

Modeling and Forecasting Daily Temperature in Bandung

Solichatus Zahroh, Yuyun Hidayat, Resa Septiani Pontoh

Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Indonesia
solichatuszahroh@gmail.com, yuyun.hidayat@unpad.ac.id, resa.septiani@unpad.ac.id

Agus Santoso

Department of Statistics, Faculty of Sciences and Technology, Universitas Terbuka, Indonesia
aguss@ecampus.ut.ac.id

Sukono

Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Indonesia
sukono@unpad.ac.id

Abdul Talib Bon

Department of Production and Operations, University Tun Hussein Onn Malaysia, Malaysia
talibon@gmail.com

Abstract

Temperature is a very important weather element whose value can change at any time. Forecasting daily temperature is important information for the society, so forecasting with high accuracy is required. The use of neural network methods in forecasting has become increasingly popular lately. Therefore, this study uses the Long Short Term Memory (LSTM) model to forecast temperatures in Bandung. This method doesn't require parametric assumptions and can be used for data with long time periods, as there is a cell state to overcome vanishing gradients in the Recurrent Neural Network method. This method is used because the data to be inputted is daily historical data for 5 years. The temperature allows for unexpected fluctuations, so that earlier data is expected to detect motion patterns that weren't executed by the most recent data. Based on the results of the LSTM analysis, the Mean Absolute Percentage Error (MAPE) for the minimum temperature prediction is 5.61% and MAPE for the maximum temperature prediction is 3.84%. Thus, forecasting temperatures using LSTM can provide high accuracy results for short-term prediction using data with a long period of time.

Keywords

Temperature, Bandung, Forecasting, Long Short Term Memory

1. Introduction

Of course, human life is very much influenced by the physical state of the environment and also by the climatic conditions. Attention to weather and climate information is increasing with the growth of natural phenomena, not uncommonly or commonly referred to as extreme weather (Dewantara 2012). An extreme weather that has occurred lately is the increase in temperature during the day and the decrease in temperature at night. The city of Bandung is located at 06° 57' LS 107° 37' BT is located in a country with a tropical climate, so that it receives the sun all year round, which affects the high temperatures. Figure 1 shows the maximum and minimum temperature curves of the last 5 years. The data curve shows that the data is stationary. However, temperature is a weather element whose value fluctuates rapidly (Hafni et al., 2015), so data may not be stationary in the future.

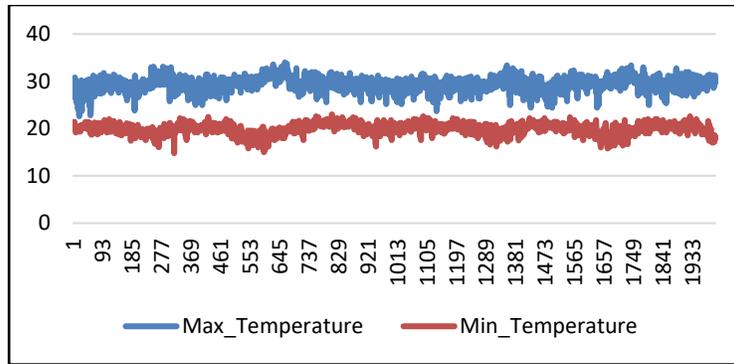


Figure 1. Daily Maximum and Minimum Temperature Charts (January 2014 - June 2019) BMKG Bandung

In the field of prediction, neural networks are becoming increasingly popular due to their practicability (Sesa et al., 2017). In addition, temperature values can fluctuate unexpectedly, so there are cases where old data can help the model identify trends and general movements that do not match the latest data (Zaytar and Amrani 2016). The method that can be predicted using sequence data over a long period of time is Long Short Term Memory. Moreover, the Long Short Term Memory has been successfully applied in various fields, especially time sequence problem, such as stock market prediction (Lai et al., 2019), flood forecasting (Le et al., 2019), sales forecasting (Helmini et al., 2019), travel time prediction (Duan et al., 2016), and weather prediction (Kumar et al., 2018). Temperature forecasts using other methods have previously been carried out for several regions in Indonesia. Long Memory models to forecasting temperatur in Karimun, Kepulauan Riau (Caraka et al., 2016), forecast of air temperature in Yogyakarta using fuzzy model (Priyana and Abadi, 2011), prediction of short-term air temperature in Banda Aceh using autoregressive integrated moving average method (Anwar, 2017), and comparative analysis of Singular Spectrum Analysis and Fourier Series Analysis on air temperature data in Surabaya (Christienova et al., 2018). The main objective of this research is to obtain a prediction model for daily temperature in Bandung using a neural network, the accuracy will be compared to the Numeric Weather Prediction method based on the latest data and information from observations and satellite monitoring.

2. Methodology

The commonly used forecasting method is the ARIMA model. However, the ARIMA model may have high sensitivity to extremes for long periods of time. Experiments are needed to explain the variability and extrapolation of seasonal factors as ARIMA equations when the parameters change due to time shifts (Fromm 1978). In ARIMA, overfitting may occur if too many variables are included in the selected model or complex nonlinear models are used to estimate linear relationships or simple non-linear relationships (Liu 2000). In other words, there is an overfitting because there is a dependency on too much training data (Koehrsen 2018). A popular way to overcome this is the Artificial Neural Network (ANN), which is believed to have high accuracy, was introduced by McCulloch and Pits (1943). ANN is divided into two groups, Feedforward Neural Network (FFNN) and Recurrent Neural Network (RNN) (Puspitaningrum 2006). In the RNN, there is a vanishing gradient problem during the backpropagation process as training data periods increase (Ashrovy 2017). Vanishing gradient is a situation in which the gradient value used for weight renewal of neurons during an RNN training session is 0 or near 0. This is because using an activation function that maps input to output affects the sensitivity of the neuron layer in the cell weight renewal process. So if the gap between current and past data is created, RNN can not connect the input. This is commonly referred to as long-term dependence. This proves that RNN is not suitable for long-term data patterns, so another method is required, Long Short Term Memory (LSTM).

2.1 Long Short Term Memory

The LSTM method was introduced by Hochreiter and Schmidhuber (1997). LSTM is a variant of RNN that solves problems in RNN by adding cell states or memory cells with constant errors so that errors can be reproduced without vanishing gradients. There are three gates in the LSTM, namely an input gate, a forget gate and an output gate. Before modeling, the data is normalized using the min-max scaling technique.

$$X'_i = \left[\frac{(x_i - x_{min})}{x_{max} - x_{min}} \right]$$

where X'_i is input data after normalization, x_i is input data, x_{min} is minimum value of the data, and x_{max} is maximum value of the data.

Hereafter, the data is divided into training and testing data. The size of the test data is usually about 20% of the total data, although this value depends on the amount of data and the size of the predictions (Hyndman and Athanasopoulos 2018). Training data is used for modeling and test data for model validation. The determination of parameters, such as learning rate, batch size, number of hidden layers, number of neurons in the hidden layer, and maximum epoch may be random or algorithmic. Each gate function is calculated and executed until the maximum epoch or expected error target is reached.

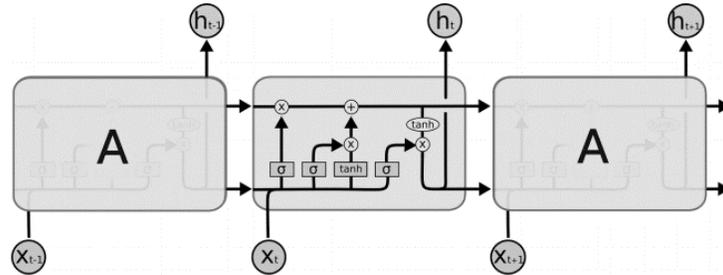


Figure 2. Structure of Long Short Term Memory. Reproduced from Olah, 2015

Figure 2 shows the structure of the LSTM. The first gate is a forget gate with a sigmoid activation function, which is a gate that determines whether information from the current input (x_t) and the previous output (h_{t-1}) is appropriate or not.

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

where σ is sigmoid activation function, W_f is weight of the forget gate, h_{t-1} is hidden state output at time $t-1$, x_t is current input at time t , and b_f is bias of the forget gate. The value of the forget gate is performed by the multiplication operation with the previous cell state. The output of this gate is 0 to 1 because the activation function used is sigmoid. If the value of f_t is close to 0, the information is removed from the previous cell, whereas if the value is close to 1, the information is passed.

Besides the forget gate, there is an input gate (a layer of neurons with a sigmoid activation function and also a tanh activation function). This gate determines which part is renewed and after passing through the sigmoid layer and determines the degree of importance after passing through the tanh layer.

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

where σ is sigmoid activation function, W_i is weight of the input gate, h_{t-1} is hidden state output at time $t-1$, x_t is current input at time t , and b_i is bias of the input gate. The value of the input gate is performed by the multiplication operation with the intermediate cell state (\tilde{C}_t).

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$

where \tilde{C}_t is intermediate cell state, \tanh is tanh activation function, W_c is weight of the cell state, h_{t-1} is hidden state output at time $t-1$, x_t is current input at time t , and b_c is bias of the cell state. The previous cell state which is performed by the multiplication operation with the forget gate is updated with the output of the intermediate cell state by the multiplication operation, so that a new cell state is obtained.

$$c_t = (i_t * \tilde{C}_t + f_t * c_{t-1})$$

The last gate is the output gate (a layer of neurons with sigmoid activation function). This gate determines how much information that is output by the cell.

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

where σ is sigmoid activation function, W_o is weight of the output gate, h_{t-1} is hidden state output at time $t-1$, x_t is current input at time t , dan b_o is bias of the output gate. Cell state with tanh activation function combined with the output gate with the sigmoid activation function by the multiplication operation, which then becomes the hidden state output.

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

Input gate, output gate, and forget gate has independent weight and bias (Olah 2015). Hence, the network will learn how much of the last output that must remain stored, how much current inputs to be saved, and how much cell states must be sent to the output.

2.2 Model Evaluation Criteria

After obtaining a predictive model from the training process, the predictive model is evaluated by testing data. Prediction models are tested to obtain the value of the accuracy of the predictions of the model. In general, model selection criteria that are commonly used is the Mean Absolute Percentage Error (MAPE). A model becomes the best model if it has the smallest MAPE value. MAPE is a measure that measures the difference between the prediction value and the actual value in percent.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{X_t - \widehat{X}_t}{X_t} \right| \times 100\%$$

where n is the number of observations, X_t is the actual value, and \widehat{X}_t is the prediction value.

3. Result and Discussion

The data used are daily maximum and minimum temperature data from 1 January 2014 to 30 June 2019. The dataset used is secondary data obtained in 2019 from the Meteorology, Climatology and Geophysics Agency (BMKG) Bandung. The data was used in total is 2007, where divided into training and testing data with a ratio of 80:20 to obtain 1605 training data and 402 testing data.

This study uses the Python software in modeling and imported Numpy, Matplotlib, and Pandas libraries to process, manage, and visualize data. The model shown in Figure 2 is based on Tensorflow and Keras, which were developed for machine learning studies.

Several training options and parameters of the model were changed to select the best results. The parameters of the neuron network were also updated depending on a loss function of an iteration step.

Table 1. LSTM Model Setup.

Parameters	Detail
Learning rate (optimizer: adam)	0.0001 (default)
Batch size	1
Number of hidden layer	1
Number of neuron in hidden layer	4
Number of epoch	50

The MAPE was used as a score function to evaluate the quality of the predictions. The scores were calculated using the test data with the following results as an example

Table 2. The result of the LSTM Model

Number of epoch	MAPE of Testing	
	Max_Temp	Min_Temp
10	3.84%	5.61%
20	3.97%	5.70%
30	4.00%	5.81%
40	3.99%	5.82%
50	3.99%	5.81%

The following graph shows the loss function (mean squared error) for each epoch change (the smallest MAPE). From the Figure 3, we can see that the loss function for minimum temperature is approximately to 0.0225 and the loss function function for maximum temperature is approximately to 0.015.

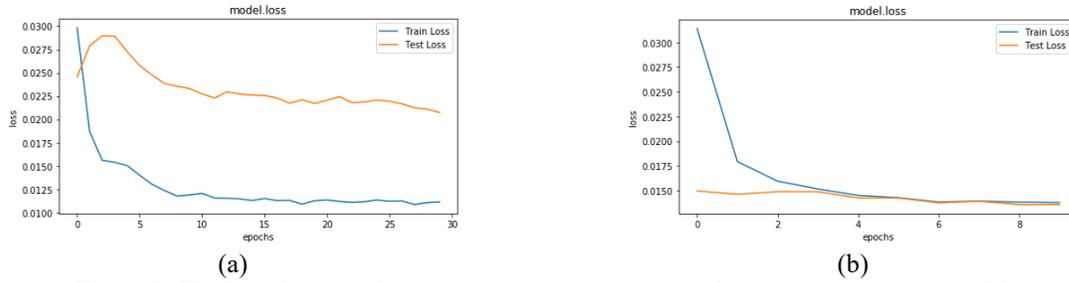


Figure 3. The loss function for minimum temperature (a) and maximum temperature (b)

Figure 4 shows the actual and predicted data for 10 epochs (the smallest MAPE), the data used in the graph is the testing data for model evaluation.

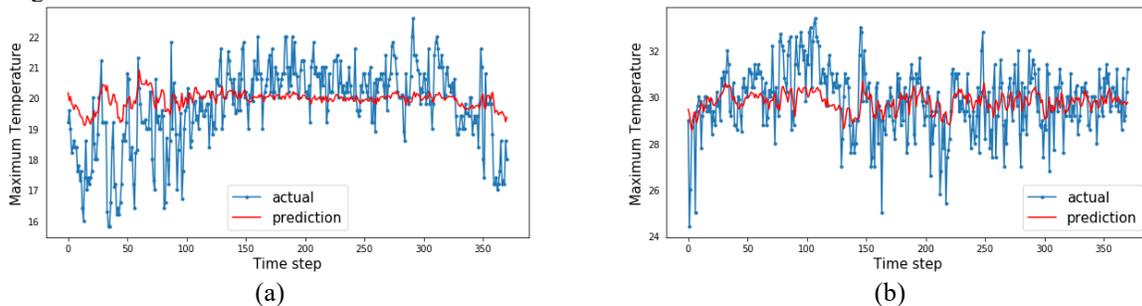


Figure 4. The actual and predicted value for minimum temperature (a) and maximum temperature (b)

4. Conclusion

The temperature is uncertain every day. Therefore, the forecast of the temperature becomes an important information for the society. The accuracy is certainly influenced by the process in the input data, especially data with a long period. RNN can not be used for data with a longer period of time and LSTM can overcome this by adding a cell state so that the old information you need can still be used. The results of this study show that LSTM can provide a small error in the prediction of temperature. In this study, we conducted a trial and error for 10, 20, 30, 40, 50 epochs and obtained the optimum results with the lowest MAPE is 10 epochs. Four neurons in the hidden layer and 10 epochs with the stochastic gradient descent algorithm can deliver MAPE of 3.84% for maximum temperatures and 5.61% for minimum temperatures in Bandung. It is to be hoped that this method can be used for predicting air temperatures in the future by combining numerical models and becoming a standard for weather forecast, specifically for temperatures.

References

- Anwar, S., Peramalan suhu udara jangka pendek di Kota Banda Aceh dengan metode Autoregressive Integrated Moving Average (ARIMA) . *Malikussaleh Journal of Mechanical Science and Technology*, vol. 5, no. 1, pp. 6-12, 2017.
- Ashrovy, R., Recurrent Neural Networ-Part One. Available: <https://medium.com/@ashrovy/recurrent-neural-network-part-one-822f1341fec>, Oktober 2, 2017.
- Caraka, R. E., Mahmud, Z., Sugiyarto, W., and Yasin, H., Long Memory models to forecasting temperatur. *Prosiding Seminar Hari Meteorologi Dunia* , pp. 222-236, 2016.
- Christienova, S. I., Pratiwi, E. W., and Darmana, G., Comparative analysis of Singular Spectrum Analysis and Fourier Series Analysis on air temperature data in Surabaya. *Berkala MIPA*, pp. 94-106, 2018.
- Dewantara, BS., Peramalan Cuaca Tradisional, Available: <http://bagassdsite.blogspot.com/2012/10/peramalan-cuaca-tradisional.html>, October 24, 2012.
- Duan, Y., Lv, Y., Wang, FY., Travel time prediction with LSTM neural network. *IEEE 19th International Conference on Intelligent Transportation System*, pp. 1053-1058, 2016.
- Fromm, G., Comments on "An overview of the objectives and framework of seasonal adjusment" by Shirley Kallek. *Seasonal Analysis of Economic Time Series*, vol. 3, pp. 26-29, 1978
- Hafni, W., Pujiastuti, D., and Harjup, W., Analisis variabilitas temperatur udara di Daerah Kototabang periode 2003-2012. *Jurnal Fisika Unand* , vol. 4, no.2, pp. 185-192, 2015.

- Helmini, S., Jihan, N., Jayasinghe, M., Perera, S., Sales forecasting using multivariate long short term memory network models. *Peer J Preprints*, pp. 1-16, 2019.
- Hochreiter, S., and Schmidhuber, J., Long short term memory. *Neural Computation* , pp. 1-32, 1997.
- Hyndmann, RJ, and Athanasopoulos, G, Forecasting: Principles and Practice, 2nd Edition, OTexts, Australia, 2018.
- Koehrsen, W., Overfitting vs. Underfitting: A Conceptual Explanation. Available: <https://towardsdatascience.com/overfitting-vs-underfitting-a-conceptual-explanation-d94ee20ca7f9>, January 28, 2018.
- Kumar, J., Goomer, R., Singh, AK., Long Short Term Memory Recurrent Neural Network based workload forecasting model for cloud datacenters. *Procedia Computer Science*, pp. 676-682, 2018.
- Lai, CY., Chen, RC., Caraka, RE., Prediction average stock price market using LSTM. pp. 1-6, 2019.
- Le, XH., Ho, HV., Lee, G., Jung, S., Application of Long Short Term Memory Neural Network for flood forecasting. *Water*, pp. 2-19, 2019.
- Liu, Y., Overfitting and forecasting: linear versus non-linear time series models. *Iowa State University Digital Repository*, pp. 1-103, 2000.
- McCulloch, W. S., & Pits, W. H., A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, pp. 115-133, 1943.
- Olah, C., Understanding LSTM Networks. Available: colah.github.io/posts/2015-08-Understanding-LSTMs, August 27, 2015.
- Priyana, J., and Abadi, A. M., Peramalan suhu udara di Yogyakarta dengan menggunakan model fuzzy. *Prosiding Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA*, pp 253-260, 2011.
- Puspitaningrum, D., *Pengantar Jaringan Saraf Tiruan*, ANDI, Yogyakarta, 2006.
- Toffalis, C., A better measure of relative prediction accuracy for model selection and model estimation. *Journal of the Operational Research Society*, pp. 1352-1362, 2015.
- Sesa, S., Suyono, H., Hasanah, RN., Peramalan beban listrik jangka menengah menggunakan jaringan syaraf tiruan pada sistem kelistrikan Kota Ambon, *elektronik Jurnal Arus Elektro Indonesia*, pp. 15-25, 2017.
- Zaytar, M. A., and Amrani, C. E., Sequence to sequence weather forecasting with Long Short Term Memory Recurrent Neural Networks. *International Journal of Computer Applications*, vol. 143, no. 11, pp. 7-11, 2016.

Biographies

Solichatus Zahroh is Graduates of the Statistics study program, Faculty of Mathematics and Natural Sciences, Padjadjaran University, Bandung Indonesia.

Yuyun Hidayat is a lecturer at the Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran. He currently serves as Deputy of the Head of Quality Assurance Unit, a field of Statistics, with a field of Quality Control Statistics.

Resa Septiani Pontoh is a lecturer at the Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, the field of Statistics, with a field of Forecasting.

Agus Santoso is a lecturer in the Department of Statistics, Faculty of Sciences and Technology, Universitas Terbuka. Now is serving as Dean of the Faculty of Science and Technology, Universitas Terbuka. The field of applied statistics, with a field of concentration of Educational Measurement.

Sukono is a lecturer in the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran. Currently as Chair of the Research Collaboration Community (RCC). The field of applied mathematics, with a field of concentration of financial mathematics and actuarial sciences.

Abdul Talib Bon is a professor of Production and Operations Management in the Faculty of Technology Management and Business at the Universiti Tun Hussein Onn Malaysia since 1999. He has a PhD in Computer Science, which he obtained from the Universite de La Rochelle, France in the year 2008. His doctoral thesis was on topic Process Quality Improvement on Beltline Moulding Manufacturing. He studied Business Administration in the Universiti Kebangsaan Malaysia for which he was awarded the MBA in the year 1998. He's bachelor degree and diploma in Mechanical Engineering which his obtained from the Universiti Teknologi Malaysia. He received his postgraduate certificate in Mechatronics and Robotics from Carlisle, United Kingdom in 1997. He had published

*Proceedings of the International Conference on Industrial Engineering and Operations Management
Riyadh, Saudi Arabia, November 26-28, 2019*

more 150 International Proceedings and International Journals and 8 books. He is a member of MSORSM, IIF, IEOM, IIE, INFORMS, TAM and MIM.