

Weather Forecasting Using Merged Long Short-Term Memory Model (LSTM) and Autoregressive Integrated Moving Average (ARIMA) Model

¹Afan Galih Salman, ²Yaya Heryadi, ²Edi Abdurahman and ³Wayan Suparta

¹Computer Science Department, School of Computer Science,

Bina Nusantara University, Jakarta, Indonesia 11480, Indonesia

^{1,2}Computer Science Department, BINUS Graduate Program-Doctor of Computer Science,

Bina Nusantara University, Jakarta, Indonesia 11480, Indonesia

³Department of Electrical Engineering, Faculty of Science and Technology,

Sanata Dharma University, Yogyakarta 55282, Indonesia

Article history

Received: 04-04-2018

Revised: 11-06-2018

Accepted: 26-06-2018

Corresponding Author:

Afan Galih Salman

Computer Science Department,

School of Computer Science,

Bina Nusantara University,

Jakarta, Indonesia 11480,

Indonesia

Email: asalman@binus.edu

Abstract: Weather forecasting is an interesting research problem in flight navigation area. One of the important weather data in aviation is visibility. Visibility is an important factor in all phases of flight, especially when the aircraft is maneuvering on or close to the ground, i.e., during taxi-out, take-off and initial climb, approach and landing and taxi-in. The aim of these study is to analyze intermediate variables and do the comparison of visibility forecasting by using Autoregressive Integrated Moving Average (ARIMA) and Long Short Term Memory (LSTM) Model. This paper proposes ARIMA model and LSTM model for forecasting visibility at Hang Nadim Airport, Batam Indonesia using one variable weather data as predictor such as visibility and combine with another variable weather data as moderating variables such as temperature, dew point and humidity. The models were tested using weather time series data at Hang Nadim Airport, Batam Indonesia. This research compares the Root Mean Square Error (RMSE) resulted by LSTM model with the RMSE resulted by ARIMA model. The results of this experiment show that LSTM model with/without intermediate variable has better performance than ARIMA Model.

Keywords: Weather Forecasting, Visibility, Autoregressive Integrated Moving Average, Long Short-Term Memory, Root Mean Square Error

Introduction

Weather forecasting attracts great attention of researchers from various research communities due to its effect to the global human life. The current wide availability of massive weather observation data motivated researchers to explore pattern hidden in the large dataset and make weather prediction. The advent of information and computer technology in the last decade also support the research of weather forecasting to get more accurate results.

Weather forecasting is an interesting research problem in flight navigation area (Maunder *et al.*, 2000). One of the important weather data in aviation is visibility. Visibility is an important factor in all phases of flight, especially when the aircraft is maneuvering on or close to the ground, i.e., during taxi-out, take-off and initial climb, approach and landing and taxi-in.

Aircraft departure and arrival is limited by the visibility (or RVR) to an extent that depends on the sophistication of ground equipment, the technical equipment fitted to the aircraft and the qualification of the flight crew. Many aerodromes and aircraft are fitted with equipment that makes a landing possible in very low visibility conditions. However, in very low visibility, it may prove impossible for the pilot to navigate the aircraft along the runway and taxiways to the aircraft stand.

The widely adopted Autoregressive-Moving-Average (ARIMA) model was proposed by Box and Jenkins (1976). This statistical model consists of two polynomial parts, one for the auto-regression and the other for moving average. Many extension of this model have been proposed in literature such as: Autoregressive Integrated Moving Average (ARIMA), Auto-Regressive Conditional Heteroscedasticity (ARCH).

ARIMA method (Box and Jenkins, 1976) has assumption that the time series under study are generated from linear processes in order to make a successful prediction.

There are three advantages of linear models assumption that are it is easily to be understood, it is enable to be analyzed in great detail, it is easily to be explained and implemented.

Many significant efforts to solve weather forecasting problem use statistical modeling including machine learning techniques with successful results. Afsin proposed a neural network fuzzy wavelet model for long term rainfall forecasting (Afsin *et al.*, 2011) and Belayneh proposed standard precipitation index drought forecasting using wavelet neural networks and support vector regression (Belayneh and Adamowski, 2012). Salman proposed Recurrent Neural Network (RNN) used heuristically optimization method for rainfall prediction based on weather dataset comprises of ENSO (Salman *et al.*, 2015). ConvLSTM with the Trajectory GRU (TrajGRU) model to predict the future rainfall intensity in a local region over a relatively short period of time that can actively learn the location-variant structure for recurrent connections. TrajGRU is more efficient in capturing the spatiotemporal correlations than ConvGRU (Shi *et al.*, 2015). ConvLSTM is a variant of LSTM (Long Short-Term Memory) containing a convolution operation inside the LSTM cell. Seongchan proposed a model to predicts the amount of rainfall from weather radar data using convolutional LSTM (ConvLSTM). This model showed that two-stacked ConvLSTM reduced RMSE by 23.0% over the linear regression (Kim *et al.*, 2017).

A recurrent convolutional neural network forecast meteorological attributes, such as temperature, air pressure and wind speed used visualization system. These system helped the user to quickly assess, adjust and improve the network design (Roesch and Günther, 2017). A hybrid approach model that combines trained predictive models with a deep neural network that simulate the joint statistics of a set of weather-related variables. The result show how the base model can be enhanced by spatial interpolation that uses learned long-range spatial dependencies (Grover *et al.*, 2015).

Although many models have been proposed over the past ten years, there is no single model which predict weather variables with high accuracy. In addition, most of prominent weather forecasting models only used predictor variables as input. The novelty of the proposed method for forecasting a weather variable is the used of the moderating variables and the merged Long Short-term Memory Model. This research will explore several weather variables as moderating variable to forecast visibility variable. Therefore, the purposes of this research are developing a merged-LSTM model and

ARIMA model to predict visibility variable, analyzing influence of the intermediate variable to predict visibility and comparing the Root Mean Square Error (RMSE) resulted by these two models.

Research Method

Dataset and Data Preprocessing

This dataset for this research was obtained from Weather Underground (<https://www.wunderground.com/>) which collects weather data including temperature, dew point, humidity and visibility on Hang Nadim Airport in Indonesia Area. The range of data for this study was obtained from year 2012 to year 2016 that consist of 40,026 time series data.

The main data preprocessing applied in this research are normalizing (2.1), rescaling into range [0,1] (2.2) and smoothing using Moving Average (MA) with lag = 9 (2.3). Consider weather time series data in T time interval $X = [x_1, x_2, \dots, x_T]$:

$$x_t = \frac{x_t - \bar{x}}{s_x} \quad (2.1)$$

$$x'_t = \frac{x_t - x_{min}}{x_{max} - x_{min}} \quad (2.2)$$

$$x''_t = \frac{1}{9}(x'_t + x'_{t-1} + \dots + x'_{t-8}) \quad (2.3)$$

Where:

x_t = Observation at t

x'_t = Normalized data at t

x''_t = The result of data smoothing using moving average at t

Correlation between two weather variables are measured using coefficient correlation (r) that was computed using Equation 2.4:

$$r = \frac{1}{T-1} \sum_{t=1}^T \left(\frac{x_t - \bar{x}}{s_x} \right) \left(\frac{y_t - \bar{y}}{s_y} \right) \quad (2.4)$$

where, $-1 \leq r \leq 1$; s_x and s_y are standard deviation variable X and Y respectively which were computed using the following formula:

$$s_x = \sqrt{\frac{\sum_{t=1}^T (x_t - \bar{x})^2}{T-1}} \text{ and } s_y = \sqrt{\frac{\sum_{t=1}^T (y_t - \bar{y})^2}{T-1}} \quad (2.5)$$

However, after being preprocessed by moving average method, the data distribution looked a bit smoother.

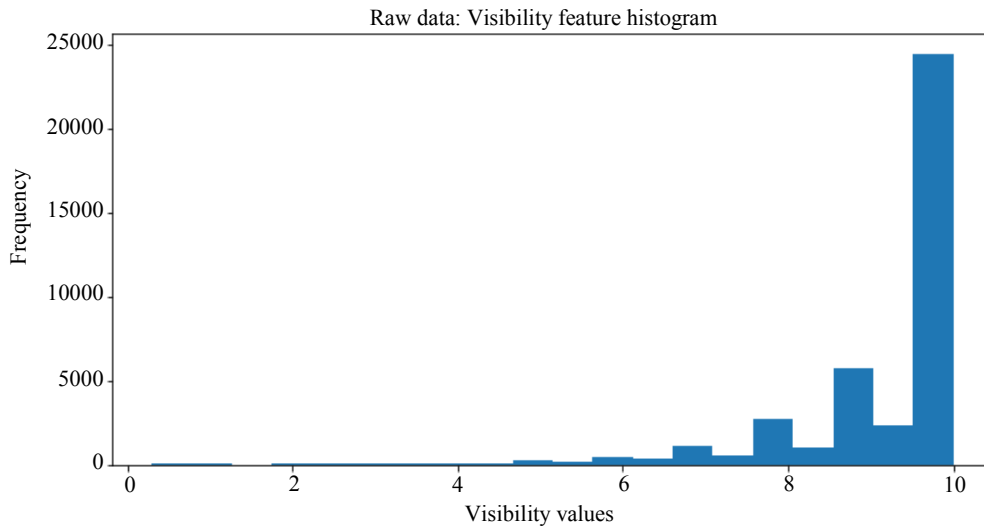


Fig. 1: Visibility data before data preprocessing

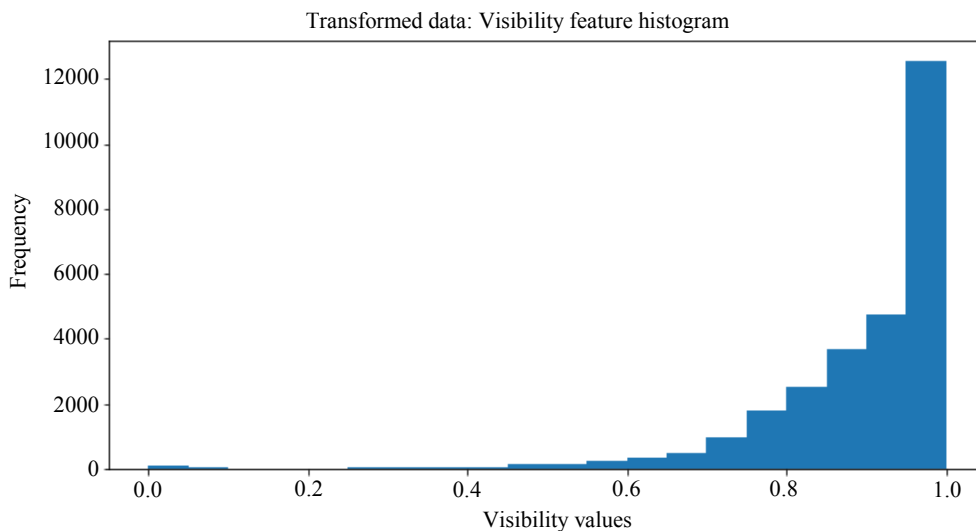


Fig. 2: Visibility data after data preprocessing

Figure 1 and 2 shows the result of visibility data before and after data preprocessing

Finally, for the purposes of model cross-validation, the total training data was divided randomly into 27,915 (70%) training and 11,906 (30%) testing dataset.

ARIMA Model

The acronym ARIMA stands for Auto-Regressive Integrated Moving Average. The stationarized series in the forecasting equation are called "autoregressive" terms. The forecast errors are called "moving average" terms.

An Autoregressive Integrated Moving Average (ARIMA) is statistical properties and the well-known Box-Jenkins methodology (Box and Jenkins, 1976).

A nonseasonal ARIMA model is classified as an "ARIMA(p,d,q)" model, where:

- p is the number of autoregressive terms
- d is the number of nonseasonal differences needed for stationary
- q is the number of lagged forecast errors in the prediction equation

An ARIMA model can be viewed as a "filter" that tries to separate the signal from the noise and the signal is then extrapolated into the future to obtain forecasts.

Tseng propose a hybrid forecasting model, which combines the seasonal time Series ARIMA (SARIMA) and the neural network Back Propagation (BP) models,

known as SARIMABP. The SARIMABP model was also able to forecast certain significant turning points of the test time series (Tseng *et al.*, 2002).

Tektas presents a comparative study of statistical and neuro-fuzzy network models for forecasting the weather of Göztepe, İstanbul, Turkey using Adaptive Network Based Fuzzy Inference System (ANFIS) and Auto Regressive Moving Average (ARIMA) models (Tektas 2010), Mallick and Jain (2017) try to predict weather for a particular period by using the strategy of Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing (ETS) and Naseem *et al.* (2017) propose ARIMA to predict air quality in Islamabad urban area using meteorological variables as predictors.

LSTM Model

This study use LSTM model which is a deep learning model proposed by Schmidhuber (Hochreiter and Schmidhuber, 1997). The model has been successfully used in many research fields such as, large scale image classification (Real *et al.*, 2017), video classification (Yoo, 2017), natural language processing (Elkaref and Bohnet, 2017), anomaly detection (Luo *et al.*, 2017; Lee *et al.*, 2018). LSTM was used as a foundation for weather forecasting model because of several reasons that are (1) the model ability to solve long lag relationship in time series data (2) the model ability to address vanishing gradient problems that occur in the training deep structure neural networks (Gers *et al.*, 2000).

LSTM is a recurrence Neural Network that first introduced by (Hochreiter and Schmidhuber, 1997) as a specific Recurrent Neural Network (RNN) architecture

that was designed to model temporal sequences. Better than the conventional RNN, LSTM is able to sort error backflow problem so that this algorithm only use the error feedback that can make more accurate prediction whereas all unsupported feedback are removed. This algorithm has sorting capability due to LSTM contains special units called memory blocks in the recurrent hidden layer. The LSTM overcome the weakness in conventional RNN that show backpropagation algorithm in RNNs cause error signals that flows backward in time tend to explode or diminish; therefore, the temporal evolution of the backpropagated error exponentially depends on the size of the weight. In other words, the strength of LSTM is special units called memory blocks in the recurrent hidden layer.

From the diagram in Fig. 3, it can be seen that each LSTM block receives the following signals: Input signal (x), input gate signal (i), recurrent signal (h) and forget gate signal (f); and produces output gate signal (o).

The memory blocks contain memory cells with self-connections storing the temporal state of the network in addition to special multiplicative units called gates to control the flow of information. Each memory block in the original architecture contained three gate types which are namely:

- Input gate: The input gate controls the flow of input activations into the memory cell
- Output gate: Output gate controls the output flow of cell activations into the rest of the network
- Forget gate: Scales the internal state of the cell before adding it as input to the cell through the self-recurrent connection of the cell, therefore adaptively forgetting or resetting the cell's memory

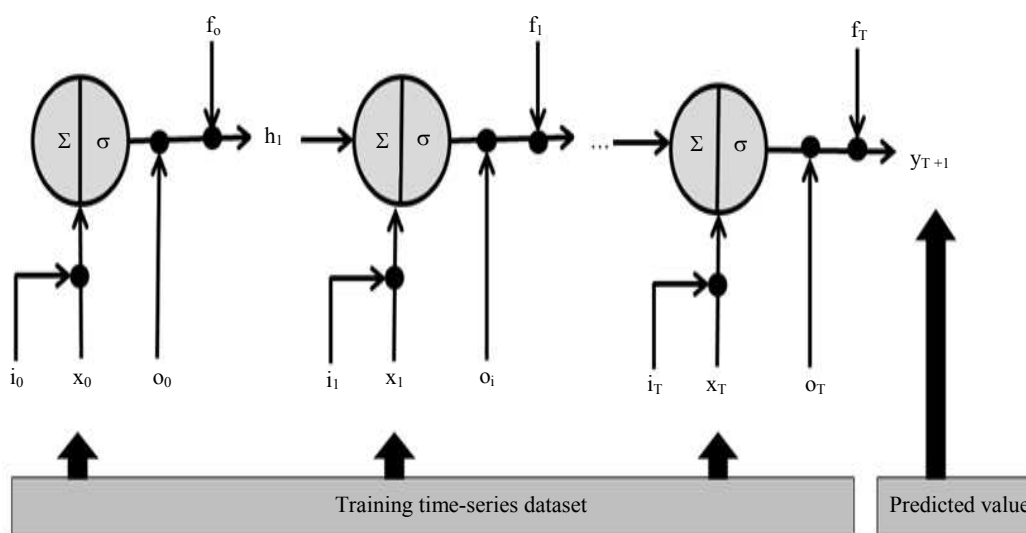


Fig. 3: Architecture of Unfolded Long Short-Term Memory Model (Adapted from: Hochreiter and Schmidhuber, 1997)

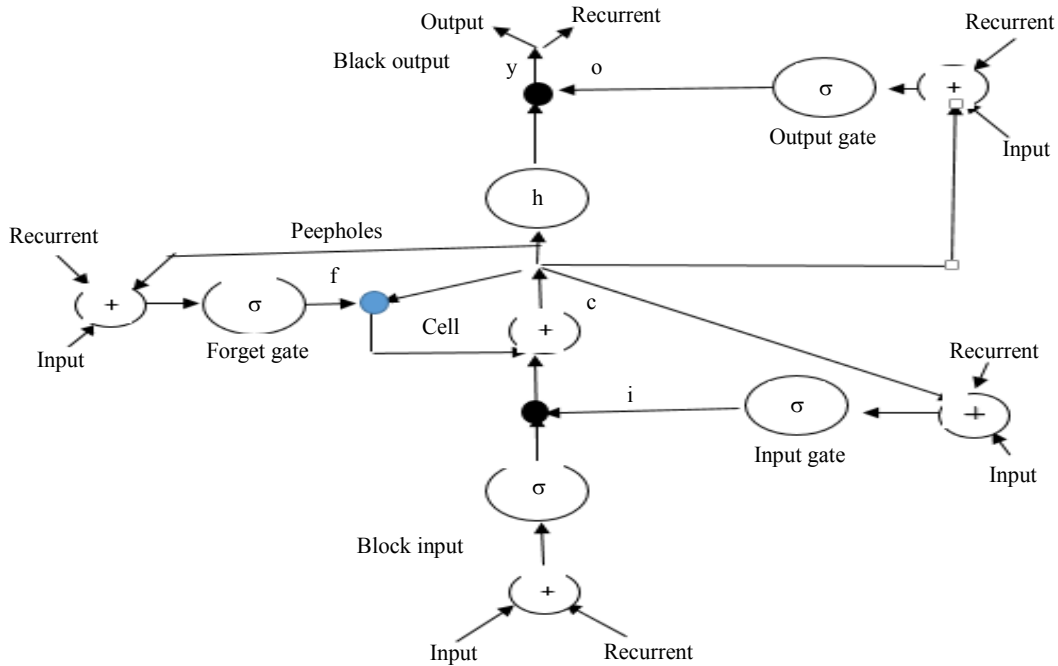


Fig. 4: Computation flow diagram of long short-term memory model (Source: Hochreiter and Schmidhuber, 1997)

In addition, the modern LSTM architecture contains peephole connections from its internal cells to the gates in the same cell to learn precise timing of the output.

In order to facilitate analysis so that the process goes smoothly, LSTM architecture is often unfolded over t (time)-dimension which can be represented by the following diagram.

Output from each LSTM cell (Fig. 4), h_t , is computed using the following formula:

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \quad (2.6)$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \quad (2.7)$$

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \quad (2.8)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \quad (2.9)$$

$$h_t = o_t \circ \sigma_h(c_t) \quad (2.10)$$

where, f_t be forget gate's activation vector; i_t be input gate's activation vector; o_t be output gate's activation vector; $W_f, W_i, W_o, U_f, U_i, U_o$ are weight matrices to be learned during model training; σ be activation function; and \circ be element-wise multiplication.

Given a weather variable as the predictor variable and another weather variable as moderating variable, the general structure of merged-LSTM model can be illustrated below.

The detail structure of the merged LSTM model are shown in Fig. 5. As can be seen from Fig. 5, the proposed model is a stacked LSTM with subsequent layers having 200, 100, 90 and 50 nodes of hidden layers. The last part of the model is a fully connected neural network with 1 output nodes.

Given a merged of two LSTMs (merged-LSTM) model. The objective function of the merged-LSTM, \mathcal{L} , can be formulated as follows:

$$\mathcal{L} = \frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2 \quad (2.11)$$

$$\hat{y}_t = \sigma\left(\sum_{i=1}^m w_i x_t + b\right) \quad (2.12)$$

$$x_t = \frac{1}{2}(h(P_t) + h(I_t)) \quad (2.13)$$

Where:

x_t = Be the input signal to Fully Connected (FC) part of merged-LSTM as the average of predicted values from each LSTM, b be bias

\hat{y}_t = Be predicted value, y_t be actual value

N = Be the total number of training samples

σ = be activation function, x_t input to FC

$h(P_t)$ = be output of LSTM-1 whose input is the predictor variable

$h(I_t)$ = be output of LSTM-2 whose input is the moderating variable(s)

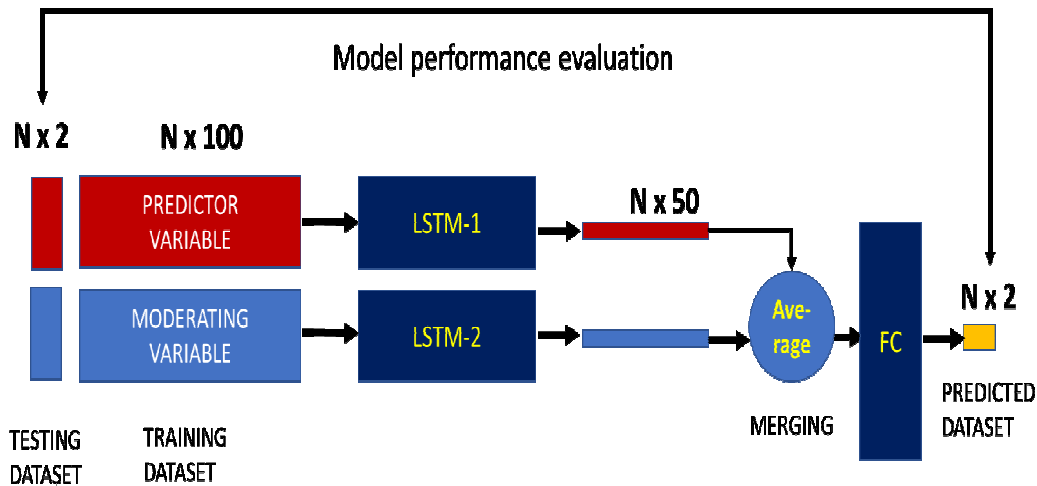


Fig. 5: The Structure of the merged-LSTM model

In this study two LSTM models and ARIMA model were explored: (1) An LSTM model as a single model which was trained using visibility time series to forecast visibility, (2) a merged-LSTM model which was trained by two time series: Visibility and intermediate variables to forecast visibility and (3) ARIMA model was trained using visibility time series to forecast visibility.

Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) is a frequently used measure of the difference between values predicted by a model and the values actually observed from the environment that is being modelled. Root mean square error is commonly used in climatology, forecasting and regression analysis to verify experimental results.

The formula is:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_{obs,i} - X_{model,i})^2}{n}} \quad (2.15)$$

where, X_{obs} is predicted values and X_{model} is modelled values at time/place i and n is the number of samples.

Model Training and Cross-Validation

In this study, the LSTM and the merged-LSTM model were supervised trained using Adam algorithm to obtain model parameter prediction that optimized a predetermined objective function. In this model training process, model cross-validation used leave-one-out technique with 70:30 proportion of training and testing datasets. The proportions of training and testing dataset are purposively set out. Model performance was measured using Root Mean Square Error (RMSE) metrics as formulated in Equation 2.15.

The ARIMA model uses 17.000 data. In this model training process, model cross-validation uses Leave-one-out technique with 70:30 proportion of training and testing datasets.

Results

The performances of each model forecasting visibility variable were summarized in Table 1.

As described in Table 1, the merged-LSTM that is used visibility as predicted variable without intermediate variable achieve lower RMSE than ARIMA model. The value of RMSE of LSTM are 0.00009 lower than ARIMA's RMSE that is 0.948.

LSTM model that is used visibility as predicted variable and dew point as moderating variable tends to achieve lower RMSE than RMSE of ARIMA which is used only visibility as the input time series. The RMSE of the mentioned model was 0.00007 while ARIMA model achieved 0.948. This value of 0.00007 is the lowest RMSE.

The lowest RMSE is also achieved by LSTM model which is used intermediate combination of temperature and dew point and combination of temperature and humidity.

Table 1 shows that the merged LSTM with or without intermediate model has lower RMSE value than ARIMA model. The merged LSTM has better performance than ARIMA.

Figure 6 shows prediction result of LSTM Model combination of temperature and dew point compare to the test (actual) time series. As described by Fig. 6, the comparison has slight deviation.

The comparison of prediction result of the ARIMA Model and the test (actual) time series is shown by Fig. 7. The figure shows that deviation between predicted and actual test data is wide.

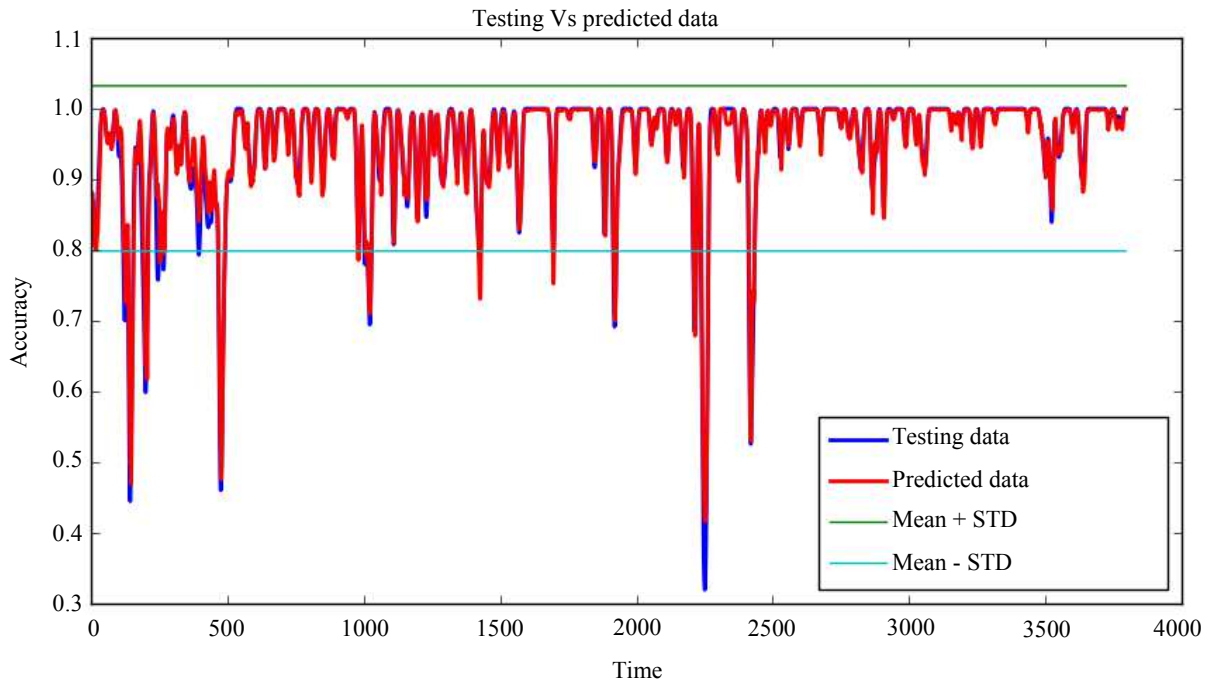


Fig. 6: Training accuracy of merged-LSTM

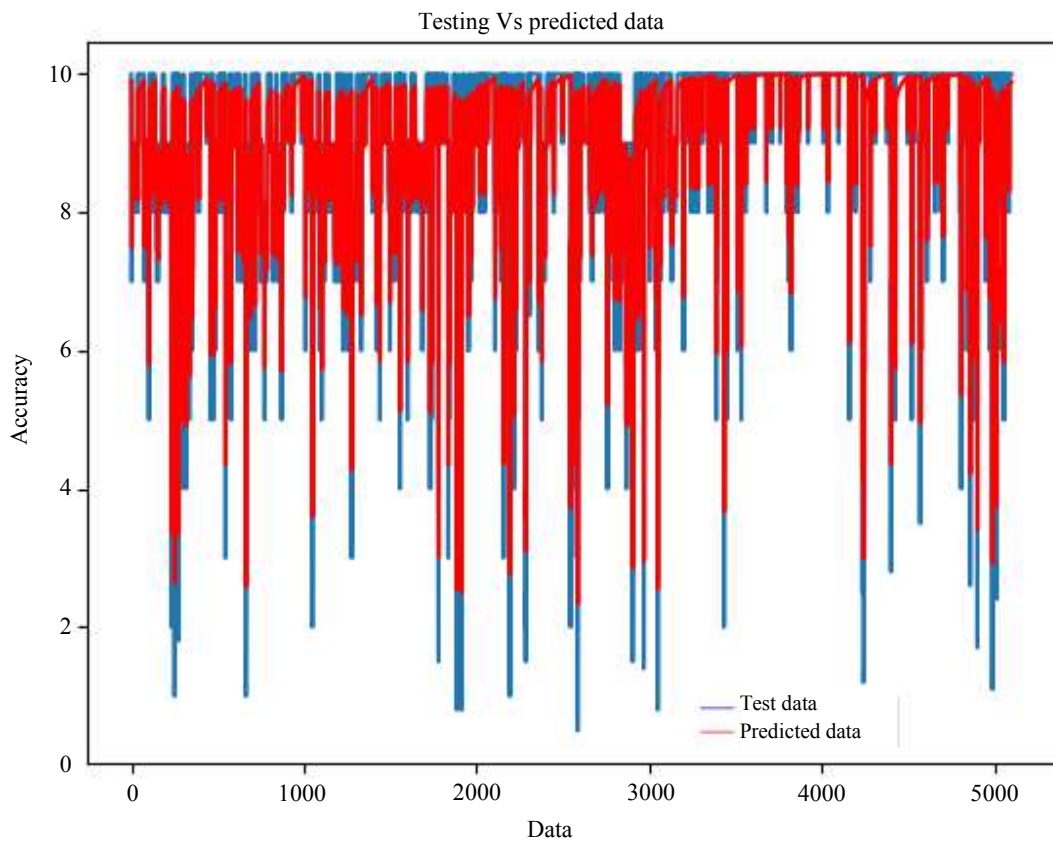


Fig. 7: Training accuracy of ARIMA mode

Table 1: Performance Evaluation Accuracy training of ARIMA, LSTM and merged-LSTM* to Forecast Visibility

Model	RMSE
ARIMA	0.948
LSTM without intermediate	0.00009
LSTM with intermediate Temp	0.00007
LSTM with intermediate Dew point	0.00009
LSTM with intermediate Humidity	0.00008
LSTM with intermediate Temp and Dew point	0.00007
LSTM with intermediate Temp and Humidity	0.00007
LSTM with intermediate Dew point and Humidity	0.00008

This study indicates that visibility forecasting by all model LSTM with intermediate variable or without intermediate variable yields better accuracy than visibility forecasting using ARIMA.

Conclusion

LSTM and ARIMA model have been used in time series forecasting. This study compares the LSTM Model and ARIMA Model in the weather forecasting. The visibility variable of the weather is used in this study because it is the most important weather variable that has big impact on all phases of flight, especially when the aircraft is maneuvering on or close to the ground LSTM model of this research set visibility as the predictor variable. Another weather such as temperature, humidity and dew point variable is used as intermediate variable.

Whereas ARIMA model forecast visibility variable only without intermediate variable. The result show that The ARIMA model yields RMSE value of 0.948. LSTM model yields the lowest RMSE value that is 0.00007. This lowest value resulted by LSTM which is used intermediate variable that are temperature variable only, combination of temperature and dew point and combination of temperature and humidity.

The LSTM model with or without intermediate variable has better performance than The ARIMA Model.

The most important findings of this research are combination of the weather variable that has the most impact on the accuracy of weather forecasting and the research artifacts (scripts and dataset). This artifacts will be available for other researchers in the same domain.

Acknowledgment

This research is supported by Doctoral of Computer Science, Bina Nusantara University.

Authors Contribution

Afan Galih Salman: Lead research project, coordinate developer, doing experiment, data analysis and writing the manuscript.

Yaya Heryadi: Advise research project, design the application, data analysis, writing manuscript, proof reading.

Edi Abdurahman: Advise research project, design the experiment, data analysis and writing manuscript.

Wayan Suparta: Advise research project, design the research methodology, data analysis, writing manuscript.

Ethics

Authors confirm that this manuscript has not been published elsewhere and that no ethical issues are involved.

Conflict of Interest Declaration

Authors declare that there is no conflict of interest regarding the publication of this manuscript.

References

- Afshin, S., H. Fahmi, A. Alizadeh and F. Kaveh, 2011. Long term rainfall forecasting by integrated artificial neural network-fuzzy logic-wavelet model in Karoon basin. *Scientific Res. Essay*, 6: 1200-1208. DOI: 10.5897/SRE10.448
- Belayneh, A. and J. Adamowski, 2012. Standard precipitation index drought forecasting using neural networks, wavelet neural networks and support vector regression. *Applied Comput. Intell. Soft Comput.*, 2012: 794061-794061. DOI: 10.1155/2012/794061
- Box, G.E.P. and G.M. Jenkins, 1976. *Time Series Analysis: Forecasting and Control*. 1st Edn., Holden-Day, San Francisco, CA, ISBN-10: 0816211043, pp: 575.
- Elkaref, M. and B. Bohnet, 2017. A simple LSTM model for transition-based dependency parsing. *arXiv preprint arXiv:1708.08959*.
- Gers, F.A., J. Schmidhuber and F. Cummins, 2000. Learning to forget: Continual prediction with LSTM. *Neural Comput.*, 12: 2451-2471. DOI: 10.1162/089976600300015015
- Grover, A., A. Kapoor and E. Horvitz, 2015. A deep hybrid model for weather forecasting. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 10-13, ACM, Sydney, NSW, Australia, pp: 379-386. DOI: 10.1145/2783258.2783275

- Hochreiter, S and J. Schmidhuber, 1997. Long short term memory, neural computation. MIT.
- Kim, S., S. Hong, M. Joh and S.K. Song, 2017. Deep rain: ConvLSTM network for precipitation prediction using multichannel radar data. Proceedings of the 7th International Workshop on Climate Informatics, Sept. 20-22.
- Lee, T.J., J. Gottschlich, N. Tatbul, E. Metcalf and S. Zdonik, 2018. A zero-positive machine learning system for time-series anomaly detection. arXiv preprint arXiv:1801.03168.
- Luo, W., W. Liu and S. Gao, 2017. Remembering history with convolutional LSTM for anomaly detection. IEEE International Conference on Multimedia and Expo, Jul. 10-14, IEEE Xplore Press, Hong Kong, China, pp: 439-444. DOI: 10.1109/ICME.2017.8019325
- Mallick, B. and G. Jain, 2017. A study of time series models ARIMA and ETS. IJ. Modern Educ. Comput. Sci., 4: 57-63. DOI: 10.5815/ijmecs.2017.04.07
- Maunder, J., R.W. Katz and A.H. Murphy, 2000. Economic value of weather and climate forecasts. Climatic Change, 45: 601-606. DOI: 10.1023/A:1005617405583
- Naseem, F., A. Rashid, T. Izhar, M.I. Khawar and S. Bano *et al.*, 2017. An integrated approach to air pollution modeling from climate change perspective using ARIMA forecasting. J. Applied Agric. Biotechnol., 2: 38-45.
- Real, E., S. Moore, A. Selle, S. Saxena and Y.L. Suematsu *et al.*, 2017. Large-scale evolution of image classifiers. arXiv preprint arXiv:1703.01041.2017
- Roesch, I. and T. Günther, 2017. Visualization of neural network predictions for weather forecasting. Eurographics Proceedings, The Eurographics Association.
- Salman, A.G., B. Kanigoro and Y. Heryadi, 2015. Weather forecasting using deep learning techniques. Proceedings of the International Conference on Advanced Computer Science and Information Systems, Oct. 10-11, IEEE Xplore Press, Depok, Indonesia, pp: 281-285. DOI: 10.1109/ICACSI.2015.7415154
- Shi, X., Z. Chen, H. Wang and D.Y. Yeung, 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Advances in Neural Information Processing Systems.
- Tektaş, M., 2010. Weather forecasting using ANFIS and ARIMA MODELS a case study for Istanbul. Environ. Res. Eng. Manage., 2010: 5-10.
- Tseng, F.M., H.C. Yub and G.H. Tzengc, 2002. Combining neural network model with seasonal time series ARIMA model. Technol. Forecast. Soc. Change, 69: 71-87. DOI: 10.1016/S0040-1625(00)00113-X
- Yoo, J.H., 2017. Large-scale video classification guided by batch normalized LSTM translator. arXiv preprint arXiv:1707.04045