International Proceeding
Universitas Tulungagung
2025

# RAINFALL PREDICTION IN SURABAYA USING TABNET MODEL BASED ON RADIOSONDE DATA

**Annabel Gracia Puryani[1]*, Aviolla Terza Damaliana[2], Dwi Arman Prasetya[3]**

[1,2,3] Universitas Pembangunan Nasional Veteran Jawa Timur, Indonesia
annabelgraciapuryani@gmail.com

**Abstract**

This study aims to predict daily rainfall in Surabaya using TabNet, a deep learning model for tabular data, based on radiosonde atmospheric parameters and official BMKG rainfall data. The radiosonde data include LI, KI, TT, CAPE, SWEAT, and SI measured at 00.00 and 12.00, while daily rainfall data were collected from a local station. All data were pre-processed to handle missing values, normalized, and used to construct relevant features for modeling. The dataset was divided into training and testing sets, and the TabNet model was trained to capture non-linear relationships between atmospheric parameters and rainfall. Model performance was evaluated using RMSE, MAE, MAPE, and $R^2$, achieving 6.58 mm, 3.14 mm, 17.69%, and 0.54, respectively, indicating good predictive ability. The predicted monthly rainfall for 2025 reflects seasonal patterns consistent with Surabaya's tropical climate, with higher rainfall at the beginning and end of the year and lower values in mid-year months. The results demonstrate that TabNet can effectively capture seasonal rainfall trends, providing practical benefits for flood mitigation planning, water resource management, and early warning systems. This study highlights the potential of combining radiosonde data and deep learning for accurate rainfall prediction in tropical regions.

**Keywords:** *rainfall prediction, TabNet, radiosonde data, tropical climate, deep learning*

## INTRODUCTION

Rainfall is a climate element that significantly impacts various sectors, including agriculture, water resources management, transportation, and hydrometeorological disaster mitigation (1). As a tropical region, Surabaya has rainfall patterns influenced by the Asia-Australia monsoon system, as well as local interactions with topography and the Java Sea. Therefore, accurate rainfall predictions are crucial to support early warning and regional governance planning (2).

Vertical atmospheric parameters from radiosonde data, such as Convective Available Potential Energy (CAPE), K Index (KI), Lifted Index (LI), and SWEAT Index, have been shown to have a strong relationship with the potential for convective rainfall (3). Utilizing this data in rainfall modeling provides a more comprehensive perspective than surface data alone.

Modern deep learning approaches, particularly TabNet, enable efficient tabular data analysis with an attention mechanism that selects important features at each prediction stage, making them suitable for handling multi-variable and correlated radiosonde data (4)(5). However, rainfall predictions often face seasonal bias, particularly during the monsoon transition phase. Therefore, corrections based on local climatological patterns are necessary to achieve more realistic predictions (6).

The objective of this study was to develop a TabNet-based rainfall prediction model utilizing radiosonde parameters and to evaluate model performance using RMSE, MAE, MAPE, and $R^2$ metrics. Furthermore, this study also aimed to produce a 2025 rainfall prediction corrected to reflect Surabaya's seasonal patterns. This would make the predictions more realistic and usable for supporting regional governance planning and hydrometeorological disaster mitigation.

## LITERATURE REVIEW

Rainfall prediction has been extensively studied using various methods, ranging from classical statistical approaches to machine learning. For example, Yuan et al. (2020) used a deep learning model to predict daily rainfall using a hydrometeorological dataset, demonstrating that nonlinear methods capture rainfall dynamics better than linear regression methods (7). Another study by Putra & Lestari (2022) utilized radiosonde parameters to predict convective rainfall in tropical regions, finding that atmospheric stability indices such as CAPE, KI, LI, and Total Totals (TT) significantly contribute to model accuracy (3) (8).

Following the development of deep learning, the TabNet method was introduced by Arik & Pfister (2021) as a model that incorporates sequential attention for automatic feature selection and handles nonlinear interactions between tabular variables (4). Nguyen et al. (2023) extended the application of TabNet to climatological data, demonstrating that this model outperforms LSTM and Random Forest models in rainfall prediction on multi-variable datasets (5). Hermawan et al. (2021) also reported that the use of vertical atmospheric data improves the accuracy of rainfall predictions in tropical regions with high seasonal variation (9).

However, seasonal bias remains a challenge, as models tend to underestimate rainfall during the peak season and overestimate rainfall during the dry season. To address this, a seasonal bias correction approach that references official rainfall intensity from the Meteorology, Climatology, and Geophysics Agency (BMKG) has proven effective in aligning predictions with local climatological patterns, as reported by BMKG (2022) and Hersbach et al. (2020) (6) (10).

## RESEARCH METHOD

This study uses a quantitative approach to predict daily rainfall in Surabaya based on radiosonde data and official rainfall data from the Meteorology, Climatology, and Geophysics Agency (BMKG) (10). The radiosonde data used includes atmospheric parameters such as LI, KI, TT, CAPE, SWEAT, and SI at observation times of 00:00 and 12:00. Rainfall data (CH) was obtained from BMKG stations and recorded daily. All data was collected in CSV format and then preprocessed to ensure data quality, including cleaning for missing values, normalization, and generating relevant features for modeling(110).

The processed data was then divided into a training set and a testing set. Modeling was performed using TabNet, a deep learning architecture for regression capable of capturing non-linear relationships between atmospheric parameters and rainfall. TabNet's advantage lies in its attention mechanism, which allows the model to adaptively select features at each processing stage, thereby improving interpretability and prediction accuracy.

Mathematically, each input data set is processed through several decision steps. At each step, TabNet generates a feature transformer $h_t = f_t(x_t)$ and an attentive mask $M_t = softmax\left(A_t(x_t)\right)$, which determines the features selected for the next step. The partial predictions from each step are summed to produce the final output (10):

$$\hat{y} = \sum_{t=1}^{T} linear\ (h_t \odot M_t\ )$$

In the context of rainfall regression, $\hat{y}$ represents the predicted daily rainfall. The model was trained using the Mean Squared Error (MSE) as the loss function. Model performance was evaluated using RMSE, MAE, and the coefficient of determination ($R^2$) metrics on the testing data to assess the prediction accuracy and TabNet's ability to capture seasonal rainfall patterns in the Surabaya region.

Additionally, a CSV data summary table can be displayed to show atmospheric parameters per observation time and rainfall data per date, so that readers can see the number of observations and the rainfall categories analyzed.

**Table 1 Dataset**

| Tanggal | SWEAT 00.00 | SWEAT 12.00 | . . . | SI 12.00 | TT 00.00 | TT 12.00 |
|---|---|---|---|---|---|---|
| **01/01/19** | 227,2 | 227,5 | . . . | 0 | 43,2 | 43 |
| **02/01/19** | 206,1 | 237,3 | . . . | 0 | 43,1 | 43,9 |
| . | . | . | | . | . | . |
| . | . | . | . . . | . | . | . |
| . | . | . | | . | . | . |
| **30/12/24** | 220,9 | 230,2 | . . . | 0,9 | 41,7 | 41,4 |
| **31/12/24** | $x_{1,2192}$ | $x_{2,2192}$ | . . . | $x_{10,2192}$ | $x_{11,2192}$ | $x_{12,2192}$ |

## RESULT AND DISCUSSION

The TabNet model evaluation results demonstrated a fairly good predictive ability for modeling daily rainfall in Surabaya based on radiosonde data and official rainfall from the Meteorology, Climatology, and Geophysics Agency (BMKG). Testing was conducted on test data not yet used during training, and model performance was assessed using several standard metrics. The evaluation results showed that the model had a Root Mean Square Error (RMSE) of 6.58 mm, a Mean Absolute Error (MAE) of 3.14 mm, and a Mean Absolute Percentage Error

(MAPE) of 17.69%, with a coefficient of determination ($R^2$) of 0.54. The relatively low RMSE and MAE values indicate that the model's predictions closely match daily observations, while the $R^2$ value indicates that approximately 54% of the rainfall variation can be explained by the TabNet model. Therefore, this model is quite effective in capturing the non-linear relationship between atmospheric parameters and rainfall, while also being able to adjust for seasonal patterns in tropical regions like Surabaya.

Before presenting the monthly prediction results, it is important to clarify that rainfall intensity is classified into specific categories to facilitate interpretation and communication of the results. The rainfall categories used in this study include:

**Table 2 Rain Type**

| Rain Type | Intensity (mm/hour) |
|---|---|
| No Rain | 0 |
| Light Rain | 1 mm – 5 mm |
| Moderate Rain | 5 mm -10 mm |
| Heavy Rain | 10 mm - 20 mm |
| Very Heavy Rain | >20 mm |

This classification serves as a reference for assessing whether the monthly predictions from the TabNet model align with the climatological patterns of the Surabaya region, particularly in capturing the seasonal variations that typically occur between the rainy and dry seasons. The monthly rainfall predictions for 2025, after evaluation and seasonal correction, are shown in the following table:

**Table 3 Result Prediction**

| Month | Prediction |
|---|---|
| January 2025 | 12,05 |
| February 2025 | 7,65 |
| March 2025 | 6,05 |
| April 2025 | 2,52 |
| May 2025 | 0 |
| June 2025 | 0 |
| July 2025 | 0 |
| August 2025 | 0 |
| September 2025 | 0 |
| October 2025 | 2,30 |
| November 2025 | 5,14 |
| December 2025 | 10,77 |

These monthly predictions illustrate a clear seasonal trend, with the early and late months of the year experiencing relatively higher rainfall intensity compared to the middle months. The table shows that January, February, and December exhibit heavy to moderate rainfall, while May through September experience light to moderate rainfall, consistent with the dry season pattern. Transition months, such as March, April, and October, have moderate to light rainfall predictions, indicating a transition from the rainy to the dry season and vice versa. These results demonstrate that TabNet is not only capable of producing quantitative predictions that closely match observations but also captures seasonal rainfall trends in tropical regions.

## CONCLUSION

Based on the research results, it can be concluded that the TabNet model is capable of predicting daily rainfall in Surabaya with fairly good accuracy. Model evaluation showed an RMSE of 6.58 mm, MAE of 3.14 mm, MAPE of 17.69%, and $R^2$ of 0.54, indicating that the model adequately captures seasonal trends and the non-linear relationship between radiosonde atmospheric parameters and rainfall. Monthly predictions for 2025 also show a seasonal pattern consistent with tropical climatology, with the early and late months of the year tending to have moderate to heavy rainfall, while the middle months are relatively light. These results demonstrate that using TabNet for rainfall prediction based on radiosonde data is an effective and realistic method.

Practically, this research has several implications. First, the resulting rainfall predictions can be used by relevant parties, such as the BMKG (Meteorology, Climatology, and Geophysics Agency), local governments, and water resource managers, for flood risk mitigation planning, irrigation management, and early warning system development. Second, the seasonal patterns detected through this prediction can help communities and the agricultural sector adapt their activities to predicted rainfall conditions.

However, this study has several limitations. One is the use of data from a single BMKG station in Surabaya, so the predictions may not be representative of microclimate variations across the wider region. Furthermore, the model only uses radiosonde atmospheric parameters at 00:00 and 12:00, thus ignoring other intraday variations. External factors such as long-term climate change, El Niño or La Niña phenomena, and air quality data that can influence rainfall are also not included in the model.

Based on these limitations, several suggestions for future research are available. Future research could expand the data coverage by including several BMKG stations in the Surabaya or East Java areas to increase the representativeness of the predictions. Furthermore, it could utilize radiosonde data with a higher observation frequency or add additional variables such as humidity, wind, or satellite data to improve model accuracy. Further research could also explore combining TabNet with other methods, such as ensemble learning or long-term prediction models, to better capture seasonal trends and rainfall extremes.

Thus, this research makes important scientific and practical contributions to the fields of meteorology and disaster risk management, while also opening up opportunities for the development of more accurate rainfall prediction models in the future.

## REFERENCES


Arik, S. O., & Pfister, T. (2021). TabNet: Attentive interpretable tabular learning. arXiv preprint arXiv:1908.07442. https://arxiv.org/abs/1908.07442

BMKG. (2022). Data curah hujan dan iklim Indonesia. Badan Meteorologi, Klimatologi, dan Geofisika. https://www.bmkg.go.id

Putra, A. D., & Lestari, S. (2022). Using radiosonde atmospheric parameters to predict convective rainfall in tropical regions. Journal of Atmospheric Research, 280, 106–119. https://doi.org/10.1016/j.atmosres.2022.106119

Nguyen, T., Tran, V., & Pham, H. (2023). Deep learning for climatological data: TabNet applications in rainfall prediction. Environmental Modelling & Software, 165, 105–127. https://doi.org/10.1016/j.envsoft.2023.105127

Yuan, L., Chen, Y., & Li, X. (2020). Deep learning approach for daily rainfall prediction using hydrometeorological datasets. Water, 12(7), 1945. https://doi.org/10.3390/w12071945

Hersbach, H., Bell, B., Berrisford, P., et al. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hermawan, I., Saputra, R., & Wulandari, D. (2021). Impact of vertical atmospheric data on rainfall prediction accuracy in tropical regions. International Journal of Climatology, 41(12), 5552–5567. https://doi.org/10.1002/joc.7185



Hermawan, I., Saputra, R., & Wulandari, D. (2021). Impact of vertical atmospheric data on rainfall prediction accuracy in tropical regions. International Journal of Climatology, 41(12), 5552–5567. https://doi.org/10.1002/joc.7185



Samet, R., & Shojaei, A. (2020). Mathematical modeling and optimization in civil engineering: A review. Journal of Civil Engineering and Management, 26(5), 417–432. https://doi.org/10.3846/jcem.2020.12635

Shang, Y., & Abourizk, S. M. (2018). Modeling construction project decision processes using digital simulations. Automation in Construction, 94, 75–84. https://doi.org/10.1016/j.autcon.2018.06.005

Sharma, A., & Tiwari, M. K. (2019). Linear programming techniques in construction management. Journal of Construction Engineering and Management, 145(2), 04018132. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001586

Taha, H. A. (2017). Operations Research: An Introduction (10th ed.). Pearson Education.

Zavadskas, E. K., Turskis, Z., & Kildienė, S. (2016). Multi-criteria decision-making (MCDM) methods in civil engineering: Part I–A state-of-the-art survey. Archives of Civil and Mechanical Engineering, 16(1), 1–13. https://doi.org/10.1016/j.acme.2015.10.003