

Time Series Forecasting for Average Temperature with the Long Short-Term Memory Network in Deli Serdang Geophysics Station

Nora Valencia Sinaga¹, Feriomex Hutagalung², Martha Manurung³, Eva Darnila⁴

^{1, 2, 3}Indonesian Agency for Meteorological, Climatological and Geophysics

⁴Department of Informatics, Universitas Malikussaleh, Lhokseumawe

Article Info

Article history:

Received September 15, 2021

Revised September 25, 2021

Accepted October 01, 2021

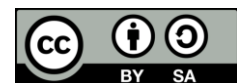
Keywords:

Forecasting,
LSTM,
Average Temperature

ABSTRACT

An understanding of trends analysis, and prediction of time series of average temperature as one of parameter weather and climate data for climate variables. It is the central process in assessing the state of the climate of a region and provides an overall estimate about the variations in the climate variables. Explore weather trends using normal and local yearly average temperatures, compare and make observations. In this study, we try to analyze local and normal average temperature data in Deli Serdang geophysic Station based on observation station in situ. The main goal of this study to compare the normal temperature to local station and to predict the average temperature data in BMKG Geophysics Station, Deli Serdang, North Sumatra using Long Short-Term Memory Model (LSTM). Based on the result of normal data science of exploring temperature with local temperature correlation, we got the display of training curve, residual plot and the scatter plot are shown using these codes. Based on the temperature series data from Geophysic station, the MSE value is 0.83 and the R2 value is 0.86.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponden Author:

Martha R Manurung
Badan Meteorologi Klimatologi dan Geofisika, Indonesia
Email: naficero7@yahoo.com

1. INTRODUCTION

Climate change is currently one of humanity's most pressing issues, owing to the environmental consequences of increased greenhouse gas emissions, which are the result of human activity. Temperature is one of the most important meteorological variables that is linked to climatic events and used to describe the state of the atmosphere. The greenhouse effect's increase in global temperature is the primary driver of rising sea levels, decreasing snow and ice cover, and changing water precipitation trends. The creation of mathematical physical models to obtain patterns that allow for the prediction of climatic variables has led to the development of mathematical physical models to assist government and institutions in being prepared to avoid economic and human losses. Rising temperatures in Deli Serdang have had an impact on agriculture, health, and food security.

Temperature is a critical element in all aspects of climate change, and it is the most important weather factor influencing fire behavior. Heat waves are more likely to occur more frequently and persist longer when temperatures rise. Warmer temperatures can potentially trigger a cascade of additional global changes [1], [2]. Because rising air temperatures affect the oceans, weather patterns, snow and ice, as well as plants and animals, it's a win-win situation. We evaluated the temperature average timeseries gathered from single station in BMKG Geophysics station, Deli Serdang North Sumatra, in this study. The primary purpose of this research is to examine intermediate variables, evaluate, and forecast temperature average timeseries from 2008 to 2020. If the time series were available, the LSTM model with/without intermediate variable performed better.

In general, the findings of this study will show how average temperature can be validated and predicted as a parameter in climate data in the Medan area based on each station, and can be used as a reference in further analysis of climate data sources that can be monitored by network climate in the Medan area based

on the local observation that is the source. The purpose of this study is to compare the normal temperature to local station and to predict the average temperature data in BMKG Geophysics Station, Deli Serdang, North Sumatra using Long Short-Term Memory Model.

2. DATA AND METHOD

We used temperature data from the BMKG Geophysics Station, Deli Serdang climate library for this investigation. The temperature data was collected from 2008 to 2020, using the average temperature for Deli Serdang as the category. The temperature series data selection gives a solid foundation for analyzing the performance of observed that have been installed and examined in earlier reports. Table 1 shows a list of the temperatures recorded at each site.

Table 1. List of the Station in BMKG Geophysics Station

No	Latitude	Longitude	Code Station
1	3.5	98.56	TSI

As in a prior study [3]–[5], we focused on analyzing temperature data as time series recorded from each station in the Medan area and using LSTM to train, test, and forecast the time series.

Figure 1, to summarize, the LSTM cell's input is a time series set of data x that runs through many sigmoid activation gates σ . To determine the cell states, each gate calculates a certain function. We merely gave a very brief overview of how LSTM works. It is still much better to take a deep learning program to learn more about LSTM. The input gate controls the flow of input activations into the memory cell, the output gate controls the cell activation flow, and the forget gate filters the information from the input and previous output and chooses which one should be remembered or not. Aside from the three gates, the LSTM cell has a cell update layer, which is normally part of the cell state. Three variables enter each LSTM cell: the current input x_t , the previous output h_{t-1} , and the previous cell state c_{t-1} . On the other hand, two variables emerge from each LSTM cell: the current output h_t and the current cell state c_t .

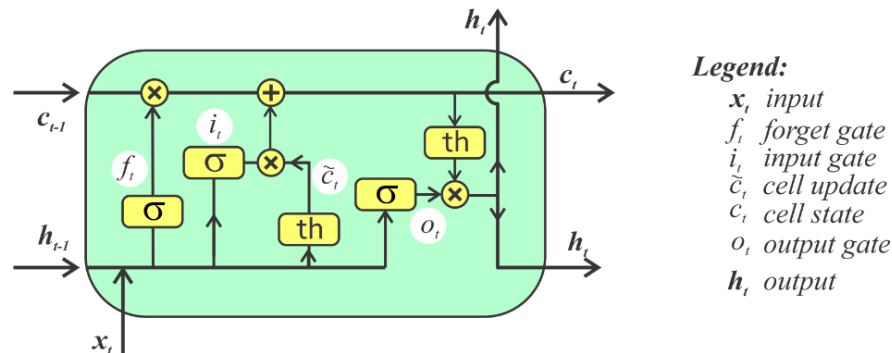


Fig. 1. The LSTM Model [6]

The current output h_t computation is the LSTM cell's final stage. The current output is calculated by multiplying the output gate layer and the tanh layer of the current cell state C_t by o_t times. The current output h_t has traveled through the network as either the prior state for the next LSTM cell or the input for the neural network output layer. The concepts of mean squared error and R-squared will be covered in this section. The average of the sum of squared differences between the actual value and the projected or estimated value is the mean squared error (MSE). Mean squared deviation is another name for it (MSD). This is how it is mathematically represented [7].

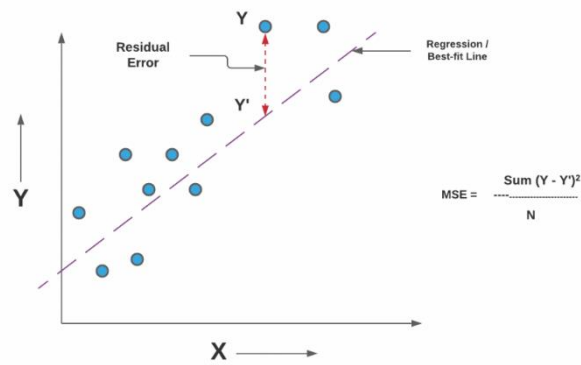


Fig. 2. Mean Squared Error Representation [7]

MSE is always positive or larger than zero in value. A number around 0 indicates that the estimator / predictor is of higher quality (regression model). The fact that the predictor has an MSE of zero (0) indicates that it is a perfect predictor. When the MSE value is squared, the result is root mean squared error (RMSE). The actual value is represented by Y, while the anticipated value is represented by Y' in the above equation. The diagrammatic representation of MSE [7] is shown below.

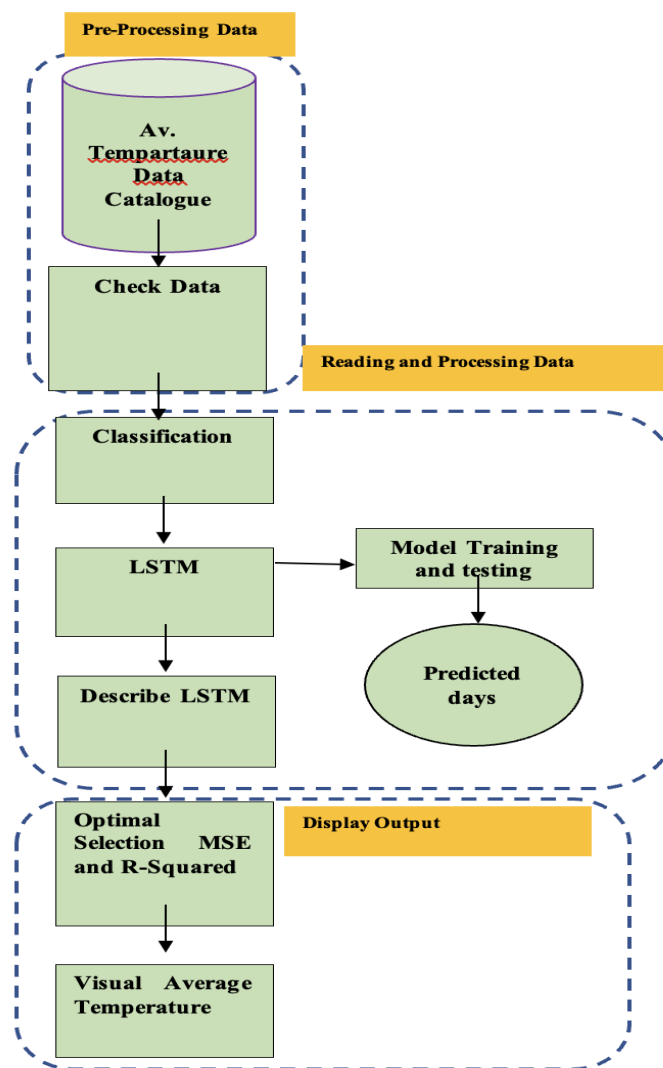


Fig. 3. Forecasting Average Temperature Flowchart

3. RESULT AND DISCUSSION

The original average temperature data was computed to ensure that our model prediction was correct. However, the projected (n days) shows that the inaccuracies are frequently caused by an unexpected surge or drop in the data, as seen in days 350-360. However, the model can correctly follow the data trend based on the first 75 days. Figure 4 depicts a graphic representation of all data from synoptic BMKG Geophysics Deli Serdang (Id Station 96037). Based on the temperature series data from Geophysic station, the MSE value is 0.83 and the R2 value is 0.86.

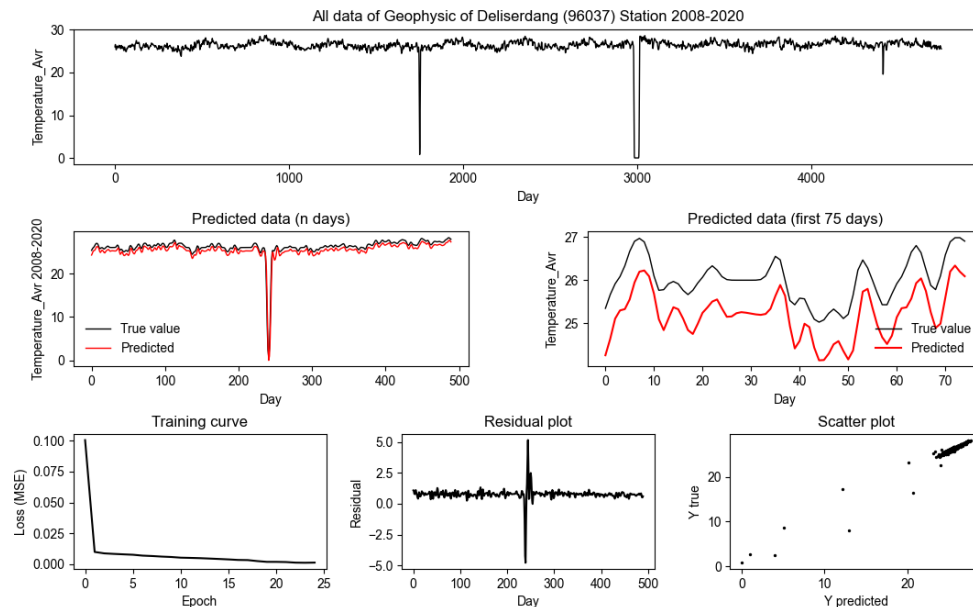


Fig. 4. Stacked LSTM prediction results in 96037 Station

The properties of temperature in the 2008-2020 timeframe are separated into MSE and R2 values using the LSTM technique. The MSE value of 96037 is 0.83, and the R2 value is 0.83, as shown in Figure 4.

4. CONCLUSION

We discovered that LSTM is an excellent method for forecasting average temperature data in this investigation. Based on the representation of BMKG Geophysics station, a number not too close to zero indicates that the estimator / predictor is of low quality (regression model). We need to evaluate the available average temperature series. However, we can see from this that there are a few things to remember as LSTM lessons. To begin with, having more input days does not necessarily imply that the model will be more accurate. Aside from that, temperature data conditioning could help improve the model's accuracy. Finally, even though we haven't demonstrated it, LSTM requires a specific amount of data to work. From there, we may envisage LSTM being utilized to forecast weather and average temperature trends.

REFERENCE

- [1] Wikipedia, "Effects of climate change - Wikipedia," https://en.wikipedia.org/wiki/Effects_of_climate_change, 2021. https://en.wikipedia.org/wiki/Effects_of_climate_change (accessed Mar. 10, 2021).
- [2] NASA, "Global Warming," <https://earthobservatory.nasa.gov/features/GlobalWarming>, 2010. <https://earthobservatory.nasa.gov/features/GlobalWarming> (accessed Mar. 10, 2021).
- [3] S. Afshin, H. Fahmi, A. Alizadeh, H. Sedghi, and F. Kaveh, "Long term rainfall forecasting by integrated artificial neural network-fuzzy logic-wavelet model in karoon basin," *Sci. Res. Essays*, vol. 6, no. 6, pp. 1200–1208, 2011, doi: 10.5897/SRE10.448.
- [4] M. I. Hutapea, Y. Y. Pratiwi, I. M. Sarkis, I. K. Jaya, and M. Sinambela, "Prediction of relative humidity based on long short-term memory network," *AIP Conf. Proc.*, vol. 2221, no. March, 2020, doi: 10.1063/5.0003171.
- [5] A. G. Salman, Y. Heryadi, E. Abdurahman, and W. Suparta, "Weather forecasting using merged Long Short-Term Memory Model (LSTM) and Autoregressive Integrated Moving Average (ARIMA) Model," *J. Comput. Sci.*, vol. 14, no. 7, pp. 930–938, 2018, doi: 10.3844/jcssp.2018.930.938.
- [6] Stanford, "Understanding LSTM Networks," https://web.stanford.edu/class/cs379c/archive/2018/class_messages_listing/content/Artificial_Neural_Network_Technology_Tutorials/OlahLSTM-NEURAL-NETWORK-TUTORIAL-15.pdf, pp. 1–13, 2015, [Online]. Available: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>.

-
- [7] A. Kumar, "Mean Squared Error or R-Squared, Data Analytics," <https://vitalflux.com/mean-square-error-r-squared-which-one-to-use/>, 2020. <https://vitalflux.com/mean-square-error-r-squared-which-one-to-use/> (accessed Mar. 10, 2021).