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

Putu Harry Gunawan¹, Devi Munandar², Anis Zainia Farabiba³
¹²School of Computing, Telkom University, Indonesia
³The Indonesian Institute of Sciences, Bandung, Indonesia

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 whether 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.

Keywords: Air temperature, Deep learning, Long short-term memory, Prediction

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 instance, 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](image)

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],

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

(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 $\sigma$ is sigmoid activation function as shown in (2).
Given any variable $z$, then the sigmoid function is defined as,

$$
\sigma(z) = \frac{1}{1 + e^{-z}}.
$$

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:

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

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

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

$$
tanh(x) = 2\sigma(2x) - 1.
$$

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:

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

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).

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

$$
h_t = O_t \times \tanh(c_t).
$$

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

(Putu Harry Gunawan¹, Devi Munandar², Anis Zainia Farabiba³)

3.1. Research area and data collection

In this research, air temperature data (°C) 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 instance, 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 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) 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:

\[ X'_t = \left( \frac{x_t - \min(x)}{\max(x) - \min(x)} \right) \]  

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](image)

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.

![Figure 4. Structure of two hidden LSTM layer](image)
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.

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

\[
MAE = \frac{1}{n} \sum_{j=1}^{n} |Y_j - x_j|, \quad (11)
\]

\[
MSE = \frac{1}{n} \sum_{j=1}^{n} (Y_j - x_j)^2, \quad (12)
\]

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

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>
<thead>
<tr>
<th>Layers</th>
<th>Optimizer</th>
<th>Input Days</th>
<th>MAE</th>
<th>MSE</th>
<th>RMSE Train</th>
<th>RMSE Test</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ADAM</td>
<td>1</td>
<td>0.074</td>
<td>0.008</td>
<td>1.02</td>
<td>1.03</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.073</td>
<td>0.009</td>
<td>0.99</td>
<td>1.04</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td>0.068</td>
<td>0.008</td>
<td>0.97</td>
<td>0.99</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>SGD</td>
<td>1</td>
<td>0.085</td>
<td>0.011</td>
<td>1.16</td>
<td>1.16</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.084</td>
<td>0.010</td>
<td>1.14</td>
<td>1.15</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td>0.081</td>
<td>0.010</td>
<td>1.13</td>
<td>1.12</td>
<td>0.13</td>
</tr>
<tr>
<td>2</td>
<td>ADAM</td>
<td>1</td>
<td>0.085</td>
<td>0.010</td>
<td>1.16</td>
<td>1.15</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.083</td>
<td>0.010</td>
<td>1.12</td>
<td>1.13</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td>0.077</td>
<td>0.009</td>
<td>1.08</td>
<td>1.09</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>SGD</td>
<td>1</td>
<td>0.089</td>
<td>0.011</td>
<td>1.22</td>
<td>1.21</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.089</td>
<td>0.011</td>
<td>1.21</td>
<td>1.2</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td>0.089</td>
<td>0.012</td>
<td>1.22</td>
<td>1.21</td>
<td>-0.02</td>
</tr>
</tbody>
</table>
Long Short-Term Memory Approach for Predicting Air Temperature in Indonesia
(Putu Harry Gunawan¹, Devi Munandar², Anis Zainia Farabiba³)

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.

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

Figure 6. Observed and predicted profile

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


