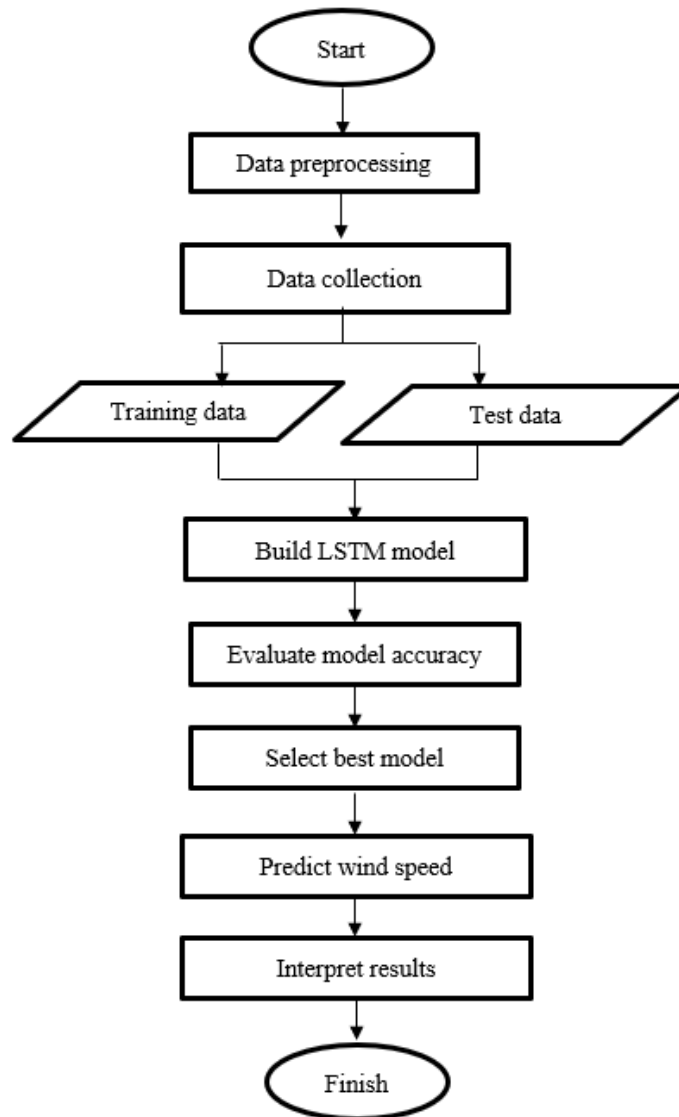


Figure 2. Flowchart of the research methodology for LSTM-based wind speed forecasting, illustrating the sequential stages of meteorological data collection, preprocessing, training–testing data partitioning, model development, performance evaluation, optimal model selection, wind speed prediction, and interpretation of results.

3. RESULTS AND DISCUSSION

3.1 Data Preprocessing

The meteorological variables have different units and numerical ranges. Min–max scaling maps each variable to a specified range, typically $[0, 1]$, and can improve numerical conditioning during LSTM training. Missing observations and date formatting must be handled separately from scaling. To avoid information leakage, minimum and maximum values should be estimated from the training subset and then applied without refitting to the test subset. The manuscript should specify whether this procedure was followed.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

where X' is the normalized value, X is the original value to be normalized, X_{min} is the minimum value of all data, X_{max} is the maximum value of all data.

Table 2. Representative samples of normalized meteorological data used for LSTM-based wind speed prediction, including air temperature, relative humidity, atmospheric pressure, rainfall, and wind speed, with all variables scaled to the range [0, 1].

Temperature (°C)	Humidity (%)	Air Pressure (hPa)	Rainfall (mm)	Wind Speed (knots)
0.439455	0.646992	0.310209	0.009421	0.482574
0.362896	0.694179	0.366181	0.143338	0.597855
0.379888	0.644783	0.388648	0.255720	0.463807
0.469437	0.606668	0.338589	0.006729	0.335121
0.466224	0.633169	0.314939	0.115074	0.352838

The training subset was used to estimate model parameters, whereas the held-out test subset was used to evaluate predictions. Three training-to-test ratios (70:30, 80:20, and 90:10) were considered. Because the observations form a time series, the chronological ordering of the partitions and the use of any separate validation subset should be clarified.

3.2 Model Training

The configurations reported in Table 3 use 50, 100, and 150 LSTM units with 50, 100, and 150 training epochs. These settings differ from the 25, 50, and 75 units and 50 and 100 epochs stated in Section 2; the authors should verify the actual experimental grid.

Table 3. Performance comparison of LSTM configurations with different numbers of hidden units and training epochs using an 80:20 training–test partition, evaluated in terms of Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

Neuron (number of units)	Epoch	RMSE	MAE
50	50	1.2965	1.0175
	100	1.3405	1.0481
	150	1.3267	1.0362
100	50	1.3524	1.0431
	100	1.3432	1.0378
	150	1.32	1.0261
150	50	1.3679	1.0468
	100	1.3415	1.0438
	150	1.3572	1.0514

For the 80:20 training–test partition, Table 3 presents the predictive performance of different Long Short-Term Memory (LSTM) configurations, with RMSE values ranging from 1.2965 to 1.3679 knots and MAE values ranging from 1.0175 to 1.0514 knots. Among the evaluated configurations, the model incorporating 50 LSTM units and 50 training epochs achieved the lowest prediction errors, with an RMSE of 1.2965 knots and an MAE of 1.0175 knots. These results suggest that this configuration provides the most favorable predictive performance among the tested alternatives, indicating that increasing the number of neurons or training epochs does not necessarily improve model accuracy. As illustrated in Figure 3, the predicted wind-speed series generally follows the temporal variations observed in the actual measurements, particularly in capturing broader fluctuations and prevailing patterns. However, the predicted values exhibit smoother variations than the observations, with noticeable discrepancies during periods characterized by abrupt increases or decreases in wind speed. This behavior suggests that the LSTM model captures dominant temporal patterns but has limited ability to reproduce short-term fluctuations and extreme wind-speed events. The relatively small differences between the evaluated configurations also indicate that model performance is moderately sensitive to architectural complexity and training duration. Nevertheless, the reported RMSE and MAE values alone cannot establish whether the model generalizes reliably to unseen observations or is free from overfitting. Further assessment using chronological validation, training and validation loss curves, and

independent testing would strengthen the reliability of these findings and support the selection of an optimal forecasting configuration.

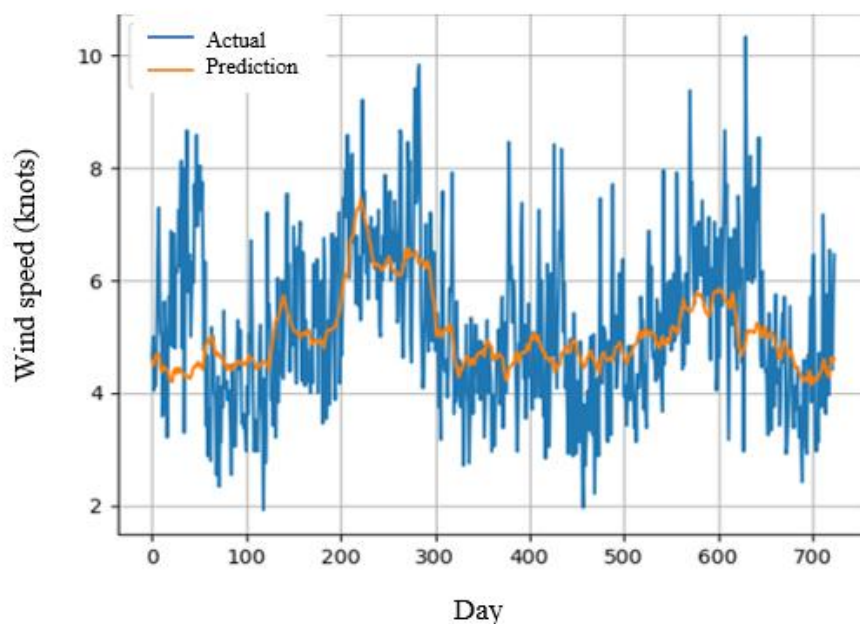


Figure 3. Comparison of observed and LSTM-predicted wind speeds in the test dataset, illustrating the model's ability to capture general temporal variations while exhibiting smoother predictions and reduced sensitivity to short-term wind-speed fluctuations.

Table 4 presents representative examples of predicted wind speeds and their corresponding observed values, demonstrating that the LSTM model can reasonably reproduce individual observations. Although these examples indicate relatively close agreement between predicted and actual wind speeds, the limited number of samples does not fully represent the model's overall predictive performance. Therefore, a more comprehensive assessment was conducted using the complete test dataset, as summarized in Table 3. The resulting Root Mean Square Error (RMSE) of 1.2965 knots and Mean Absolute Error (MAE) of 1.0175 knots indicate an average absolute prediction error of approximately 1 knot. These findings suggest that the model can capture general wind-speed variations, although discrepancies may remain during periods of rapid or substantial fluctuations.

Table 4. Representative comparison of observed and LSTM-predicted wind speeds in the test dataset, with values expressed in knots.

Actual	Prediction
4.625000	4.399096
5.000000	4.383305
4.041667	4.452825
4.250000	4.581909
4.125000	4.655954

3.3 Prediction Result

After identifying the best-performing LSTM configuration, wind-speed forecasts were generated for three prediction periods: 7, 14, and 31 days in January 2025. These forecasting periods were selected to examine the model's predictive behavior across different temporal horizons and assess its ability to reproduce observed wind-speed variations. The forecasting results were evaluated using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), with the corresponding performance metrics presented separately for each period. Comparing these results provides insights into how prediction errors vary across the evaluated periods. However, differences in the number of observations, evaluation dates, and prevailing meteorological conditions must be considered when

interpreting the results. A lower prediction error over a longer evaluation period does not necessarily indicate that longer-range forecasts are inherently more accurate, as performance may be influenced by the variability of observed wind speeds. Therefore, the reported metrics should be interpreted within their respective evaluation periods rather than as definitive evidence of superior long-term forecasting capability.

3.3.1 7-day prediction

Table 5 presents the predicted wind speeds for the first seven days of January 2025, with values ranging from approximately 3.77 to 4.43 knots. The forecast indicates a gradual decrease in wind speed throughout the evaluation period, with relatively small day-to-day variations. These changes suggest that the LSTM model predicts relatively stable wind conditions during the first week of January, without pronounced fluctuations or sudden increases in wind speed. The relatively smooth forecasting pattern may reflect the model's tendency to capture dominant temporal trends while attenuating short-term variability. However, the absence of substantial fluctuations in the predicted values does not necessarily indicate that extreme wind events did not occur during this period. A comparison with observed wind-speed measurements is required to determine whether the predicted variations accurately represent actual meteorological conditions. Therefore, these results should be interpreted as model-generated estimates rather than definitive evidence of stable atmospheric conditions.

Table 5. Daily wind-speed forecasts generated by the LSTM model for 1–7 January 2025, expressed in knots.

Date	Wind Speed Forecast (knots)
01/01/2025	4.4329133
02/01/2025	4.0891576
03/01/2025	4.1359143
04/01/2025	3.8746557
05/01/2025	3.8327928
06/01/2025	3.7942553
07/01/2025	3.777103

The table above shows the forecast for the first seven days of 2025, indicating that wind speeds will range from 3.77 to 4.43 knots. The daily variations are relatively small, ranging from 0.04 to 0.35 knots, suggesting that wind speed fluctuations during this period will not be extreme.

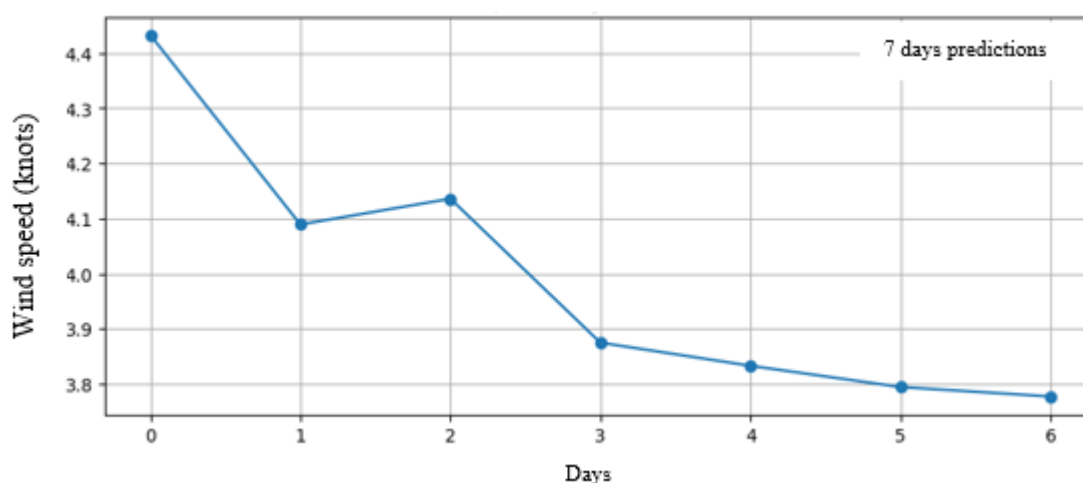


Figure 4. Predicted daily wind-speed variations for 1–7 January 2025 using the LSTM model, illustrating the overall decreasing trend and short-term fluctuations in forecast wind speed.

Table 5 and Figure 4 present the predicted wind speeds for 1–7 January 2025, with values ranging from approximately 3.8 to 4.4 knots. The forecast exhibits a gradual decrease in wind speed, indicating that the LSTM model produces relatively smooth temporal variations during the seven-day prediction period. However, the reliability of these predictions must be assessed by comparing the

forecast values with the corresponding observed wind speeds. As presented in Table 6, the model achieved a Root Mean Square Error (RMSE) of 1.4253 knots and a Mean Absolute Error (MAE) of 1.2261 knots. These results indicate that the predicted wind speeds differed from the observed values by approximately 1.23 knots on average, while the higher RMSE suggests the presence of relatively larger errors on certain days. Although the model captures general wind-speed patterns, these error values indicate that discrepancies remain between predicted and observed conditions. Therefore, the seven-day forecasting performance should be evaluated using both quantitative error metrics and the temporal agreement between paired observations and predictions, rather than relying solely on the smoothness of the forecast curve.

Table 6. Prediction errors of the LSTM model for the 7-day wind-speed forecast (1–7 January 2025), evaluated using RMSE and MAE, expressed in knots.

RMSE	1.4253
MAE	1.2261

The predictive performance of the LSTM model over the seven-day forecasting period was evaluated using two statistical metrics: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). As presented in Table 6, the model achieved an MAE of 1.2261 knots and an RMSE of 1.4253 knots. The MAE indicates that the predicted wind speeds deviated from the corresponding observed values by approximately 1.23 knots on average. Meanwhile, the higher RMSE reflects the greater influence of relatively large prediction errors on certain days. These results suggest that although the model captures general wind-speed variations, noticeable discrepancies remain between predicted and observed values. Therefore, further improvements in model configuration and validation may be necessary to enhance forecasting accuracy, particularly during periods of rapid wind-speed fluctuations.

Table 7. Daily wind-speed forecasts generated by the LSTM model for the 14-day period from 1 to 14 January 2025, expressed in knots.

Date	Wind Speed Forecast (knots)
01/01/2025	4.557855
02/01/2025	4.4687696
03/01/2025	4.445021
04/01/2025	4.1685915
05/01/2025	4.043037
06/01/2025	3.9710412
07/01/2025	3.929226
08/01/2025	3.8979506
09/01/2025	3.8744998
10/01/2025	3.8576355
11/01/2025	3.8463838
12/01/2025	3.8394656
13/01/2025	3.8358054
14/01/2025	3.8344173

3.3.2 14-day prediction

Table 7 and Figure 5 present the predicted wind speeds for the first 14 days of January 2025, indicating a gradual decrease from approximately 4.56 knots at the beginning of the forecasting period to 3.83 knots on the final day. The predicted wind-speed series exhibits a relatively smooth downward trend, characterized by progressively smaller day-to-day variations and an absence of pronounced fluctuations. This pattern suggests that the LSTM model generates increasingly stable predictions as the forecasting horizon extends. However, the absence of abrupt changes in the predicted values does not necessarily indicate that actual atmospheric conditions remained stable or that extreme wind events did not occur during this period. The progressively flatter forecast may reflect the model's tendency to smooth short-term variability, reduce sensitivity to rapid meteorological changes, or converge toward typical wind-speed values. Nevertheless, these possible explanations remain tentative because the

underlying forecasting behavior cannot be established from the predicted curve alone. Further examination of observed wind-speed variations would provide additional evidence regarding the model's ability to represent short-term atmospheric fluctuations.

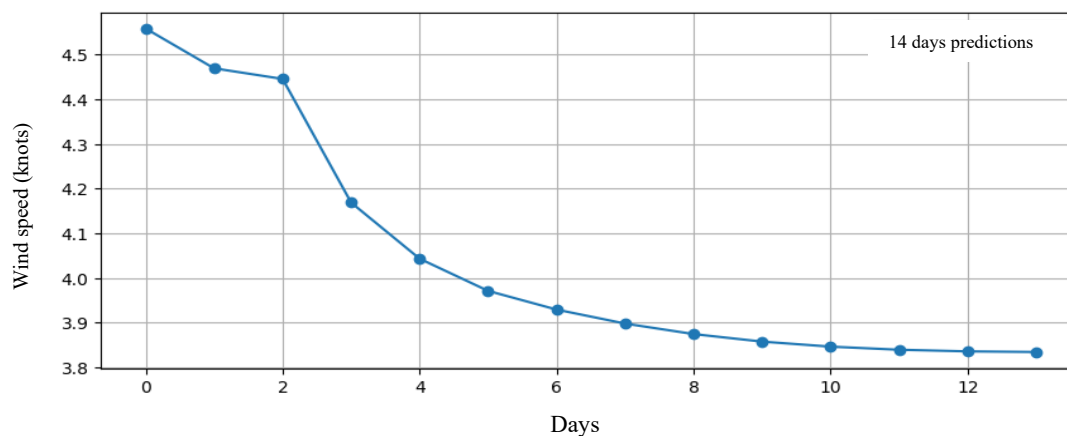


Figure 5. Predicted daily wind-speed variations generated by the LSTM model for 1–14 January 2025, illustrating a gradual decrease in wind speed followed by a tendency toward relatively stable predicted values.

The predictive performance of the 14-day forecast was quantitatively evaluated using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), as summarized in Table 8. The model achieved an RMSE of 1.3178 knots and an MAE of 1.0943 knots, indicating that the predicted wind speeds differed from their corresponding observed values by approximately 1.09 knots on average. The higher RMSE compared with MAE suggests that relatively large prediction errors occurred on certain days, contributing disproportionately to the overall evaluation. Although these results indicate that the LSTM model captures aspects of the observed wind-speed variability, the remaining prediction errors demonstrate limitations in reproducing individual daily fluctuations. Consequently, model performance should be assessed using both statistical error metrics and comparisons between predicted and observed temporal patterns. Such an evaluation is essential for determining forecasting reliability and identifying opportunities for further model improvement.

Table 8. Prediction errors of the LSTM model for the 14-day wind-speed forecast (1–14 January 2025), evaluated using RMSE and MAE, expressed in knots.

RMSE	1.3178
MAE	1.0943

3.3.3 31-day prediction

Table 9 and Figure 6 present the predicted wind speeds for the 31-day period in January 2025, with values ranging from approximately 3.36 to 4.39 knots. At the beginning of the forecasting period, the predicted wind speed reaches 4.39 knots on January 1, followed by a gradual decline to approximately 3.36 knots around January 13–15. Subsequently, the predicted values stabilize, exhibiting only minor fluctuations throughout the remaining days. This increasingly uniform pattern suggests that the LSTM model produces smoother predictions as the forecasting horizon extends. Such behavior may reflect the model's internal dynamics, the multi-step forecasting strategy, or a tendency to converge toward typical wind-speed values. However, the underlying mechanism cannot be established without further examination of the forecasting procedure and associated uncertainties.

The predictive performance of the 31-day forecast was evaluated using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), as summarized in Table 10. The model achieved an RMSE of 1.2117 knots and an MAE of 0.8886 knots, indicating an average absolute deviation of approximately 0.89 knots between predicted and observed wind speeds. These values represent the lowest reported prediction errors among the three evaluation periods. Nevertheless, this finding should be interpreted cautiously because the 7-, 14-, and 31-day evaluations involve different numbers of observations and potentially different meteorological conditions. Consequently, the smaller errors obtained over the

longer evaluation period do not necessarily demonstrate that increasing the forecasting horizon improves model accuracy.

Table 9. Daily wind-speed forecasts generated by the Long Short-Term Memory (LSTM) model for 1–31 January 2025, showing predicted wind-speed variations throughout the 31-day period, expressed in knots.

Date	Wind Speed Forecast (knots)
01/01/2025	4.3906183
02/01/2025	4.1608477
03/01/2025	4.0933785
04/01/2025	3.7455792
05/01/2025	3.639412
06/01/2025	3.5543554
07/01/2025	3.4986804
08/01/2025	3.4510653
09/01/2025	3.4174886
10/01/2025	3.3940547
11/01/2025	3.3787735
12/01/2025	3.3692133
13/01/2025	3.3638785
14/01/2025	3.361607
15/01/2025	3.3614764
16/01/2025	3.362767
17/01/2025	3.3650782
18/01/2025	3.3680508
19/01/2025	3.3713691
20/01/2025	3.374668
21/01/2025	3.3778377
22/01/2025	3.3807063
23/01/2025	3.3834538
24/01/2025	3.386031
25/01/2025	3.388338
26/01/2025	3.39019
27/01/2025	3.3915088
28/01/2025	3.3927488
29/01/2025	3.3940563
30/01/2025	3.3949683
31/01/2025	3.3957891

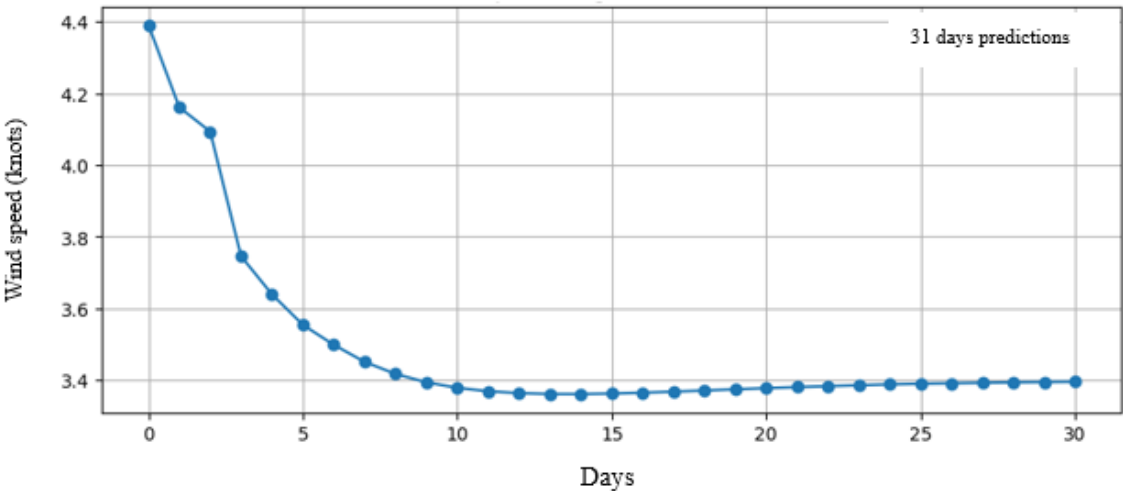


Figure 6. Predicted daily wind-speed variations generated by the Long Short-Term Memory (LSTM) model for 1–31 January 2025, illustrating an initial decline in wind speed followed by a gradual stabilization of predicted values toward the end of the forecasting period.

Overall, the results indicate that the LSTM model can reproduce general wind-speed patterns during January 2025, although its predictions become increasingly smooth over time. A more rigorous assessment would require consistent forecast initialization, evaluation over comparable verification periods, and repeated rolling-origin testing. Furthermore, quantifying forecast uncertainty and examining errors at individual lead times would provide stronger evidence of the model's reliability and limitations for extended wind-speed forecasting.

Table 10. Prediction errors of the Long Short-Term Memory (LSTM) model for the 31-day wind-speed forecast (1–31 January 2025), evaluated using RMSE and MAE, expressed in knots.

RMSE	1.2117
MAE	0.8886

3.4 SPSS regression test

3.4.1 F-test

If $\text{Sig.} < 0.05$, then the regression model as a whole is significant, which means that at least one independent variable affects the dependent variable. If $\text{Sig.} > 0.05$, then the regression model as a whole is not significant, which means that the independent variables collectively do not affect the dependent variable. If $f_{\text{calculated}} > f_{\text{table}}$, then the regression model is significant. If $f_{\text{calculated}} < f_{\text{table}}$, then the regression model is not significant.

Table 11. Analysis of variance (ANOVA) results for the multiple linear regression model examining the combined association of air temperature, relative humidity, atmospheric pressure, and rainfall with observed wind speed.

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2095.757	4	523.939	209.574	.000 ^b
	Residual	9107.557	3643	2.500		
	Total	11203.315	3647			

a. Dependent Variable: WIND SPEED

Predictors: (Constant), RAINFALL, AIR TEMPERATURE, AIR PRESSURE, HUMIDITY

Because $f_{\text{calculated}} > f_{\text{table}}$, the four independent variables together affect wind speed. Based on the p-value results of the F test, it shows that $p\text{-value} < 0.001$, so it can be concluded that temperature, humidity, air pressure, and rainfall together affect wind speed.

3.4.2 T-test

If the Sig value is < 0.05 , then the independent variable significantly affects the dependent variable. If the Sig value is > 0.05 , then the independent variable does not significantly affect the dependent variable. If the $t_{\text{calculated}} > t_{\text{table}}$, then the independent variable is significant. If the $t_{\text{calculated}} < t_{\text{table}}$, then the independent variable is not significant.

The results of the value substitution from SPSS can be seen as follows:

$$Y = -145.518 - 0.194X_1 - 0.113X_2 + 0.163X_3 + 0.003X_4 + e$$

Constant (-145.518), meaning that if the temperature, humidity, air pressure, and rainfall are 0, then the predicted wind speed is -145.518 knots. This negative value means that the constant is only the starting point of the equation. Air Temperature (X_1 , coefficient - 0.194, sig 0.000), meaning that for every 1°C increase in air temperature, wind speed will decrease by 0.194 knots. Humidity (X_2 , coefficient - 0.113, sig 0.000), meaning that for every 1% increase in humidity, wind speed will decrease by 0.113 knots. Air pressure (X_3 , coefficient + 0.163, sig 0.000), meaning that for every 1 hPa increase in air pressure, wind speed will increase by 0.163 knots. Rainfall (X_4 , coefficient + 0.003, sig 0.164),

meaning that for every 1 mm increase in rainfall, wind speed will increase by 0.003 knots. However, this is not significant because Sig. > 0.05.

Table 12. Multiple linear regression coefficients and individual significance tests for the relationships between meteorological predictors and wind speed, including unstandardized coefficients, standardized coefficients, t-statistics, and p-values.

Coefficients ^a		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
Model		B	Std. Error	Beta		
1	(Constant)	-145.518	19.563		-7.438	.000
	AIR	-.194	.030	-.120	-6.398	.000
	TEMPERATURE (°C)					
	HUMIDITY (%)	-.113	.005	-.413	-20.634	.000
	AIR PRESSURE (hPa)	.163	.019	.143	8.604	.000
	RAINFALL(mm)	.003	.002	.022	1.392	.164

Accurate wind speed prediction is crucial for supporting port operations, maritime safety, and renewable energy development in coastal regions such as Tanjung Perak, Surabaya. This study employs the Long Short-Term Memory (LSTM) deep learning method to forecast wind speed using meteorological variables including temperature, rainfall, air pressure, and air humidity as input features. The LSTM model is capable of capturing nonlinear and temporal dependencies within atmospheric data, making it well-suited for time series prediction of climatic parameters. Understanding the complex interactions among atmospheric variables is essential, as rainfall and temperature fluctuations often coincide with changes in wind dynamics and pressure systems (Aminoto, 2024; Aminoto & Faqih, 2024; Aminoto et al., 2024). Previous studies utilizing climate model outputs from CORDEX-SEA and ERA5 datasets have demonstrated the importance of high-resolution data in analyzing spatial and temporal variability of regional climate phenomena. Building upon these findings, this research aims to enhance the accuracy of short-term wind speed prediction and provide valuable insights for optimizing port operations, navigation planning, and local weather forecasting in the Tanjung Perak area.

4. CONCLUSION

Among the LSTM configurations reported in Table 3, 50 units and 50 epochs with an 80:20 training–test partition produced the lowest test errors (RMSE = 1.2965 knots; MAE = 1.0175 knots). For the specific January 2025 intervals examined, the reported errors were lowest for the 31-day period (RMSE = 1.2117 knots; MAE = 0.8886 knots). Because these periods differ in length, this finding does not demonstrate that longer forecast horizons are generally more accurate. Multiple linear regression indicated significant conditional associations of wind speed with temperature, humidity, and air pressure ($p < 0.001$), whereas rainfall was not significant ($p = 0.164$). Further work should evaluate chronological data splitting, information leakage, sensitivity to forecast lead time, and comparisons with persistence or other baseline models before operational use.

ACKNOWLEDGMENTS

The authors thank the Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG), particularly the Tanjung Perak Surabaya Maritime Meteorological Station, for providing the meteorological observations used in this study.

AUTHOR CONTRIBUTIONS

Conceptualization, Y.M.S., T.A., and H.H.; methodology, Y.M.S. and T.A.; software, Y.M.S.; validation, Y.M.S., T.A., and H.H.; formal analysis, Y.M.S.; investigation, Y.M.S.; resources, Y.M.S.;

data curation, Y.M.S.; writing—original draft preparation, Y.M.S.; writing—review and editing, Y.M.S., T.A., and H.H.; visualization, Y.M.S.; supervision, T.A. and H.H.; project administration, Y.M.S. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research received no external funding.

DATA AVAILABILITY STATEMENT

The meteorological observations analyzed in this study were obtained from the Tanjung Perak Surabaya Maritime Meteorological Station of the Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG) for 2015–2024. Access to the original observations is subject to applicable BMKG data-access and sharing policies. Specify whether the processed dataset and analysis code can be provided by the corresponding author upon reasonable request or are deposited in a repository.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

ARTIFICIAL INTELLIGENCE (AI) USAGE DECLARATION

During the preparation of this manuscript, OpenAI ChatGPT was used to assist with English-language proofreading, grammatical correction, and clarity of academic expression.

REFERENCE

- Akbar, R., Santoso, R., & Warsito, B. (2023). Prediksi tingkat temperatur Kota Semarang menggunakan metode long short-term memory (LSTM). *Jurnal Gaussian*, 11(4), 572–579. <https://doi.org/10.14710/j.gauss.11.4.572-579>.
- Aminoto, T. (2024). *Performa Model Hujan CORDEX-SEA pada Aspek Spatial-Temporal dan Implementasi Faktor Bobot pada Rata-rata Ansambel*. [Doctoral dissertation, Institut Pertanian Bogor]. IPB Repository.
- Aminoto, T., & Faqih, A. (2024). Tren curah hujan di asia tenggara berdasarkan model CORDEX-SEA dan data ERA5. *Journal Online of Physics*, 10(1), 7–13. <https://doi.org/10.22437/jop.v10i1.37499>.
- Aminoto, T., Faqih, A., Koesmaryono, Y., & Dasanto, B. D. (2024). A Comparison of the Performance of the Weighted Ensembles Means in CORDEXSEA Precipitation Simulations. *Agromet*, 38(1), 19-35. <https://doi.org/10.29244/j.agromet.38.1.19-35>.
- Anwar, S. (2017). Peramalan Suhu Udara Jangka Pendek di Kota Banda Aceh dengan Metode Autoregressive Integrated Moving Average (ARIMA). *Malikussaleh Journal of Mechanical Science and Technology*, 5(1), 6–12. <https://doi.org/10.29103/mjmst.v5i1.10882>.
- Catursari, S. A. (2024). *Analisis pengaruh tekanan udara, kelembaban udara, dan suhu udara, terhadap curah hujan di Kota Surabaya*. UIN Maulana Malik Ibrahim. *Undergraduate thesis, Universitas Islam Negeri Maulana Malik Ibrahim Malang*. Etheses UIN Malang.
- 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 (Jurnal Edukasi Dan Penelitian Informatika)*, 8(3), 405–411. <https://doi.org/10.26418/jp.v8i3.56606>.
- Friadi, R., & Junadhi. (2019). Sistem kontrol intensitas cahaya, suhu dan kelembaban udara pada greenhouse berbasis raspberry PI. *JTIS*, 2(1), 30-37. <https://doi.org/10.36085/jtis.v2i1.217>.
- 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>.
- Luthfiarta, A., Febriyanto, A., Lestiawan, H., & Wicaksono, W. (2020). Analisa Prakiraan Cuaca dengan Parameter Suhu, Kelembaban, Tekanan Udara, dan Kecepatan Angin Menggunakan Regresi Linear Berganda. *JOINS (Journal of Information System)*, 5(1), 10–17. <https://doi.org/10.33633/joins.v5i1.2760>.

- Muflih, G. Z., Sunardi, S., & Yudhana, A. (2019). Jaringan Saraf Tiruan Backpropagation untuk Prediksi Curah Hujan di Wilayah Kabupaten Wonosobo. *MUST: Journal of Mathematics Education, Science and Technology*, 4(1), 45–56. <https://doi.org/10.30651/must.v4i1.2670>.
- Novrita, R. R., Sudarti, & Yushardi. (2021). The Analisis Potensi Energi Angin Tambak Untuk Menghasilkan Energi Listrik: Analisis Potensi Energi Angin Tambak Untuk Menghasilkan Energi Listrik. *Journal of Research and Education Chemistry*, 3(2), 96–112. [https://doi.org/10.25299/jrec.2021.vol3\(2\).7165](https://doi.org/10.25299/jrec.2021.vol3(2).7165).
- Prasita, V. D., Rosana, N., Agustina, N. A., & Kusuma, A. (2023). *Pemanfaatan Sumberdaya Lahan Pesisir Berbasis Daya Dukung Lingkungan Dalam Menghadapi Perubahan Iklim Global*. Hang Tuah University Press
- Prayogo, B., Forda Nama, G., & Muhammad, M. A. (2021). Rancang Bangun Prototipe Sistem Monitoring Mini Stasiun Cuaca pada BMKG Provinsi Lampung. *Jurnal Informatika Dan Teknik Elektro Terapan*, 9(1). <https://doi.org/https://doi.org/10.23960/jitet.v9i1.2265>.
- Putri, F. A., Sudarti, & Yushardi. (2023). Analisis Kecepatan Angin Di Kota Surabaya Berbantuan Database Online-Bmkg. *Jurnal Teknologi Pendidikan Dan Pembelajaran*, 1(2), 217–223.
- Putri, M. H. (2024). Faktor tekanan dan kelembapan udara terhadap cuaca ekstrem di Surabaya pada Oktober 2023. *Maliki Interdisciplinary Journal*, 2(12), 1384–1392.
- Sunarmi, N., Kumailia, E. N., Nurfaiza, N., Nikmah, A. K., Aisyah, H. N., Sriwahyuni, I., & Lailly, S. N. (2022). Analisis faktor unsur cuaca terhadap perubahan iklim di Kabupaten Pasuruan pada tahun 2021 dengan metode Principal Component Analysis. *Newton-Maxwell Journal of Physics*, 3(2), 56–64. <https://doi.org/10.33369/nmj.v3i2.23380>.
- Supriyadi, E. (2021). *Prediksi Parameter Cuaca Menggunakan Deep Learning Long-Short Term Memory (Lstm)*. *Jurnal Meteorologi dan Geofisika*, 21(2), 55–67. <https://doi.org/10.31172/jmg.v21i2.619>.
- Tanjung, R., Listiani, A., & Lestari, F. (2024). Prediksi Multivariate Time Series Parameter Cuaca Menggunakan Long Short-Term Memory (LSTM). *Prosiding Seminar Nasional Sains Data*, 4(1), 445–456. <https://doi.org/10.33005/senada.v4i1.253>.
- Wiranda, L., & Sadikin, M. (2019). Penerapan long short term memory pada data time series untuk memprediksi penjualan produk PT. Metiska Farma. *Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI*, 8(3), 184–196. <https://doi.org/10.23887/janapati.v8i3.19139>.
- Yuliara, I. M. (2016). *Regresi linier berganda*. Denpasar: Universitas Udayana.