

Analysis of VAE-LSTM performance in detecting anomalies in average daily temperature data in Jakarta 2000-2023



INDRI RAMDANI¹, YENNI ANGRAINI^{1*}, INDAH WATI¹

¹Statistics and Data Science Study Program, IPB University, Bogor, Indonesia

*Corresponding Author:
y_angraini@apps.ipb.ac.id

Received: 12 October 2024
Revised: 21 April 2025
Accepted: 14 May 2025

Abstract. Climate change is happening worldwide, so global climate conditions are a major concern. In densely populated urban areas such as Jakarta, it is impossible to avoid the impacts of climate change, particularly the daily changes in air temperature. Therefore, a sophisticated and efficient approach is needed to find inconsistencies in daily air temperature data to provide critical information for sustainable urban planning and efforts to reduce risks. This research will combine two innovative approaches for hybrid anomaly detection. The method combines generative methods and can extract complex features, such as variational autoencoder (VAE), along with the temporal coding capabilities of long-short-term memory, a type of Recurrent Neural Network (LSTM). The data used in this study is the average daily air temperature data in Jakarta, obtained from the Kemayoran Meteorological Station and provided by the Meteorology, Climatology, and Geophysics Agency (BMKG). The data used is daily from April 2000 to December 2023. The threshold used to detect anomalies was 229.5, which resulted in excellent performance, namely F1-Score 0.985, Recall 1.000, and Precision 0.971. The VAE-LSTM model identified all dates with significant temperature anomalies, including January 21, 2014, February 22, 2014, November 12, 2014, and February 9, 2015. These dates are significant as they represent extreme weather events that can have severe implications for urban planning and climate change adaptation. The anomalies fall into the categories of point and contextual anomalies. This study contributes to climate research by providing evidence of the effectiveness of deep learning-based hybrid models in detecting complex and context-sensitive temperature anomalies.

Keywords: Air temperature, Anomaly detection, Deep Learning, LSTM, VAE

INTRODUCTION

Anomalies are data patterns that do not conform to a well-defined concept of normal behavior. Anomalous data can occur in various aspects of life in nature, such as temperature, wind speed, rainfall, and sunspot intensity. Anomalies may result from certain activities, including system failures or sensor errors in recording information. There are various approaches to anomaly detection, one of which uses statistical-based models such as autoregressive (AR), moving average (MA), and autoregressive moving average (ARMA) models [1]. However, these models are less effective for complex data or non-linear patterns [2].

The development of science and technology has encouraged the development of various new methods that aim to overcome the limitations of conventional methods. One prominent innovation is deep learning [3]. Examples of methods that use deep learning technology are convolutional neural network (CNN), long short-term memory (LSTM), recurrent neural network (RNN),

autoencoder (AE), variational autoencoder (VAE), and stacked autoencoders (SAE).

Anomalies can be found in various aspects of life, including air temperature conditions. Global climate conditions in recent decades have become a significant concern due to climate change occurring in various parts of the world. The Earth's temperature has increased by 1.1°C and is predicted to increase by a global average of 2.8°C by 2100 [4]. Over the past 30 years (1991-2020), the temperature in the Indonesian region increased by about 0.9°C. Based on WorldClim data, generally, regions in Indonesia have temperatures between 22 - 26°C, while in Jakarta, it is between 28 - 31°C (KLHK 2020). Observation data from the Jakarta station shows a consistent increase in temperature over the past 134 years, at a rate of about 0.02°C per year [5]. The increase in average temperature that exceeds the global temperature is partly due to increased urbanization in Jakarta [6]. The impact of unusual daily air temperature variations can significantly impact people's well-being,



This is an open access article under the terms of the Creative Commons Attribution (CC-BY) License, which permits use and distribution in any medium, provided the original work is properly cited. © 2025 The Authors. Jurnal Natural published by Faculty of Mathematics and Natural Science (FMIPA), Universitas Syiah Kuala.

the environment, and city infrastructure. Therefore, Jakarta was selected as the study area due to its higher than average urban temperatures, long-term warming trend, and vulnerability to impacts of climate change. The 2000-2023 period was chosen to reflect recent patterns and extreme events relevant for anomaly detection using machine learning methods.

Various studies have been conducted to detect anomalies in air temperature data. One of them is research by [7], which analyses the occurrence of average temperature anomalies as a representation of climate change in Semarang City. The results show that the LSTM model is effective for detecting climate anomalies, especially with stacked and dense structures. LSTM is effective for modeling sequential data but may need help to capture complex input data distributions, especially in high-dimensional spaces. As a result, its ability to model sequential data is reduced, affecting its ability to capture long-term dependencies [8]. Therefore, a sophisticated and efficient approach is needed to detect daily air temperature data anomalies to provide critical information for sustainable urban planning and risk mitigation efforts.

This research is not just another study on anomaly detection. It's an innovative approach that combines two cutting-edge techniques—a hybrid anomaly detection method that combines generative techniques, and complex feature extraction capabilities through VAE, along with the temporal coding capabilities of LSTM. VAE is an unsupervised generative network model consisting of an encoder and a decoder. The encoder observes the training data distribution and generates a compression value from the training data, while the decoder reconstructs the compressed data [9]. This innovative approach promises to significantly advance our understanding and detection of anomalies in daily air temperature data.

The VAE-LSTM hybrid model can identify anomalies that may occur on various time scales. Research by [10] proved the effectiveness of an anomaly detection algorithm using the VAE-LSTM method. The results showed that the VAE-LSTM method was significantly superior to three other standard methods: the VAE, LSTM-AD, and ARMA. Additionally, research by [8] achieved RHR anomaly detection with a precision of 0.993, recall of 0.534, and F1-score of 0.693, and was able to identify Covid-19 infection in 74% of participants. Therefore, the main objective of this study is to assess the performance of the VAE-LSTM model in detecting daily average air temperature anomalies in Jakarta.

METHODOLOGY

Data

The data used in this study are secondary data of average daily air temperature in Central Jakarta obtained from Kemayoran Meteorological Station and provided by the Meteorology Climatology and Geophysics Agency (BMKG) <https://dataonline.bmkg.go.id/home>. The data used is daily data for the period 1 April 2000 to 31

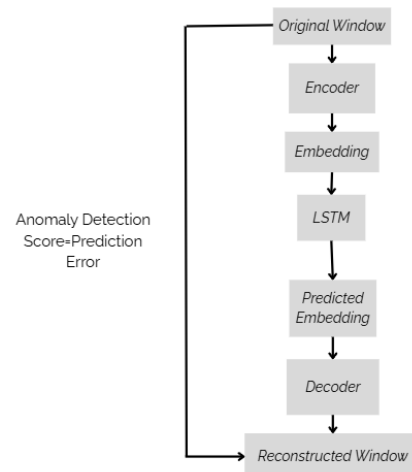


Figure 1. VAE-LSTM Modeling Structure.

December 2023, which includes both daytime and nighttime temperature observations.

Anomalies in Time Series Data

Several studies have attempted to establish the characteristics of data that exhibit anomalies. One study defines an anomaly as a pattern of data that does not conform to well-defined normal behavior. Anomalies are classified into various categories according to the type of unusual event. There are three main categories of anomalies: point anomalies, contextual anomalies, and collective anomalies [11].

Air Temperature

Air temperature is a measure of the activity of molecules in the atmosphere, reflecting the average kinetic energy of the movement of the molecules [12]. The concept of air temperature refers to the hot or cold condition of the air. According to Government Regulation Number 46 of 2012, climatological stations measure air temperature with a mercury thermometer. The thermometer is placed in a white meteorological cage 1.2-1.5 meters above the ground surface covered with grass. The air temperature data includes daily, monthly, seasonal, and annual average temperatures. The daily average temperature can be obtained by two methods, namely by adding up the maximum and minimum temperatures on that day, then dividing by two, or by recording the temperature every hour for 24 hours on that day and then taking the average [13]. The study's average temperature was the average value for 24 hours.

Variational Autoencoder (VAE)

The variational autoencoder (VAE) is a directed probability graphical model approximated by an artificial neural network. The architecture of a VAE is similar to that of an autoencoder. VAE is a variant of autoencoder that is based on Bayesian inference. In VAE, a latent concept represents the hidden representation of the data. Latent space is where points represent the distribution of features hidden in the data. The encoder, represented as $q_\phi(z|x)$, is tasked with extracting the posterior distribution of the latent variable

based on the input data z based on the input data x . While the decoder, represented as $p_\theta(x|z)$, is tasked with generating data x from the latent representation z . The parameters of the neural network, namely the encoder and decoder, are expressed by φ and θ [14]. VAE optimizes the parameters φ and θ by maximizing the evidence lower bound (ELBO) of the log-likelihood.

$$L_{vae} = -D_{KL}(q_\varphi(z|x)||p_\theta(z)) + E_{q_\varphi(z|x)}[\log p_\theta(x|z)]$$

$$\leq \log p(x|1)$$

where D_{KL} denotes the Kullback-Leibler (KL). The standard Gaussian distribution $N(0,1)$ is a common choice for the prior distribution $p_\theta(z)$ [14].

Long Short-Term Memory (LSTM)

One recurrent neural network (RNN) type that can learn long-term dependencies is LSTM or long short-term memory. The formula for each gate and cell state is as follows [14].

$$\begin{aligned} c_t &= f_t * c_{t-1} + i_t * \tilde{c}_t(2) \\ f_t &= \sigma(w_f[h_{t-1}, x_t] + b_f)(3) \\ i_t &= \sigma(w_i[h_{t-1}, x_t] + b_i)(4) \\ o_t &= \sigma(w_o[h_{t-1}, x_t] + b_o)(5) \\ h_t &= o_t * \tanh(c_t)(6) \\ \tilde{c}_t &= \tanh(w_c[h_{t-1}, x_t] + b_c)(7) \end{aligned}$$

With c is the cell state, $\tilde{c}_t$ is the candidate cell state, h is the hidden output, w are weighted in the form of a matrix of size $((t-1) \times (2t-2))$, b is the bias vector of size $((t-1) \times 1)$, x is the input. The f , i and o gates, crucial components in this context, are the forget, input, and output gates (Liu et al. 2022). σ is a sigmoid function, with $\sigma(x) = 1/(1 + e^{-x})$ [14].

Variational Autoencoder Based on Long Short-Term Memory

VAE models often fail to detect long-term anomalies because they cannot analyze data beyond a short local window. This limitation can be overcome by combining VAE as a local feature extractor with an LSTM model to handle long-term trends [10]. This combination allows the system to utilize the strengths of each model to improve the ability to detect anomalies and predict long-term trends more effectively and accurately.

Evaluation Metrics

This research will use the evaluation metrics of precision, recall, and F1-score. F1-score is a harmonic mean score weighted by precision and recall. Precision represents the proportion of actual positive predictions to total positive predictions, and recall represents the proportion of actual positive predictions to total positive actuals [9].

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (10)$$

where TP is the number of anomalous points correctly identified, FP is the number of normal points incorrectly detected as anomalous points, and FN is the number of normal points incorrectly detected as normal points [15].

Data Analysis

Figure 1 shows an illustration of the VAE-LSTM modeling process for anomaly detection. The process starts with feeding the normalized data into the VAE encoder. The encoder compresses the data into a low-dimensional latent representation called embedding. The embedding produced by the encoder becomes the input for the LSTM model, which is used to predict the next embedding. This prediction result is then used as input for the VAE decoder. The VAE decoder is responsible for reconstructing the embedding predicted by the LSTM. The prediction error is calculated by comparing the decoder's reconstructed window with the encoder's original window. The difference is used to determine the anomaly score.

Data analysis in this study uses Python software in the following stages.

1. Preprocessing the data consists of viewing and changing the type of data entered into the application programming and imputing the data. Three imputation methods were used in this research.

- a. Inversed Square Distance Method

This method uses adjacent stations to calculate the average temperature of the missing days on a single day. The equation for the Inversed Squared Distance method is as follows [16].

$$p_x = \frac{\sum_{i=1}^n \frac{p_i}{L_i^2}}{\sum_{i=1}^n \frac{1}{L_i^2}} \quad (11)$$

with p_x is the missing average temperature, p_i is the average temperature at the surrounding stations in the same period, and L_i is the distance between the stations.

- b. Mean Substitution Method

This method uses the average value of all known values to fill in the missing data [17]. This method is applied when the missing data is on one day, and the nearest station is also missing. The process is done by calculating the average of the month.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (12)$$

with $\bar{x}$ is the average, x_i is the i -th data, and n is the number of data.

- c. Forecasting with LSTM

This method is used for missing data sequentially in a time range, so that the missing values are filled with forecasting using the LSTM model.

2. Explored the data using time series plots and boxplots to determine the patterns and characteristics of the daily average temperature data.
3. Labeling the data by detecting anomalies using the Z-score method.

4. Normalize the training and test datasets using standard normalization (Z-score).

$$\bar{x}_n = \frac{x_n - \mu}{\sigma} \quad (13)$$

where μ is the mean, σ is the standard deviation, and n is the air temperature index. Normalization ensures that the data is in the right scale for inclusion in the modeling.

5. Partition the data into training data and test data. The training data contains no anomalies, and the remaining time series with anomalies are kept as test data for evaluation.
6. Conduct VAE model training.
 - a. The VAE model developed in this study uses a 4-layer convolutional encoder and a 6-layer convolutional decoder, each with 512 neurons. The training was conducted over 100 epochs with a batch size of 32 and a learning rate of 0.0001.
 - b. The training data $T = \{x_1, x_2, \dots, x_n\}$ will be divided into sequences using time windows.
 - c. The VAE model training process by creating rolling windows based on the training data. A window with the following functions.

$$w_t = [x_{t-p+1}, \dots, x_t] \quad (14)$$

w_t is a window with length p ends at time t , where x_t is the reading at time t .

- d. A training window input sequence can describe all training data as follow.

$$W = \{w_1, w_2, \dots, w_{n-p+1}\} \quad (15)$$

W It is a collection of several windows w_t .

- e. The encoder in the VAE model can take data from time windows W_i and extract the features of those windows, and then the data is stored in latent space.
- f. The compressed data passes through a decoder, reconstructing the initial time window w_t . The model can take the new time window w_t and reconstruct it into w_t' using the learned parameters of the encoder and decoder.
7. Conduct LSTM model training.
 - a. The training data is divided into k sequences of non-overlapping rolling windows. The LSTM model is created by operating on VAE embedding. The sequence of windows ending at time t is shown as follows.

$$W_t = [w_{t-(k-1) \times p}, w_{t-(k-2) \times p}, \dots, w_t] \quad (16)$$

W_t is a collection of multiple windows w , consisting of k non-overlapping windows, and each window has length p .

- b. Next comes the embedding of each window into the model.

$$E_t = [e_t^1, \dots, e_t^{k-1}] \quad (17)$$

With E_t is the corresponding embedding in W_t and e_t^i is the i -th window embedding in W_t .

- c. Computes all the embedding sequences E_t in the training set using the encoder trained from the VAE model.
- d. The training of the LSTM model is done with the embedding of the VAE. Here, the LSTM model receives an initial sequence of $(k-1)$

embeddings, and then the LSTM model predicts the next $(k-1)$ embeddings.

8. The application of the VAE-LSTM model using test data for real-time anomaly detection is as follows:

- a. The VAE-LSTM model reviews the testing sequence W_t , which consists of $k \times p$ previous readings that are tracked from time t .
- b. Estimation of embedding sequence E_t in W_t is performed with the use of an encoder from the VAE model.
- c. Prediction of the next $(k-1)$ embedding with LSTM is made by modeling the first $(k-1)$ embedding to the LSTM model.
- d. Reconstruction of the last $(k-1)$ windows ($\hat{w}_{t-(k-i) \times p}$) by using the decoder of the VAE model.

$$\hat{w}_{t-(k-i) \times p} = \text{Decoder } \hat{e}_t^i, \quad i = 2, \dots, k \quad (18)$$

- e. Calculation of the score function dt is performed to measure the goodness of the model in reconstructing these windows by calculating the Prediction error of each window. Equation dt is the sum of the squares of the difference between the reconstructed window ($\hat{w}_{t-(k-i) \times p}$) with the original window ($w_{t-(k-i) \times p}$) for each original window.

$$d_t = \sum_{i=2}^k (\hat{w}_{t-(k-i) \times p} - w_{t-(k-i) \times p})^2 \quad (19)$$

- f. Anomalies in the data at a time can be characterized by the value of the d_t

9. Evaluation of model performance through histogram plots of VAE ELBO Loss and LSTM reconstruction error histogram plots using performance metrics, namely precision, Recall, and F1-score.
10. Determination of the threshold θ contained in the score function dt by using the validation set. The best threshold θ is selected using the evaluation metrics of precision, recall, and F1-score.
11. Application of the VAE-LSTM model in real time: if the d_t exceeds the threshold θ , then an anomaly warning will be signaled $y_t = 1$.
12. Interpreting anomaly detection results involves identifying specific time points detected as anomalies, analysing their causes and impacts, and ensuring the relevance of the results in the environmental context.

RESULTS AND DISCUSSION

Data Imputation

This research data preprocessing includes imputation to fill in missing data at Kemayoran Meteorological Station. The daily average temperature data in Jakarta from 2000 to 2023 has a total of 8,765 data, but there are 28 missing data in various conditions. The first condition is that the data is missing at only one point but is available at the nearest station to the Kemayoran Meteorological Station. To overcome this, the inverse squared distance method will be used. The second condition involves missing data at one point and missing

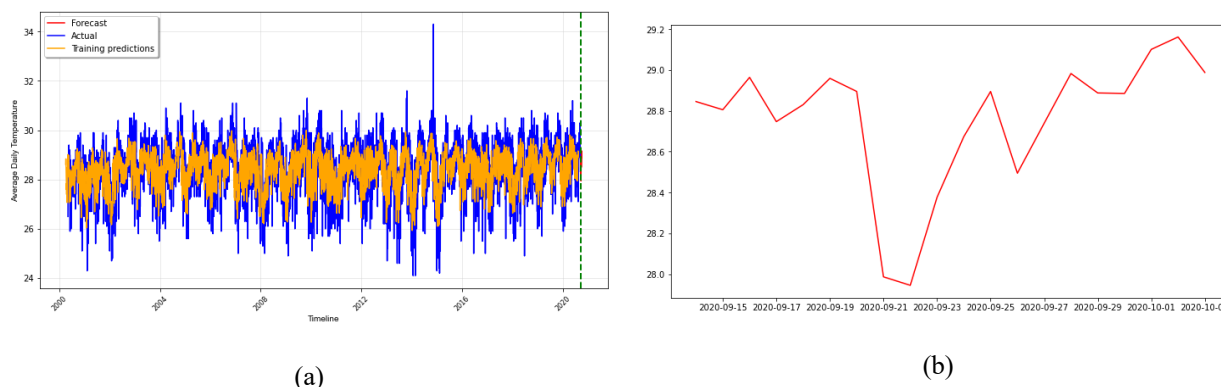


Figure 2. Comparison plot of actual data with predicted data (a); forecasting average daily temperature data (b).

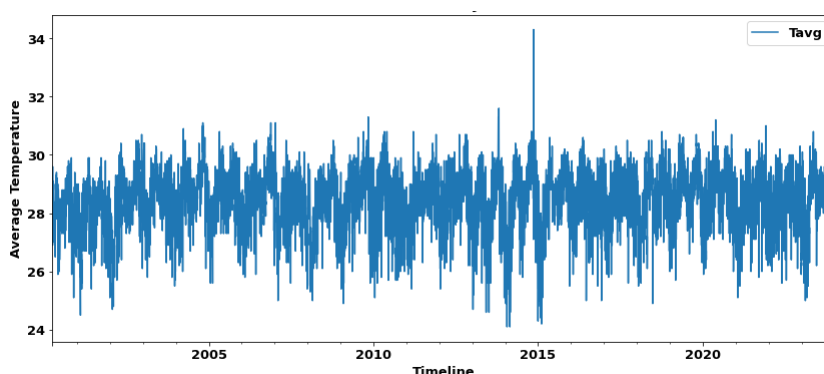


Figure 3. Time series plot of daily average temperature data in Jakarta for 2000-2023.

at all stations nearby. This condition will be overcome by using the mean substitution method. The last condition is data that is missing at all stations and missing sequentially at several periods, and the condition is solved by forecasting using the LSTM model.

The inverse squared distance method is a straightforward calculation. It involves summing the division between the average temperature and the square of the distance for each nearest station, then dividing the result by the total inverse square distance of each nearest station. This method is used to handle the 12 missing data points. For instance, to calculate the temperature at Kemayoran Station on 28/01/2015, we divide the temperature at Tanjung Priok Station (27.6°C) by the square of the distance (10^2), and the temperature at Halim Perdana Kusuma Station (27.8°C) by the square of the distance (20^2). The results of these divisions are then summed up and divided by the total of the inverse of the square of the distance of each station, resulting in 27.6°C , which is used as the average daily temperature data at Kemayoran Station on 28/01/2015.

Mean substitution will be performed for missing data at one point and missing at all nearby stations. There are three data that will be imputed using mean substitution, namely data on 12/06/2020, 09/04/2023, and 23/10/2023. The calculation of mean substitution is done by calculating the average temperature in one month. For example, the date 12/06/2020 will be imputed by calculating the average temperature during June 2020.

Forecasting using LSTM for the case of missing daily air temperature data at all stations from 15 September 2020 to 27 September 2020. The data used to perform LSTM modeling, namely daily data from 1 April 2000 to 13 September 2020 (7472 observations). This data will be divided into training data and test data with a proportion of 80:20. The training data will contain observations from 1 April 2000 to 11 August 2016 (5977 observations), while the test data will cover the period 12 August 2016 to 13 September 2020 (1495 observations). The next step is to normalize the data using MinMaxScaler into the range (0,1). After normalization, the data will be prepared to train the LSTM model and perform forecasting to fill in the missing values.

The LSTM model architecture used in this study consists of 2 to 10 LSTM layers and one dense layer, and the number of neurons per layer ranges from 32 to 512. The training process was conducted for 100 epochs with batch size 183, learning rate 0.001, and patience 10, using ADAM as the optimizer and mean squared error (MSE) as the loss function. Hyperparameter tuning was performed using RandomSearch, resulting in an optimal architecture with six layers and the number of neurons 352, 320, 512, 32, 32, and 384, respectively.

Figure 2(a) shows a comparison of actual data and predicted data, so that evaluation results for the LSTM model show a MAPE value of 2.66%, which falls into the excellent category for forecasting. Figure 2(b) shows the result of forecasting is carried out for the next 20-day period, from 14 September 2020 to 3 October 2020. The

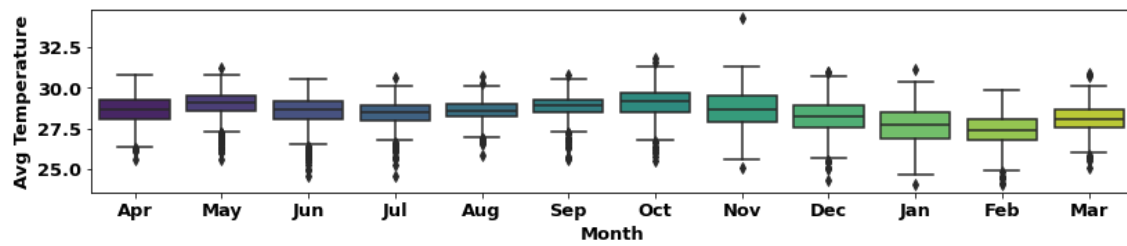


Figure 4. Boxplot of daily average temperature by month.

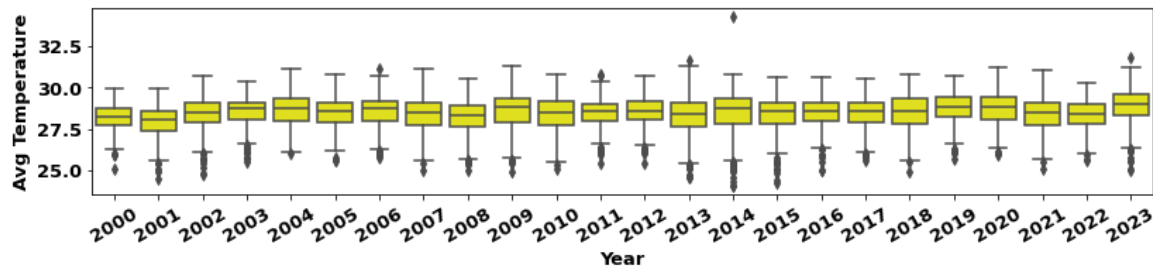


Figure 5. Boxplot of daily average temperature by year.

Table 1. Descriptive statistics of daily average temperature data

Statistics	Average Temperatures (°C)
Mean	28.46
Standar deviation	1.01
Minimum	24.10
1 st Quartile	27.90
Median	28.60
3 rd Quartile	29.20
Maximum	34.30

forecast results will be used to fill in the missing data from 15 September 2020 to 27 September 2020.

Data Exploration

Based on WorldClim data, regions in Indonesia generally have temperatures ranging from 22-26°C, while in Jakarta, it ranges from 28-31°C [5]. Figure 3 illustrates the fluctuation of daily average air temperature in Jakarta over the period 2000 to 2023, with temperatures ranging from 24°C to 34°C. There was a notable peak around 2014-2015, but in general, there was no clear upward or downward trend.

According to the Meteorological, Climatological and Geophysical Agency (BMKG), Indonesia experiences the rainy season from October to March, while the dry season occurs from April to September. This can be seen in Figure 4, which shows that April to September has a higher median than October to March, indicating higher average air temperatures during the dry season. In addition, there are some outliers that indicate the possibility of extreme values, especially in November, which has a very extreme upper outlier.

Figure 5 illustrates the variability of the median daily temperature values for each year. The years 2000 and 2001 stand out as a period where the median daily temperature tends to be lower than other years. However, in 2014, there is a point that stands out as an extreme outlier, signaling the potential for an anomaly. This potential anomaly becomes more pronounced when compared with Figure 4, which also shows indications of an abnormality in November 2014, emphasizing the need for further investigation.

Table 1 provides a statistical overview of the daily average temperature in Jakarta. The daily average temperature in Jakarta shows a mean of 28.46°C, with a minimum value of 24.19°C, a median of 28.60°C, a maximum value of 34.30°C, and a standard deviation of 1.01. The small standard deviation indicates that the data has a relatively small spread to the mean. This can be seen from Figure 3, which shows a relatively consistent average temperature. It is known that Jakarta has an average daily temperature between 28-31°C [5], so the minimum and maximum values in this research data are thought to be an anomaly.

Anomaly Detection using Z-score

Anomaly detection using a z-score is performed to determine the division of data when VAE-LSTM modeling is used, and data containing anomalies are labeled to measure the evaluation of VAE-LSTM models. When using a z-score from a univariate dataset, the threshold for outliers should be determined based on the characteristics of the entire population [18]. This study used a threshold value of 4σ . This value was chosen because the daily average temperature graph shows consistency with a small standard deviation. If 3σ is used, the data points considered anomalies are not extreme enough to be categorized as anomalies. Therefore, the threshold value used is 4σ , so the data

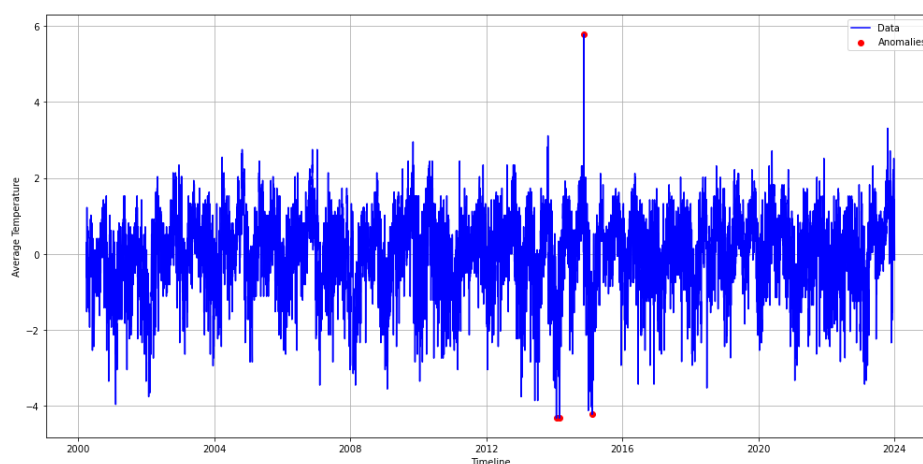


Figure 6. Anomaly detection plot using z-score

points categorized as anomalies are extreme enough. Figure 6 shows four data points categorized as anomalies, namely data on 21 January 2014, 22 February 2014, 12 November 2014, and 9 February 2015.

Training on VAE-LSTM Model

Training of the VAE-LSTM model is done in an unsupervised manner. The training data is taken from the part of the time series that does not contain anomalies, while the data containing anomalies is kept as test data. Figure 7 shows the division of data based on the z-score calculation, the training data used is from 1 April 2000 to 30 September 2013 (4931 observations). The test data was from 1 October 2013 to 31 December 2023 (3744 observations). After data division, the next step is to perform normalization using a z-score. This normalization aims to reduce the number of numerical columns in the data set to a standard scale without changing the range of values or losing data.

The VAE model, a cornerstone of our training process, is trained by dividing the training data into several sequences using non-overlapping rolling windows as visualized in Figure 8. The window length used is 24 steps, ensuring that the window shifts along the data and produces a subset of size 24. The VAE model in this study is a robust one, consisting of 4 convolutional encoder layers and 6 convolutional decoder layers, with each layer housing 512 neurons. The training process is conducted for 100 epochs with a batch size of 32, and uses a learning rate of 0.0001. The output of the VAE, a crucial part of the process, will be used as input for the LSTM model.

The encoder consists of 4 convolution layers. Each convolution layer uses a 3×3 kernel for layers 1 to 3, while layer 4 uses a 4×4 kernel. The strides used are two horizontally and one vertically. The larger strides horizontally with a value of 2 aim to efficiently down-sample data without using spatially deterministic functions such as maxpool [19]. The activation function used is Leaky ReLU, which allows gradient flow, including negative input values. The output of the encoder, latent representation, or embedding with low dimension will be the input for the LSTM model.

The Decoder section has a symmetrical shape with the encoder. The decoder consists of 6 convolution layers with strides of 1 and varying kernel sizes. The upsampling process in the decoder uses the Leaky ReLU activation function. The first layer in the decoding process will change the output shape to (32, 1, 1, 512). Next, the second layer uses a kernel size of 1×1 and changes the output shape to (32, 4, 1, 128). In layer 3, a 3×1 kernel is used and changes the output shape to (32, 8, 1, 64). Layer 4 has the same process with the same kernel size as layer 3, changing the output shape to (32, 16, 1, 32). Like before, layer five changes the output shape to (32, 32, 1, 16). Finally, layer 6 uses a kernel size 9×1 and changes the output shape to (32, 24, 1). This decoder process will produce a representation that matches the original data.

The LSTM model training process, a pivotal part of our anomaly detection process, is performed using the VAE embedding derived from the encoder process. The LSTM model, which is built based on the research of [10], uses a 3 layer LSTM with 64 neurons, and a window length order of 7. The training process is carried out for 50 epochs with a batch size of 32, and uses a learning rate of 0.0001. The optimizer used is RMSProp. The LSTM model is trained to predict the next embedding after retrieving the first embedding. After training, the LSTM is used to forecast future embeddings and assist in the detection of daily average temperature anomalies. When the LSTM prediction results are significantly different from the actual embedding, it indicates an anomaly, underscoring the importance of the LSTM model in our process.

VAE-LSTM Model Evaluation

VAE regulates the distribution of latent representations using a Gaussian distribution. This involves setting the mean and variance of the latent representation so that the model can predict the data over time. Figure 9(a) shows the distribution of anomaly scores from Jakarta's daily average temperature data. The chance density plot for the anomaly scores shows a smooth scatter with a single peak around the mean anomaly score, indicating that the model produces consistent results and performs well. This is also supported by the 25th percentile value of

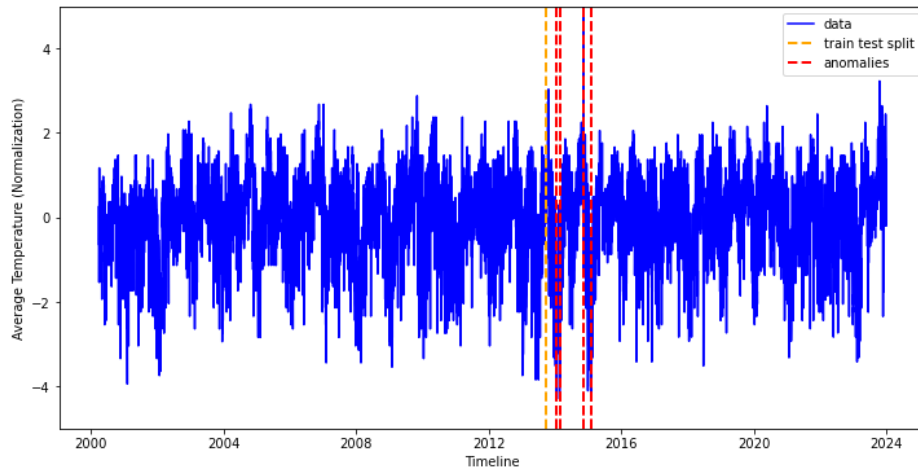


Figure 7. Splitting data after normalization.

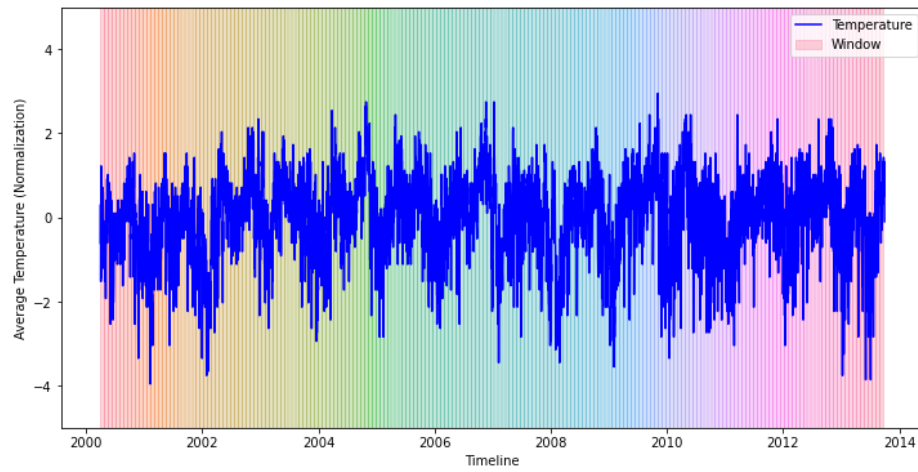


Figure 8. Training data window shift scenario using rolling window

382,963.46 and the 75th percentile value of 897,460.99, which have significant differences in value. This means the model is sensitive to data changes because it can distinguish normal and anomalous data points. The highest anomaly score of 2,025,761.63, representing 1% of the highest values in the distribution, indicates a very extreme data point.

Based on Figure 9(b), the distribution of the LSTM reconstruction errors relatively matches the Gaussian distribution of the data, with most of the reconstruction errors falling within the fitted Gaussian range. However, some reconstruction errors fall outside the fitted Gaussian range, indicating anomalies in the data. The 25th and 75th percentiles of the LSTM reconstruction error are 96.03 and 189.26, respectively. The highest anomaly score of 312.55, representing 1% of the highest values in the distribution, indicates the presence of a data point that suggests it is an anomaly.

The anomaly score in the VAE-LSTM model is calculated by labeling the anomalies in an anomaly index range. The calculation is as follows.

$$idx_{start} = idx_{anomaly_{test}[i]} - (p \times k - 1) \quad (20)$$

$$idx_{end} = idx_{anomaly_{test}[i]} + 1 \quad (21)$$

where idx_{start} is the start index that will be labeled as one, idx_{end} is the end index that will be labeled as one, $idx_{anomaly_{test}[i]}$ is the anomaly points in the test data based on the z-score method, p is the window length of 24, and k is the sequence length of 7. Equation (20) and Equation (21) show that the anomaly that will be labeled as 1 (one) starts from 167 indexes before and one index after. The initial labeling results in the z-score method obtained four anomaly points with anomaly indices [113 145 408 497]. Based on Equation (20) and Equation (21), an array is obtained that will label the indices that are anomalies into category 1 (one), listed in Table 2. Then, the labeling result on the VAE-LSTM anomaly index is the actual anomaly and will be calculated in the evaluation metrics, namely F1-Score, Recall, and Precision.

Furthermore, the best threshold is determined by calculating the evaluation metric based on the maximum and minimum range of LSTM reconstruction error on the test data. The procedure is repeated 200 times, and the result shows that the best threshold is 229.5. Excellent F1-score, recall, and precision values were obtained: an F1-score of 0.985, recall of 1.000, and precision of 0.971.

Therefore, the VAE-LSTM model shows excellent performance in anomaly detection in daily average temperature data.

Table 3 shows the evaluation metrics for the model that only uses LSTM without using the VAE model. The LSTM model produces the best threshold of 291.3, thus obtaining an F1-score of 0.982, Recall 1.000, and Precision 0.964. While the LSTM model demonstrates strong performance, the VAE-LSTM model shows superiority in terms of F1-score and Precision. This suggests that while both methods are able to detect anomalies that are also detected in the z-score method, but the LSTM model produces more new anomalies than the VAE-LSTM model. Thus, the VAE-LSTM model has a performance advantage in detecting anomalies in daily average temperature data compared to the LSTM model. This indicates that the integration of the variational autoencoder (VAE) method with LSTM is able to improve the performance of the model in anomaly detection, so that the VAE-LSTM model is more effective in identifying anomalous patterns in temperature data in Jakarta.

Anomaly Detection Using the VAE-LSTM Model

Anomaly detection using VAE-LSTM provides interesting information about the data's behavior. As shown in Figure 10, it can be seen that the VAE-LSTM model successfully detects anomalies from the anomaly window, which is at the same anomaly point as the anomaly in the z-score method. With a precision of 0.971, there are instances where the model detects anomalies that are not detected in the z-score method, as seen in Figure 10, where one window is identified as an anomaly. However, that window is not an anomaly in the z-score method.

It is known that the points containing anomalies were on 21 January 2014, 22 February 2014, 12 November 2014, and 9 February 2015. The 21 January 2014, 22 February 2014, and 9 February 2015 anomalies are classified as point anomalies because they have the lowest temperatures during the rainy season. Meanwhile, the anomaly on 12 November 2014 is classified as a contextual anomaly because it occurred during the rainy season, but the temperature was very high, reaching 34.30°C. Thus, in this study, the VAE-LSTM model is able to detect both point and contextual anomalies.

Temperatures in Jakarta are generally high throughout the year, with daily averages ranging from 28-31°C. On 21 January 2014, 22 February 2014, and 9 February 2015, the air temperature in Jakarta was very low at around 24°C. The decrease in temperature is caused by several factors, including the rainy season, lack of water vapor, the presence of radiation at night [20], and the initial presence that can block solar radiation [21]. High air pollution can also contribute to air temperature fluctuations in Jakarta.

Generally, the air temperature in Jakarta in November peaks at around 32°C, rarely falling below 30°C or rising above 34°C, with a daily low temperature of around 24°C, rarely falling below 24°C or rising above 25°C [22]. The anomaly on 12 November 2014 in Jakarta was

influenced by local weather factors and climate change that affected the overall temperature. This phenomenon can be of particular concern in predicting air temperature and detecting possible anomalies. Risk mitigation efforts can be made by monitoring temperature changes regularly and taking appropriate actions to reduce the impact.

CONCLUSION

The VAE-LSTM model has excellent performance in detecting anomalies in daily average temperature data in Jakarta. The performance of the VAE-LSTM model obtained is F1-score 0.985, recall 1.000, and precision 0.971. With such excellent performance, the VAE-LSTM model identified all dates with significant temperature anomalies, including 21 January 2014, 22 February 2014, 12 November 2014, and 9 February 2015. These anomalies fall into the categories of point and contextual anomalies. Thus, the VAE-LSTM model can detect anomalies accurately, and it shows that this approach is practical for analyzing daily average temperature data in Jakarta. This study contributes significantly by offering a deep learning based hybrid approach that outperforms conventional methods in capturing complex and context specific anomalies. The findings can support early warning system and enhance environmental monitoring for climate related risk mitigation efforts.

This research focuses on applying the method to univariate data where only point and contextual anomalies are present. Further research can test the VAE-LSTM model on datasets with different characteristics or conduct simulations to treat various types of anomalies to evaluate its flexibility and robustness. In addition, future research can integrate the VAE-LSTM model with Internet of Things (IoT) technology for real-time temperature monitoring and anomaly detection in different environments. It is also recommended to extend the study to other regions, such as areas in Eastern Indonesia, like East Nusa Tenggara (NTT), which often experience extreme temperatures, to evaluate the model's performance in diverse climatic conditions.

ACKNOWLEDGMENT

This study was supported by data provided online by the Meteorology, Climatology, and Geophysics Agency (BMKG) of Indonesia, accessible via <https://dataonline.bmkg.go.id>. The authors gratefully acknowledge BMKG for making these data publicly available, which significantly contributed to the analysis and findings of this research.

REFERENCE

- [1] Borges, H.; Akbarinia, R.; and Masegla, F. 2021. Anomaly detection in time series *Transaction on Large-Scale Data- and Knowledge-Centered System L*. Berlin: Springer Science and Business Media Deutschland GmbH. p. 46-62.

- [2] Kundu, A.; Sahu, A.; Serpedin, E.; Davis, K. 2020. A3D: Attention-based auto-encoder anomaly detector for false data injection attacks. *Electric Power Syst. Res.* **189** 1-7. DOI:10.1016/j.epsr.2020.106795.
- [3] Liu, R.; Jiang, Y.; Lin, J. 2022. Forecasting the volatility of specific risk for stock with LSTM. *Procedia Comput. Sci.* **202** 111-114. DOI:10.1016/j.procs.2022.04.015.
- [4] Lee, H.; Calvin, K.; Dasgupta, D.; Krinner, G.; Mukherji, A.; Thorne, P.; Trisos, C.; Romero, J.; Aldunce, P.; Barret, K.; et al. 2023. *Climate Change 2023 Synthesis Report*. Geneva: IPCC. DOI:10.59327/IPCC/AR6-9789291691647.
- [5] KLHK. 2020. *Roadmap Nationally Determined Contribution (NDC) Adaptasi Perubahan Iklim*. Jakarta (ID): Kementerian Lingkungan Hidup dan Kehutanan Republik Indonesia.
- [6] Siswanto, S.; van Oldenborgh, G.J.; van der Schrier, G.; Jilderda, R.; van den Hurk, B. 2015. Temperature, extreme precipitation, and diurnal rainfall changes in the urbanized Jakarta city during the past 130 years. *Int. J. Climatol.* **36**(9) 3207-3225. DOI:10.1002/joc.4548.
- [7] Wahyono, T.; Heryadi, Y.; Soeparno, H.; Abbas, B.S. 2020. Anomaly detection in climate data using stacked and densely connected long short-term memory model. *J. Comput.* **31**(4) 42-53. DOI:10.3966/199115992020083104004.
- [8] Abir, F.F.; Chowdhury, M.E.H.; Tapotee, M.I.; Mushtak, A.; Khandakar, A.; Mahmud, S.; Hasan, A.; 2023. PCovNet+: a CNN-VAE anomaly detection framework with LSTM embeddings for smartwatch-based covid-10 detection. *Eng. Appl. Artif. Intell.* **122** 1-15. DOI:10.1016/j.engappai.2023.106130.
- [9] Zhao, Y.; Zhang, X.; Shang Z.; Cao, Z. 2022. DA-LSTM-VAE: dual-stage attention-based LSTM-VAE for KPI anomaly detection. *Entropy*. 1-17. DOI:10.3390/e24111613.
- [10] Lin, S.; Clark, R.; Birke, R.; Schonborn, S.; Trigoni, N.; Roberts, S. 2020. Anomaly detection for time series using VAE-LSTM hybrid model. *ICASSP*. 4322-4326. DOI: 10.1109/ICASSP40776.2020.9053558.
- [11] Chandola, V.; Banerjee, A.; Kumar, V. 2009. Anomaly detection: a survey. *ACM Comput. Surv.* **14**(3) 1-58. DOI:10.1145/1541880.1541882.
- [12] Fadholi, A. 2013. Uji perubahan rata-rata suhu udara dan curah hujan di Kota Pangkalpinang. *Jurnal Matematika, Sains, dan Teknologi*. **14**(1) 11-25. DOI:10.33830/jmst.v14i1.309.2013.
- [13] Winarno, G.D.; Harianto, S.P.; Santoso, T. 2012. *Klimatologi Pertanian*. (Bandar Lampung: Pusaka Media)
- [14] Han, P.; Ellefsen, A.L.; Li, G.; Holmeset, F.T.; Zhang, H. 2021. Fault detection with LSTM-based variational autoencoder for maritime components. *IEEE Sens. J.* **21**(19) 21903-21912. DOI: 10.1109/JSEN.2021.3105226
- [15] Niu, Z.; Yu, K.; Wu, X. 2020. LSTM based VAE-GAN for time series anomaly detection. *Sensors*. **20**(3738) 1-13. DOI:10.3390/s20133738.
- [16] Rizky H.; Nasution, Y.N.; Goejantoro, R. 2019. Analisis data curah hujan yang hilang menggunakan metode inversed square distance. *Seminar Nasional Matematika, Statistika, dan Aplikasinya*. Samarinda; p. 138-149.
- [17] Pamuji, F.Y.; Muslikh, A.R.; Arief, R.M.; Muti, D. 2024. Komparasi metode mean dan KNN imputation dalam mengatasi missing value pada dataset kecil. *Jurnal Informatika Polinema*. **10**(2) 257-264. DOI: 10.33795/jip.v10i2.5031.
- [18] Yoseph, F.; Markku, H.; Howard, D. 2019. Outliers identification model in point-of-sales data using enhanced normal distribution method. *IEEE*. 72-78. DOI:10.1109/iCMLDE49015.2019.00024.
- [19] Maulana, A.; Fatichah, C.; Suciati, N. 2020. Facial inpainting pada citra wajah unaligned menggunakan generative adversarial network dengan feature reconstruction loss. *Jurnal Ilmiah Teknologi Informasi*. **18**(2) 171-178. DOI: 10.12962/j24068535.v18i2.a1004.
- [20] Fallahnda, B. 2023. Kenapa Cuaca Hari Ini Dingin Sekali dan Penjelasan Lengkap BMKG. [Online]. Available at: <https://tirta.id/kenapa-cuaca-hari-ini-dingin-sekali-dan-penjelasan-lengkap-bmkg-gNb7>. [accessed 22 Apr. 2024].
- [21] Mutiarasari K. 2023. Faktor-faktor Penyebab Cuaca Dingin, Apa Saja?. Available at: <https://news.detik.com/berita/d-6534269/faktor-faktor-penyebab-cuaca-dingin-apa-saja>. [accessed 22 Apr. 2024].
- [22] Weather Spark. 2024. Cuaca November di Jakarta Indonesia. Available at: <https://id.weatherspark.com/m/116847/11/Cuaca-Rata-rata-pada-bulan-November-in-Jakarta-Indonesia>. [accessed 22 Apr. 2024].

How to cite this article:

Ramdani, I.; Angraini, Y.; Indahwati. (2025). Analysis of VAE-LSTM performance in detecting anomalies in average daily temperature data in Jakarta 2000-2023. *Jurnal Natural*, 25(2), 64-73.