Long Short-Term Memory Approach for Predicting Air Temperature in Indonesia

Putu Harry Gunawan[1], Devi Munandar[2], Anis Zainia Farabiba[3]

[1,3]School of Computing, Telkom University, Indonesia
[2]The Indonesian Institute of Sciences, Bandung, Indonesia

Article Info	ABSTRACT
	Air temperature is one of the main factors for describing the weather behaviour in the earth. Since Indonesia is located on and near equator, then monitoring the air temperature is needed to determine either global climate change occurs or not. Climate change can have an impact on biological growth in various fields. For instance, climate change can affect the quality of production and growth of animal and plants. Therefore, air temperature prediction is important to meteorologists and Indonesian government to provide information in many sectors. Various prediction algorithms have been used to predict temperature and produce different accuracy. In this study, the deep learning method with Long Short-Term Memory (LSTM) model is used to predict air temperature. Here, the results show that LSTM model with one layer and Adaptive Moment Estimation (ADAM) optimizer produce accuracy which is 32% of R^2, 0.068 of MAE and 0.99 of RMSE. Moreover, here, ADAM optimizer is found better than Stochastic Gradient Descent (SGD) optimizer.

Corresponding Author:

Putu Harry Gunawan,
School of Computing,
Telkom University,
Jl. Telekomunikasi No 1, Terusan Buah Batu, Bandung 40275, Indonesia
Email: phgunawan@telkomuniversity.ac.id

1. INTRODUCTION

Indonesia has three climates namely sea, tropical and seasonal climate. This is because of Indonesia is located on and near equator. Since it is located on and near equator, then it is clearly possible to get vulnerable to climate change [1, 2]. Climate change is a natural phenomenon which can affects many sectors in biological growth of plants and animals in earth. One of the causes of climate change is air temperature [3]. Here, air temperature plays important role in various fields. For instances, it can be a factor of agricultural production because it affects the development and process of plant growth. In the health sector, fluctuating temperatures can also cause various diseases. Moreover, high fluctuation of air temperature can also had an effect on the aviation sector on aircraft flight schedules [4]. Therefore, the need of predicting air temperature is important to help many sectors for Indonesia Government.

Machine Learning has a basic definition of algorithms that process data, study the data, and apply what it has been learned to make decisions. The results of learning form machine learning will provide prediction results. While here, deep learning is part of machine learning which is designed to continuously analyze data with a logical structure. This can make deep learning is similar to human in making decisions. In order to achieve this ability, deep learning uses a layered algorithm structure called the artificial neural network (ANN).

In this research, the prediction of air temperature using machine learning approach will be elaborated. Since this approach have been applied in many applications and shown a robust approach for prediction problem (see [5, 6] for examples). Although some prediction methods based on time series data are available in some literatures (see [7, 8, 9, 10, 11] for more detail. The deep learning approach will be elaborated to

approximate the air temperature in a local area in Indonesia (west of Sumatra). Here, Long Short-Term Memory (LSTM) approach will be used in this research.

LSTM consist of Recurrent Neural Network (RNN) which issued with a special gating procedure, such that it can controls the access to memory cells [12]. The benefit of these properties is LSTM is able to process data into complex system and applications such as speech recognition [13], handwriting recognition [14], generating text [15], image-to-caption generation [16], and machine translation [17].

In this study, the air temperature forecasting system is developed using the LSTM model using time series of air temperature data. Here, the measurement data is collected from weather measurement station which belongs to the Indonesian Institute of Sciences or known as LIPI. This station collects the data per hours in time series mode. The aim of this work is to find the best model of LSTM by paying attention to the error rate of RMSE (Root Mean Square Error). Therefore, it is expected that the existence of this forecasting method provides a new way in the development of air temperature prediction model and can be used as the basis for the development of subsequent research.

2. METHOD

Deep learning approach in trading model type consists of several processing layers for representing data with various levels of abstraction. Using the backpropagation algorithm, deep learning is about a machine with a task to change the internal parameters in each layers from previous layers representation [18].

Sepp Hochreiter dan Jurgen Schmidhuber [12] in 1997 were the first authors whose proposed LSTM model. Moreover, this model has been developed by several researchers such as: Alex Graves [19, 13], Hasim Sak [20], Wojciech Zaremba [21], and many others [22]. The architecture of LSTM uses RNN (Recurrent Neural Network) architecture which designed to solve the problem of vanishing gradient on RNN using gating mechanism [23].
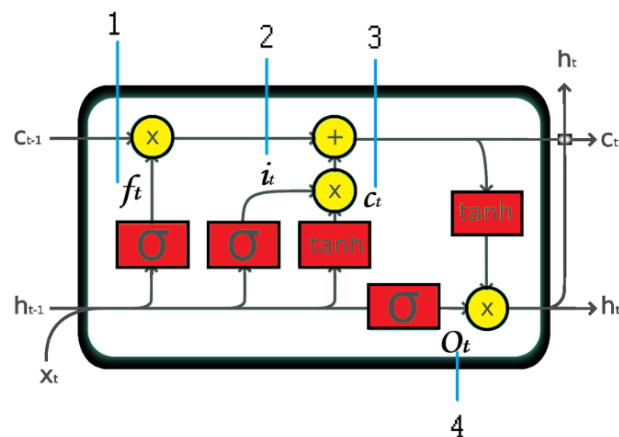

Figure 1. Structure of LSTM

LSTM model has 4 layers in its structure, which are called Forget Gate (1), Input Gate (2), Cell Gate (3), and Output Gate (4) (see Figure 1). In Figure 1, it shows that the black arrow represents the flow of information in the cell, between cells and out of cell (output h). Meanwhile, yellow circles represent operations of element-wise: additions and multiplications, and red boxes represent neural network layer.

LSTM has two outputs, one is the actual output which is passed to the next cell and becomes the output of that cell and the other one is the cell state (c_t). The basic idea of LSTM is update cell states with element-wise operations in each cell of LSTM.

Number one in Figure 1 is forget gate or neuron layer that determines the information from input (x_t) and output (h_{t-1}) which will be stored or discarded on memory cells. Moreover, the activation function is sigmoid, where the output is between 0 and 1. If the output is 1 then all data will be stored and if the output is 0 then all data will be discarded. The formula is given as follows [20, 22],

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

where f_t is forget gate, W_f is weight for forget gate, $h_{\{t-1\}}$ is value of output from before cell, x_t input, b_t is bias for forget gate, and σ is sigmoid activation function as shown in (2).

Long Short-Term Memory Approach for Predicting Air Temperature in Indonesia
(Putu Harry Gunawan[1], Devi Munandar[2], Anis Zainia Farabiba[3])

162

Given any variable z, then the sigmoid function is defined as,

$$\sigma\,(z) = \frac{1}{e^{-z}}\,. \tag{2}$$

The next step is input gate. Here, there are 2 layers in input gate. In Figure 1 the layers are numbered two and three. Neuron layer with sigmoid activation function decides which value will be updated. Then $tanh$ activation function will create a new vector that will be stored in memory cell [20, 22]. The equations are given as follows:

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

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

where i_t is input gate, C_t is represents the value that will be stored in memory cells, W_i and are value of weight, b_i and b_c are value of bias and tanh activation function is defined in (5).

$$\tanh(x) = 2\sigma(2x) - 1\,. \tag{5}$$

In cell gate, it will replace the value in previous memory cell with the new value of memory cell. Where this value is obtained from combining the values contained in forget gate and input gate [20] using the following equation:

$$C_t = f_t \times C_{\{t-1\}} + i_t \times c_t. \tag{6}$$

The last one is *output gate* (Figure 1 number 4). First process uses sigmoid activation function to decide which value will be the output. Then multiplying (7) and (5), thus resulting the output of cell (output h) as shown in (8).

$$O_t = \sigma\left(W_0 \cdot \left[h_{\{t-1\}}, x_t\right] + b_0\right), \tag{7}$$

$$h_t = O_t \times \tanh\left(c_t\right)\,. \tag{8}$$

3. METHODOLOGY AND DATASET

In this section, the methodology of this research which consists of research area, preprocessing data and LSTM network process will be elaborated.
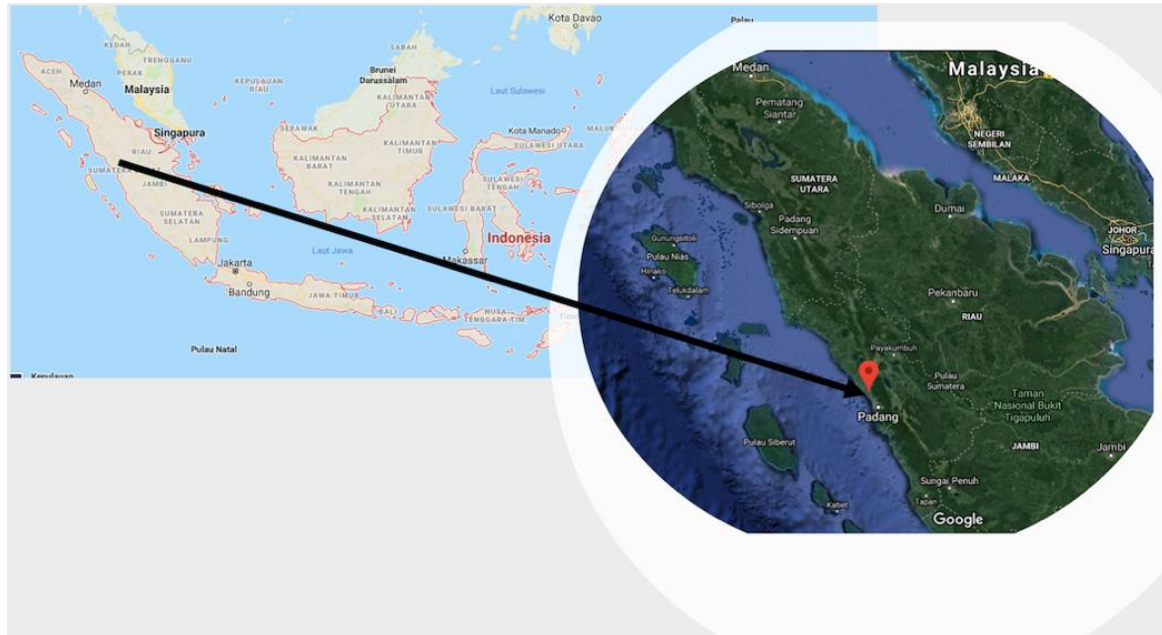

Figure 2. Location of weather monitoring stations in Muaro Anai Padang, Sumatra Indonesia

3.1. Research area and data collection

In this research, air temperature data (C^0) was taken from Indonesian Institute of Science or known as Lembaga Ilmu Pengetahuan Indonesia (LIPI) Weather Measurement Station. The station is located at west of Sumatra, in Muaro Anai, Padang. Geographically it is located at 0 47012:8" South Latitude and 100 17018:2" East Latitude, Indonesia (see Figure 2 for more detail). This area receives daily average wind speed 1.53 Km/h, temperature 25.23 C, humidity 89.44%RH, and rainfall 1.17 mm. The air temperature data for this study was collected for a period of January 2015 to December 2017. In Figure 2 the location of the observation station is shown trough maps at coordinates -0.786876, 100.288399.

3.2. Pre-processing data

The raw data from weather station are very vulnerable to several problems. For instances, missing value, noisy, outliers, inconsistent data, redundancy, etc are the general problem can be found in raw data. Therefore, the raw data require pre-processing procedures to produce high-quality data, such that the results of predicting have high accuracy [24, 25]. In this this research, pre-processing data will be done for missing value and normalization problem.

3.2.1. Missing value problem

Missing value problem is a problem where the information in raw data is incomplete. This problem can happen because of an error when inputting data, hard-to-find values, information about objects not given or indeed the information does not exist. In order to handle this problem, the general tool for solving this problem can be used. The tool can be the average value (Mean) of the attribute or find the value that often appears (Mode) [24]. This study the mean tool will be used to fill in the missing value in raw data.

3.2.2. Normalization

Normally, in order to avoid the high fluctuation data during prediction process, then the value of raw data should be normalized. Normalization is the process of reducing data into an interval [0, 1], with calculation is given as follows:

$$X'_t = \left(\frac{x_t - \min(x)}{\max(x) - \min(x)} \right).$$ (9)

where,

X'_t = value of normalization data,

x_t = observation data that will be normalized.

3.3. LSTM network process

The process of LSTM network consists of a sequence of input and target pairs $(X_1, Y_1), \cdots, (X_n, Y_n)$. Here, for each pair (X_i, Y_i) takes new input X_i, and produces as estimate for target Y_i. This process only uses one hidden layer to provide the estimate value, the diagram can be illustrated as in Figure 3.
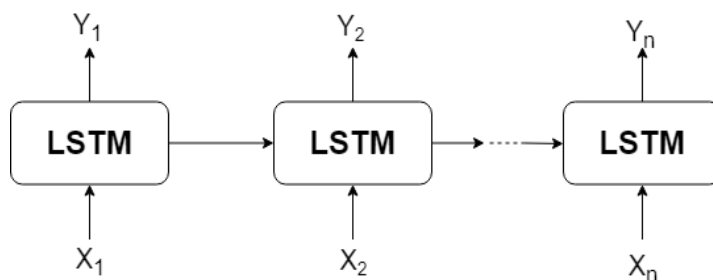

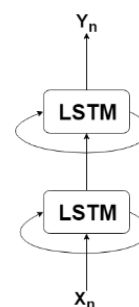Figure 3. Structure of one hidden LSTM layer

Figure 4. Structure of two hidden LSTM layer

The number of layers in LSTM can vary which depends on the model that is chosen. The large number of hidden layer is not guarantee that the model produces good results. For instance in [26], the LSTM with three hidden layers is found too complex, therefore, processing data is difficult. Figure 4 shows the structure of LSTM using two hidden layers. The input X must trough two layers before resulting the estimation target Y.

Long Short-Term Memory Approach for Predicting Air Temperature in Indonesia
(Putu Harry Gunawan[1], Devi Munandar[2], Anis Zainia Farabiba[3])

164

Here, the LSTM with two hidden layers were constructed using the similar idea as the model presented in paper [27].

4. RESULT AND DISCUSSION

In order to examine the effectiveness of the LSTM estimation, then the coefficient determination (R^2). should be calculated. This coefficient determination is obtained from (10) using observation output and network predicted output. Moreover, the measurement of errors can be calculated by Mean Absolute Error (MAE), Mean Squared Error (RMSE) and Mean Root Squared Error (RMSE) where presented in (11), (12) and (13) respectively.

$$R^2 = \left(\frac{\sum_{j=1}^{n} Y_j x_j}{\sqrt{\sum_{j=1}^{n} Y_j^2 \sum_{j=1}^{n} x_j^2}} \right)^2 , \qquad (10)$$

$$MAE = \frac{1}{n} \sum_{j=1}^{n} |Y_j - x_j| , \qquad (11)$$

$$MSE = \frac{1}{n} \sum_{j=1}^{n} (Y_j - x_j)^2 , \qquad (12)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^{n} (Y_j - x_j)^2} , \qquad (13)$$

where n is the number of test points, x and Y are the observed and predicted values and j is the index of value of the air temperature dataset.

As mentioned earlier, the LSTM model in this research uses one and two layers with 15 epochs in Muaro Anai area. The input parameters were daily averaged and fed as the input for the LSTM model. The output air temperature of the model will be averaged daily. Here, two optimizers will be investigated for each layer model. The two optimizers are called Adaptive Moment Estimation (ADAM) and Stochastic Gradient Descent (SGD) optimizer.

The comparison of ADAM and SGD optimizer for each hidden layer can be found in Table 1. Here, ADAM optimizer works better compared with SGD optimizer function in one and two hidden layers (see blue cell in Table 1 for one layer). Here the best optimizer is chosen based on the high value of R^2 from (10) and small value of errors (MAE, MSE and RSME).

As seen in Table 1, ADAM optimizer function is observed has R^2 and RMSE of 0.32 and 0.99 respectively in one-layer model. This is cased one-layer model with ADAM optimizer produces better approximation. Meanwhile ADAM and SGD optimizer in two layers produce high errors and small coefficient determination.

Table 1. Comparison of ADAM and SGD optimizer in one and two hidden layers

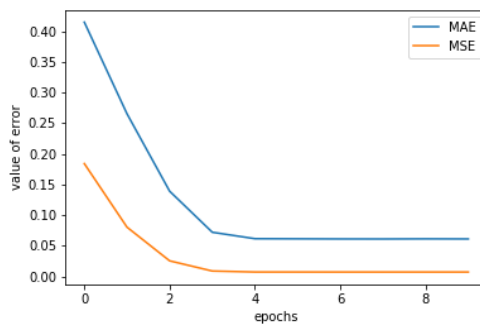| Layers	Optimizer	Input Days	MAE	MSE	RMSE Train	RMSE Test	R^2
1	ADAM	1	0.074	0.008	1.02	1.03	0.25
		2	0.073	0.009	0.99	1.04	0.24
		7	0.068	0.008	0.97	0.99	0.32
	SGD	1	0.085	0.011	1.16	1.16	0.06
		2	0.084	0.010	1.14	1.15	0.08
		7	0.081	0.010	1.13	1.12	0.13
2	ADAM	1	0.085	0.010	1.16	1.15	0.07
		2	0.083	0.010	1.12	1.13	0.11
		7	0.077	0.009	1.08	1.09	0.18
	SGD	1	0.089	0.011	1.22	1.21	-0.02
		2	0.089	0.011	1.21	1.2	-0.01
		7	0.089	0.012	1.22	1.21	-0.02

Figure 5. The error profile of MAE and MSE along the epochs

Moreover, the profile of MAE and MSE values along the increasing of epochs can be seen in Figure 5. In Figure 5 it can be seen that at epochs 4-5, there is no change in error value which means the weight value (W) is stable or no weight updated. Finally, the prediction results using LSTM one-layer model is given in Figure 6.

In Figure 6, it can be shown that the testing or predicting data is in a god agreement with the observation data. The error of this figure using ADAM optimizer and one hidden layer is shown in Table 1. The error is not too small enough, since the observation data is in fluctuation shape. However, still, this model can capture the prediction in a good result. Moreover the (R^2) between observation output and predicted output were showed by Figure 7 in scatter type.
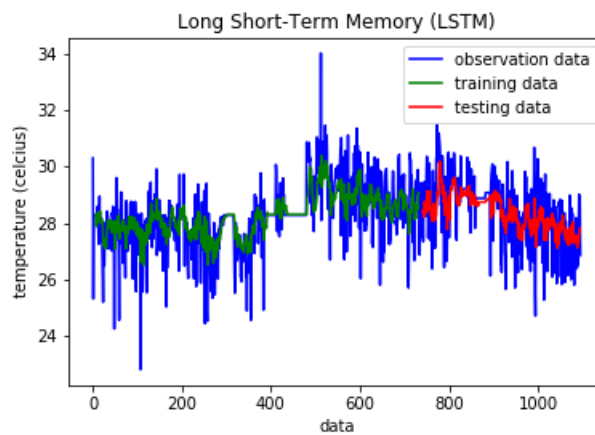

Figure 6. Observed and predicted profile

In Figure 7 the scatter plot of observation data is described by blue dots. Meanwhile the linear line is described the prediction results by LSTM. The prediction line is shown able to approximate the observation data in the middle of observation data distribution with R2 core is 32%.
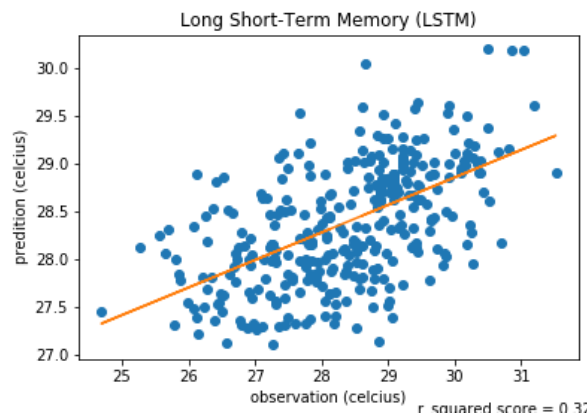

Figure 7 Scatter plot profile from air temperature prediction using LSTM with one hidden layer and ADAM optimizer

Long Short-Term Memory Approach for Predicting Air Temperature in Indonesia
(Putu Harry Gunawan[1], Devi Munandar[2], Anis Zainia Farabiba[3])

166

5.   CONCLUSION

The prediction of air temperature using Long Short-Term Memory (LSTM) model with two layers and optimizer is elaborated. Here, the LSTM model with ADAM optimizer and SGD optimizer are used for predicting the average air temperature. The result shows that LSTM model with one hidden layer and ADAM optimizer produces accuracy R^2 of 32%, MAE value 0.068 and RMSE value 0.99. These results are shown better than LSTM model with two hidden layers for both optimizers. By present study, it is concluded that LSTM model with ADAM optimizer in one hidden layer has a better potential to predict air temperature than LSTM with SGD optimizer.

6.   REFERENCES

[1] S. E. Page, R. Wust, D. Weiss, J. O. Rieley, W. Shotyk, and S. H. Limin, "A record of late pleistocene and holocene carbon accumulation and climate change from an equatorial peat bog (kalimantan, indonesia): implications for past, present and future carbon dynamics," *Journal of Quaternary Science*, vol. 19, no. 7, pp. 625–635, 2004.

[2] F. Nhita et al., "A rainfall forecasting using fuzzy system based on genetic algorithm," *in Information and Communication Technology (ICoICT), 2013 International Conference*. IEEE, 2013, pp. 111–115.

[3] A. Klein Tank, J. Wijngaard, G. Konnen, R. Bohm, G. Demaree, A. Gocheva, M. Mileta, S. Pashiardis, L. Hejkrlik, C. Kern-Hansen et al., "Daily dataset of 20th-century surface air temperature and precipitation series for the european climate assessment," *International Journal of Climatology: A Journal of the Royal Meteorological Society*, vol. 22, no. 12, pp. 1441–1453, 2002.

[4] J. Priyana and A. M. Abadi, "Peramalan suhu udara di yogyakarta dengan menggunakan model fuzzy," in *Prosiding Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA: Fakultas MIPA Universitas Negeri Yogyakarta*, 2011.

[5] C. O. Alm, D. Roth, and R. Sproat, "Emotions from text: machine learning for text-based emotion prediction," in *Proceedings of the conference on human language technology and empirical methods in natural language processing*. Association for Computational Linguistics, 2005, pp. 579–586.

[6] H. Nielsen, S. Brunak, and G. von Heijne, "Machine learning approaches for the prediction of signal peptides and other protein sorting signals," *Protein engineering*, vol. 12, no. 1, pp. 3–9, 1999.

[7] P. M. Ferreira, E. Faria, and A. Ruano, "Neural network models in greenhouse air temperature prediction," *Neurocomputing*, vol. 43, no. 1-4, pp. 51–75, 2002.

[8] B. A. Smith, R. W. McClendon, and G. Hoogenboom, "Improving air temperature prediction with artificial neural networks," *International Journal of Computational Intelligence*, vol. 3, no. 3, pp. 179–186, 2006.

[9] C. Inard, H. Bouia, and P. Dalicieux, "Prediction of air temperature distribution in buildings with a zonal model," *Energy and buildings*, vol. 24, no. 2, pp. 125–132, 1996.

[10] H. Swaid and M. E. Hoffman, "Prediction of urban air temperature variations using the analytical cttc model," *Energy and Buildings*, vol. 14, no. 4, pp. 313–324, 1990.

[11] R. F. Chevalier, G. Hoogenboom, R. W. McClendon, and J. A. Paz, "Support vector regression with reduced training sets for air temperature prediction: a comparison with artificial neural networks," *Neural Computing and Applications*, vol. 20, no. 1, pp. 151–159, 2011.

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[13] A. Graves, N. Jaitly, and A.-r. Mohamed, "Hybrid speech recognition with deep bidirectional lstm," in *Automatic Speech Recognition and Understanding* (ASRU), 2013 IEEE Workshop on. IEEE, 2013, pp. 273–278.

[14] A. Graves and J. Schmidhuber, "Offline handwriting recognition with multidimensional recurrent neural networks," in *Advances in neural information processing systems*, 2009, pp. 545–552.

[15] I. Sutskever, J. Martens, and G. E. Hinton, "Generating text with recurrent neural networks," in *Proceedings of the 28th International Conference on Machine Learning* (ICML-11), 2011, pp. 1017–1024.

[16] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and tell: A neural image caption generator," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3156–3164.

[17] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in neural information processing systems*, 2014, pp. 3104–3112.

[18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," nature, vol. 521, no. 7553, p. 436, 2015.

[19] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional lstm and other neural network architectures," *Neural Networks*, vol. 18, no. 5-6, pp. 602–610, 2005.

[20] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition*," arXiv preprint arXiv:1402.1128*, 2014.

[21] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization," *arXiv preprint arXiv:1409.2329*, 2014.

[22] A. Geron, ´ Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems." *O'Reilly Media*, Inc.", 2017.

[23] S. Hochreiter, Y. Bengio, P. Frasconi, J. Schmidhuber et al., "Gradient flow in recurrent nets: the difficulty of learning long-term dependencies," *Sepp Hochreiter Fakult at f ur Informatik*, 2001.

[24] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.

[25] M. Berry and G. Linoff, Mastering data mining: *The art and science of customer relationship management*. John Wiley & Sons, Inc., 1999.

[26] Z. Liu and C. J. Sullivan, "Prediction of weather induced background radiation fluctuation with recurrent neural networks," *Radiation Physics and Chemistry*, 2018.

[27] A. Graves, "Generating sequences with recurrent neural networks," *arXiv preprint arXiv:1308.0850*, 2013.

Long Short-Term Memory Approach for Predicting Air Temperature in Indonesia
(Putu Harry Gunawan[1], Devi Munandar[2], Anis Zainia Farabiba[3])

168