Comparison of Simple and Multiple Linear Regression Models for Monthly Rainfall Prediction in Batam Using Temperature and Relative Humidity Predictors

Vivi Rianti[(1)], Dzikrullah Akbar[(2)*], Ari Paskalis Stefanus Hutagalung[(3)], Sofy Nabilah Balqis Eka Sugiarto[(4)]

[1,2,3,4]State College of Meteorology, Climatology, and Geophysics (STMKG)

[1,2,3,4]Jl. Meteorologi No.5, Kota Tangerang, Banten 15221

e-mail: viviriantisbm@gmail.com[(1)], dzikrullah.akbar@stmkg.ac.id[(2)*], arihutagalung1603@gmail.com[(3)], cintasofy554@gmail.com[(4)]

ABSTRAK

Batam merupakan wilayah pesisir strategis yang terletak di jalur pelayaran internasional dan memiliki iklim tropis dengan intensitas curah hujan yang tinggi. Kondisi tersebut secara signifikan memengaruhi berbagai sektor penting, seperti transportasi dan pariwisata, sehingga diperlukan prediksi klimatologis yang akurat. Penelitian ini mengembangkan model regresi linier sederhana dan regresi linier berganda untuk curah hujan bulanan dengan menggunakan suhu udara dan kelembapan relatif sebagai variabel prediktor. Data yang digunakan merupakan hasil pengamatan di Stasiun Meteorologi BMKG Hang Nadim selama periode 2000 - 2024. Kinerja model dievaluasi dengan membandingkan nilai prediksi terhadap data observasi tahun 2024 menggunakan *Root Mean Square Error* (RMSE) dan koefisien korelasi Pearson (r). Hasil penelitian menunjukkan bahwa model regresi linier berganda memberikan akurasi prediksi tertinggi, dengan nilai RMSE sekitar ± 74 mm dan koefisien korelasi sebesar $r = 0.855$. Diantara model yang menggunakan satu prediktor, kelembapan relatif menunjukkan kinerja yang lebih baik dibandingkan suhu, dengan RMSE sebesar ± 85 mm ($r = 0.759$) dibandingkan RMSE sebesar ± 138 mm ($r = 0.737$). Fakta tersebut menunjukkan bahwa integrasi prediktor suhu dan kelembapan relatif lebih efektif untuk memproyeksikan curah hujan bulanan, berdasarkan analisis data klimatologi selama 25 tahun di wilayah tropis maritim seperti Batam. Peningkatan prediksi curah hujan tersebut mendukung strategi adaptasi iklim dan mitigasi bencana yang lebih tangguh di Batam.
Kata kunci: Curah Hujan; Suhu ; Kelembapan; Regresi Linear; Batam

ABSTRACT

Batam is a strategic coastal region located along an international shipping route, characterized by a tropical climate with high rainfall intensity. These conditions substantially affect key sectors such as transportation and tourism, underscoring the need for accurate climatological predictions. This study develops simple and multiple linear regression models for monthly rainfall, incorporating air temperature and relative humidity as predictor variables. The dataset comprises observations recorded at the BMKG Hang Nadim Meteorological Station for the period 2000–2024. Model performance was evaluated by comparing predicted values with observed data for 2024 using the Root Mean Square Error (RMSE) and the Pearson correlation coefficient (r). The results show that the multiple linear regression model provides the highest predictive accuracy, with an RMSE of approximately ± 74 mm and a correlation coefficient of $r = 0.855$. Among the single-predictor models, relative humidity performed better than temperature, achieving an RMSE of ± 85 mm ($r = 0.759$) compared with ± 138 mm ($r = 0.737$). These findings indicate that the integration of temperature and relative humidity predictors is more effective for monthly rainfall prediction, based on the analysis of a 25-year climatological dataset from a tropical maritime region such as Batam. This improvement in rainfall prediction supports more robust climate adaptation and disaster mitigation strategies in Batam.
Keywords: *Rainfall; Temperature; Humidity; Linear Regression; Batam*

INTRODUCTION

Rainfall is one of the climatological parameters that significantly contributes to various hydrological processes and socio-economic activities [1], [2]. Rainfall variability in tropical regions is influenced by complex interactions between atmospheric and oceanic processes involving various climatological parameters [3], [4]. Among these parameters, air temperature and relative humidity are key predictors of rainfall due to their strong influence on water vapor availability, condensation, cloud formation, and subsequent precipitation processes [4], [5]. In tropical coastal and archipelagic regions, maritime influences and climate variability make rainfall highly sensitive to changes in air temperature and relative humidity [3], [6], [7].

Batam City, located in the Kepulauan Riau Province, is a tropical coastal area situated along a strategic international shipping route. The region has experienced rapid growth in industrial, commercial, and tourism activities [8], [9]. These conditions make rainfall stability an important factor for various socio-economic activities in the region [6]. High or unpredictable rainfall may affect transportation, outdoor activities, and water resource management [1], [10], [11]. Therefore, integrating air temperature and relative humidity into multivariable statistical models can improve the accuracy of rainfall predictions, thereby supporting adaptive planning and climate risk mitigation in Batam City.

Linear regression is one of the most widely used approaches for rainfall prediction because it can quantitatively model the relationship between rainfall and climatological parameters [5], [12], [13]. Under atmospheric conditions influenced by multiple variables, multiple linear regression (MLR) generally provides better representation and prediction accuracy than simple linear regression (SLR) [14]–[16]. Although similar studies have been conducted in several regions of Indonesia [5], [13]–[18], studies comparing SLR and MLR models for rainfall prediction using long-term climatological data remain scarce, particularly in Batam City. Furthermore, evaluations that simultaneously consider prediction error metrics and the strength of statistical relationships are still rarely reported.

Based on these considerations, this study aims to develop and compare simple and multiple linear regression models for predicting monthly rainfall in Batam City. The SLR models use either air temperature or relative humidity as the predictor variable, whereas the MLR model integrates both variables simultaneously as predictors. The performance of all models is evaluated using Root Mean Square Error (RMSE) and the Pearson correlation coefficient to determine the model with the highest prediction accuracy. This study provides new insights into the relative predictive roles of temperature and relative humidity in monthly rainfall forecasting based on 25 years of climatological observations from Batam, a tropical maritime environment that remains insufficiently explored in Indonesia.

METHOD

Study Area

This study was conducted in the Batam area, particularly around Hang Nadim International Airport, which functions as a major air transportation center in the Riau Islands. The area serves as an important gateway for tourism and business activities in Batam City. Consequently, weather dynamics and rainfall variability in this region directly influence transportation activities, tourism, and local communities. In addition, the study area is situated within a tropical coastal environment influenced by atmosphere–ocean interactions, resulting in distinctive climatic characteristics with relatively high rainfall variability.

The study area is located near major transportation activity centers and rapidly developing coastal tourism areas. These geographical conditions make the area representative of Batam City's microclimatic characteristics, particularly with respect to weather variability and rainfall patterns in tropical coastal environments. Such characteristics also make this area important for climatological studies and the development of rainfall prediction models that support various strategic sectors in Batam City. Furthermore, the location of the study area in Batam City is presented in Figure 1.

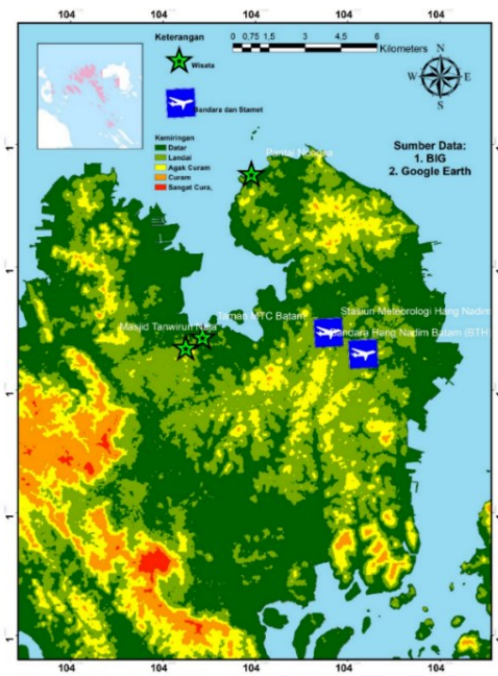

Figure 1. Study area map

Data Collection and Preprocessing

This study used secondary climatological data provoded by the Indonesia Agency of Meteorology, Climatology, and Geophysics (BMKG) and accessed via the BMKG Data Online Portal (https://dataonline.bmkg.go.id/). The observational data were collected from Hang Nadim Meteorological Station over a 25-year period from 2000 to 2024. The climatological parameters analyzed in this study consisted of total monthly rainfall, monthly average air temperature, and monthly average relative humidity. In this study, total monthly rainfall was used as the dependent variable (Y), while monthly average air temperature and monthly average relative humidity were used as independent variables, represented as X_1 and X_2, respectively.

Total monthly rainfall was calculated from the cumulative daily rainfall recorded over each month, whereas air temperature and relative humidity were derived from the monthly averages of daily observations. Climatological data from 2000 to 2023 were used as training data for developing the linear regression models, while data from 2024 were used independently as validation data to evaluate the predictive performance of the models under actual conditions.

Prior to model development, all data underwent preprocessing steps, including data cleaning, consistency verification, and aggregation at the monthly scale. During the data quality assessment, several missing daily observations were identified, primarily during dry-season months when rainfall occurrence was minimal or absent. Because these missing records occurred predominantly under low-rainfall conditions, their contribution to total monthly rainfall was considered negligible and therefore did not significantly affect the monthly rainfall totals used in the analysis. Because ordinary least squares regression is scale-invariant, normalization or scaling of the predictor variables was not required for model estimation, and all climatological variables were retained in their original units. These steps were conducted to ensure data quality and minimize potential bias arising from incomplete, inconsistent, or non-uniform data. The preprocessing stage also aimed to produce a more representative dataset suitable for linear regression analysis of rainfall prediction in the study area.

Linear Regression Modeling

Linear regression models are statistical approaches widely used in climatology to quantitatively represent the relationship between rainfall and atmospheric parameters [5], [12], [13]. In this study, total monthly rainfall was treated as the dependent variable (Y), while air temperature and relative humidity were used as predictor variables. The methods employed in this study consisted of simple linear regression (SLR) and multiple linear regression (MLR). The SLR model employed a single predictor [13], [19], namely either air temperature or relative humidity, whereas the MLR model simultaneously incorporated both variables [5], [19] to predict rainfall.

Linear regression was selected because it provides an interpretable baseline model and allows direct evaluation of the individual and combined effects of air temperature and relative humidity on rainfall variability. Furthermore, the primary objective of this study was to assess the statistical relationships between these climatological variables and monthly rainfall rather than to maximize predictive accuracy through more complex nonlinear or machine learning approaches. Therefore, linear regression was considered an appropriate method for investigating the influence of atmospheric parameters on rainfall variability in the study area. Mathematically, the SLR model can be expressed as [20], [21]

$$Y = \beta_0 + \beta_1 X \,, \tag{1}$$

where X represents either air temperature or relative humidity. Meanwhile, the MLR model is defined as [13], [22]

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \,, \tag{2}$$

where X_1 and X_2 represent air temperature and relative humidity, respectively. The parameter β_0 denotes the intercept, whereas β_1 and β_2 are the regression coefficients. Prior to interpreting the MLR model, multicollinearity among the predictor variables was assessed using the Variance Inflation Factor (VIF) [19], [22]. Furthermore, a statistical significance testing of the individual regression coefficients (β_1 and β_2) was conducted using a two-tailed t-test to evaluate the reliability and contribution of each meteorological predictor to the model [19], [22].

The linear regression models were developed separately for each month using climatological data collected between 2000 and 2023. This monthly modeling approach was adopted to account for the distinct seasonal characteristics of rainfall, temperature, and relative humidity throughout the year. By developing separate models for each month, the analysis implicitly accommodated seasonal variability without requiring an explicit seasonality decomposition procedure. This approach was applied to ensure that the relationship between rainfall and atmospheric parameters could be represented in accordance with the seasonal characteristics of each month. Subsequently, rainfall in 2024 was predicted by substituting the actual 2024 air temperature and relative humidity values into the obtained regression equations. Therefore, the performance of the linear regression models developed in this study could be evaluated.

Model Evaluation Metrics

Evaluation of prediction model performance is essential to ensure the reliability and accuracy of the models developed in this study. In regression-based prediction, two evaluation metrics commonly used are the Root Mean Square Error (RMSE) and the Pearson correlation coefficient [13], [23], [24]. The RMSE value quantifies the average magnitude of prediction errors relative to the observed data. Meanwhile, the Pearson correlation coefficient is used to assess the strength and direction of the linear relationship between predicted and observed values [13], [23], [24]. Collectively, these two metrics provide complementary information for evaluating the accuracy and consistency of prediction models. The RMSE value is defined as [23], [24]

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n}}, \tag{3}$$

where y_i represents the observed value, $\hat{y}_i$ denotes the predicted value, and n indicates the number of observations. A smaller RMSE value indicates better prediction accuracy because it reflects a smaller deviation between observed and predicted values [13], [25]. Futhertmore, the Pearson correlation coefficient (r) is formulated as [26]

$$r = \frac{\sum(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum(y_i - \bar{y})^2 \sum(\hat{y}_i - \bar{\hat{y}})^2}}, \tag{4}$$

where $\bar{y}$ and $\bar{\hat{y}}$ represent the mean values of the observed and predicted data, respectively.

The Pearson correlation coefficient ranges from -1 to 1. Values approaching 1 indicate a strong positive relationship between predicted and observed data, whereas values approaching -1 indicate a strong negative relationship. Meanwhile, values close to zero indicate a weak linear relationship [5], [23]. Furthermore, the interpretation of relationship strength based on the Pearson correlation coefficient used in this study is presented in Table 1.

Table 1. Interpretation of Pearson correlation coefficient values [5], [27]

No.	Interval	Criteria
1	0.00 – 0.19	Very Weak
2	0.20 – 0.39	Weak
3	0.40 – 0.59	Moderate
4	0.60 – 0.79	Strong
5	0.80 – 1.00	Very Strong

In addition to RMSE and Pearson correlation coefficient, residual analysis was performed to further evaluate the adequacy of the developed linear regression models. Residuals represent the differences between observed and predicted rainfall values and are calculated as [28]

$$e_i = y_i - \hat{y}_i , \tag{5}$$

where y_i denotes the observed rainfall value and $\hat{y}_i$ represents the corresponding predicted value. Residual plots were subsequently examined to assess the distribution of prediction errors [28]. A random scatter of residuals around the zero-reference line without discernible systematic patterns indicates that the linear regression models adequately represent the relationship between the predictor variables and rainfall. Moreover, the relatively constant spread of residuals suggests no apparent evidence of heteroscedasticity, indicating that the assumption of homoscedasticity is reasonably satisfied [28], [29]. Furthermore, to formally validate the statistical inferences, the normality of these regression residuals was rigorously assessed using the Shapiro-Wilk test paired with a visual evaluation via a normal Quantile-Quantile (Q-Q) plot [28], [30].

Research Workflow

To provide a systematic overview of the research stages conducted in this study, the research workflow, starting from data collection, preprocessing, regression model development, and model evaluation, is presented in Figure 2.

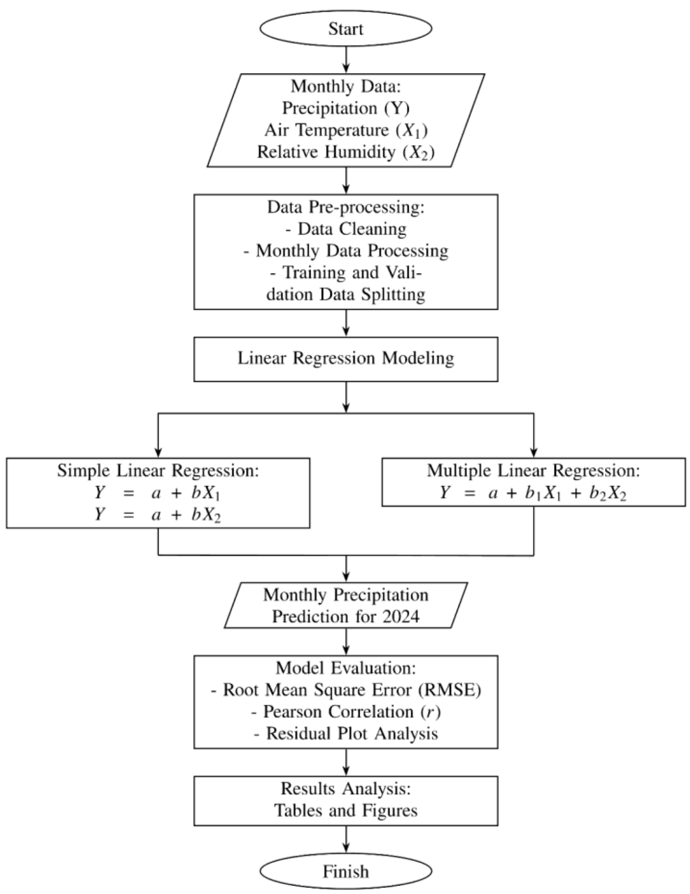

Figure 2. Research workflow of the study

RESULT AND DISCUSSION

Descriptive Statistics

Descriptive statistics for the three meteorological variables, total monthly rainfall (Y), monthly average air temperature (X_1), and monthly average relative humidity (X_2), are presented in Table 2. These statistics were derived from climate observations recorded at the Hang Nadim Meteorological Station in Batam over the 2000–2023 period, comprising 288 monthly observations. The table provides a comprehensive summary of the data, including the number of observations (N), mean, standard deviation (SD), variance, as well as the maximum and minimum values for each parameter to illustrate the climatic characteristics of the study area.

Table 2. The descriptive statistics of data (2000-2023)

Variable	N	Mean	SD	Variance	Max	Min
$\boldsymbol{Y}$	288	181.36	131.77	17363.17	989.50	0.00
$\boldsymbol{X_1}$	288	27.41	0.62	0.39	29.16	26.05
$\boldsymbol{X_2}$	288	82.98	2.69	7.22	88.71	74.52

The descriptive statistics indicate that rainfall in Batam exhibits high variability, as reflected by a standard deviation of 131.77. In contrast, air temperature and relative humidity exhibit relatively low variability, consistent with the stable maritime tropical climate in Batam.

Regression Model Analysis

Based on daily data collected over the 2000–2023 period, linear regression analysis was performed separately for each month. Two approaches were applied: simple linear regression models using either air temperature (X_1) or relative humidity (X_2) individually as predictors, and a multiple linear regression model incorporating both variables simultaneously as predictors. This approach enables a comparison of the performance of each model in predicting monthly rainfall during 2024.

The calculated regression coefficients are formulated as monthly mathematical equations. The general forms of the three regression models (X_1, X_2, and X_1 & X_2) used in the prediction process are summarized in Table 3.

Table 3. Linear regression equations for monthly rainfall prediction in Batam City, 2024

Month	Linear Regression Equations		
	Temperature Predictor (X_1)	Humidity Predictor (X_2)	Temperature & Humidity Predictors (X_1 & X_2)
January	$Y = 3686.67 – 128.20X_1$	$Y = – 3831.45 + 49.68X_2$	$Y = – 3448.03 – 9.88X_1 + 48.24X_2$
February	$Y = – 215.87 + 10.45X_1$	$Y = – 1164.65 + 15.49X_2$	$Y = – 3324.07 + 61.84X_1 + 21.48X_2$
March	$Y = 3574.37 – 123.56X_1$	$Y = – 2370.39 + 31.30X_2$	$Y = – 2714.05 + 8.22X_1 + 32.75X_2$
April	$Y = 1878.93 – 61.96X_1$	$Y = – 1110.82 + 15.28X_2$	$Y = 1106.06 – 47.40X_1 + 4.44X_2$
May	$Y = 506.94 – 11.20X_1$	$Y = – 477.96 + 8.03X_2$	$Y = – 1886.88 + 35.11X_1 + 13.13X_2$
June	$Y = 2291.30 – 76.71X_1$	$Y = – 1942.91 + 25.19X_2$	$Y = – 3139.60 + 29.26X_1 + 29.81X_2$
July	$Y = 1735.78 – 57.45X_1$	$Y = – 1738.38 + 22.66X_2$	$Y = – 4695.64 + 64.53X_1 + 36.83X_2$
August	$Y = 866.33 – 25.74X_1$	$Y = – 958.98 + 13.47X_2$	$Y = – 4293.67 + 75.70X_1 + 28.62X_2$
September	$Y = 2451.45 – 83.83X_1$	$Y = – 1464.08 + 19.40X_2$	$Y = – 597.29 – 20.35X_1 + 15.70X_2$
October	$Y = 4827.28 – 168.89X_1$	$Y = – 3049.97 + 38.99X_2$	$Y = – 1842.61 – 113.50X_1 + 17.60X_2$
November	$Y = 2915.00 – 98.43X_1$	$Y = – 2520.21 + 32.50X_2$	$Y = – 132.19 – 55.67X_1 + 22.18X_2$

December	$Y = 4073.41 - 141.43X_1$	$Y = -2075.69 + 27.73X_2$	$Y = 2548.34 - 118.03X_1 + 10.61X_2$

Linear regression equations used to predict rainfall (Y) based on air temperature (X_1) and relative humadity (X_2) reveal distinct atmospheric relationships within Batam's local climate system. The regression coefficients for temperature vary across months and are often negative, indicating that higher air temperature is associated with reduced rainfall under certain atmospheric stability conditions. This pattern is particularly evident in months such as October and December, when stronger surface heating may suppress local condensation processes. In contrast, the regression coefficients for relative humidity remain consistently positive throughout all twelve months, demonstrating that higher atmospheric moisture availability strongly supports cloud formation and precipitation. These monthly variations in the regression coefficients reflect the dynamic characteristics of Batam's maritime tropical climate.

Following the presentation of the complete set of MLR models in Table 3, a multicollinearity test using the Variance Inflation Factor (VIF) along with individual parameter significance testing via t-test was performed for each monthly model. This analysis was conducted to ensure model stability, evaluate the reliability of individual regression coefficients, and to verify that the predictor variables (X_1 and X_2) did not exhibit strong intercorrelations prior to interpreting the regression equations. If strong correlations exist between the independent variables, the resulting regression coefficient estimates may become unstable and unreliable. The calculated regression coefficients, t-statistics, p-values, and their corresponding Variance Inflation Factor (VIF) values for each month are summarized in Table 4.

Table 4. Regression Coefficients, Multicollinearity, and t-Test Significance Results

Month	Predictor	Coefficient	t-Stat	P-value	VIF
January	(Intercept)	-3448.03	-1.29	0.21	
	Temperature (X_1)	-9.88	-0.16	0.88	1.56
	Humidity (X_2)	48.24	3.11	0.01	1.56
February	(Intercept)	-3324.07	-2.60	0.02	
	Temperature (X_1)	61.84	1.82	0.08	1.30
	Humidity (X_2)	21.48	3.14	0.01	1.30
March	(Intercept)	-2714.05	-0.79	0.44	
	Temperature (X_1)	8.22	0.10	0.92	3.41
	Humidity (X_2)	32.75	1.96	0.06	3.41
April	(Intercept)	1106.06	0.45	0.66	
	Temperature (X_1)	-47.40	-0.93	0.36	4.00
	Humidity (X_2)	4.44	0.33	0.75	4.00
May	(Intercept)	-1886.88	-0.79	0.44	
	Temperature (X_1)	35.11	0.61	0.55	2.05
	Humidity (X_2)	13.13	1.13	0.27	2.05
June	(Intercept)	-3139.60	-1.84	0.08	
	Temperature (X_1)	29.26	0.73	0.47	2.28

	Humidity (X_2)	29.81	3.53	0.00	2.28
July	(Intercept)	-4695.64	-2.24	0.04	
	Temperature (X_1)	64.53	1.45	0.16	3.66
	Humidity (X_2)	36.83	3.21	0.00	3.66
August	(Intercept)	-4293.67	-2.50	0.02	
	Temperature (X_1)	75.70	1.99	0.06	3.44
	Humidity (X_2)	28.62	3.17	0.00	3.44
September	(Intercept)	-597.29	-0.21	0.84	
	Temperature (X_1)	-20.35	-0.31	0.76	3.80
	Humidity (X_2)	15.70	1.13	0.27	3.80
October	(Intercept)	1842.61	0.49	0.63	
	Temperature (X_1)	-113.50	-1.36	0.19	2.46
	Humidity (X_2)	17.60	0.86	0.40	2.46
November	(Intercept)	-132.19	-0.04	0.97	
	Temperature (X_1)	-55.67	-0.71	0.48	1.56
	Humidity (X_2)	22.18	0.92	0.37	1.56
December	(Intercept)	2548.34	0.54	0.60	
	Temperature (X_1)	-118.03	-1.07	0.30	1.47
	Humidity (X_2)	10.61	0.37	0.71	1.47

To rigorously assess the individual contribution of each meteorological variable within the regression models, a detailed evaluation of the model parameters listed in Table 4 was conducted. When evaluating the individual significance of the predictors via t-test, Relative Humidity (X_2) emerges as the most dominant and statistically reliable driver for monthly rainfall in Batam. It exhibits significant positive impacts ($p < 0.05$) across multiple months, specifically January ($t = 3.11$), February ($t = 3.14$), June ($t = 3.53$), July ($t = 3.21$), and August ($t = 3.17$), demonstrating that atmospheric moisture availability plays a critical role in precipitation variance.

In contrast, Air Temperature (X_1) shows weaker direct linear significance on a monthly basis, only reaching marginal significance ($p < 0.10$) during February ($t = 1.82$) and August ($t = 1.99$). To ensure the statistical validity of these regression coefficients, a multicollinearity test was evaluated based on the Variance Inflation Factor (VIF). The VIF values for all months consistently range from 1.30 to 4.00, which is substantially lower than the conservative critical threshold of 10, confirming that no problematic intercorrelations exist between air temperature and relative humidity [19], [22].

Rainfall Prediction for 2024

The linear regression equation used to predict rainfall (Y) based on air temperature (X_1) and relative humidity (X_2) is a multiple linear regression model that incorporates more than one independent variable. By incorporating both variables simultaneously, the model is expected to provide more accurate rainfall estimates than simple linear regression models, which use only a single explanatory variable. After obtaining the linear regression equations for each month, the next step is to calculate the predicted total monthly rainfall for the year 2024. This calculation is performed by substituting the actual air temperature and humidity values of 2024 into the established regression models. The resulting monthly rainfall predictions are then presented in Table 5.

Table 5. Predicted total monthly rainfall (mm) in Batam City, 2024

	Predicted Total Monthly Rainfall (mm)		
Month	**Temperature Predictor**	**Humidity Predictor**	**Temperature & Humidity Predictors**
January	226	467	459
February	79	66	131
March	6	95	103
April	77	128	88
May	189	214	240
June	143	188	201
July	84	92	134
August	148	178	233
September	105	178	161
October	79	260	145
November	213	316	275
December	100	173	93

The predicted total monthly rainfall for 2024 obtained from the linear regression models was subsequently compared with the actual data provided by BMKG. The comparison indicates that the model using both air temperature and relative humidity as predictors produces estimates that most closely match the observed conditions. This is reflected in the consistency between the predicted patterns and the observed data trends, particularly during months with high rainfall such as January and May. To support these findings, three graphical visualizations are presented to compare the performance of each model with the observed data.

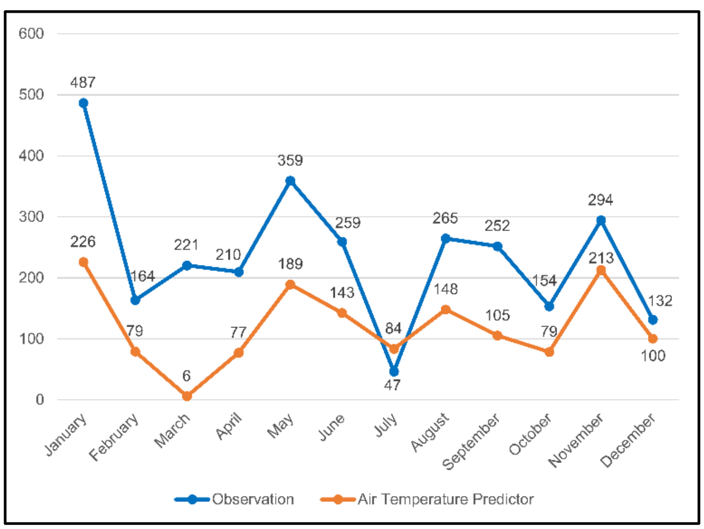

Figure 3. Comparison of predicted total monthly rainfall using air temperature as a predictor with observed data at Hang Nadim Meteorological Station in 2024

The monthly rainfall prediction results derived using air temperature as the sole predictor variable are presented in Figure 3. In general, the model exhibits relatively large deviations from the observed data, particularly during months with high rainfall. In January and May, the predicted values are significantly lower than the actual data, indicating substantial underprediction. Conversely, in July, when the observed rainfall is very low (47 mm), the model produces a higher prediction (84 mm), indicating overestimation. These discrepancies highlight the limitations of using air temperature as a single predictor in representing the complex variability of rainfall in tropical regions. Consequently, the relationship between temperature and rainfall cannot be consistently assumed to be linear throughout the year. Therefore, incorporating additional climatological variables, such as relative humidity, is necessary to improve the predictive accuracy of the models.

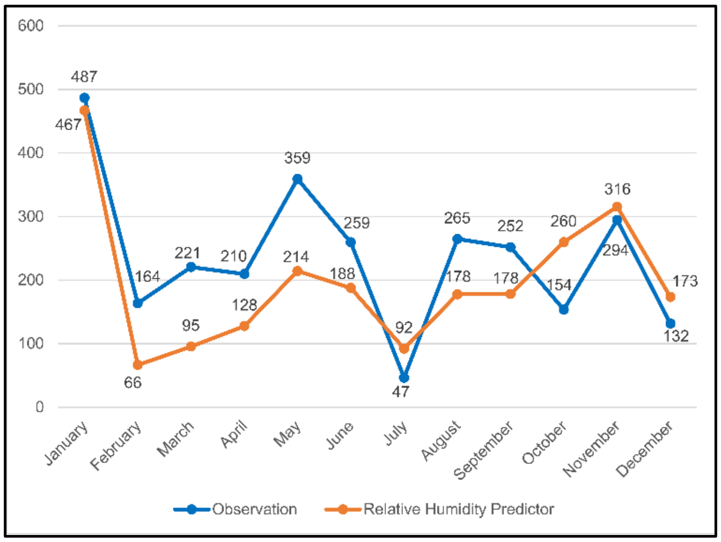

Figure 4. Comparison of predicted total monthly rainfall using relative humidity as a predictor with observed data at Hang Nadim Meteorological Station in 2024

Figure 4 shows the results of monthly rainfall predictions using air humidity as the sole predictor variable. In general, this model demonstrates more stable performance compared to the model that uses only air temperature, although notable discrepancies still occur in several months. In January, the predicted value (467 mm) is relatively close to the observed data (487 mm), indicating a fairly accurate estimate. However, in February and March, significant underprediction occurs, with differences approaching 100 mm. In contrast, the model tends to overestimate rainfall in October and November, particularly in October, where the predicted value (260 mm) exceeds the observed value (154 mm). Nevertheless, relative humidity generally captures monthly rainfall variability more effectively than air temperature because of its more direct influence on condensation processes and cloud formation.

In contrast to Figures 3 and 4, which present monthly rainfall predictions using simple linear regression (SLR) models, Figure 5 illustrates monthly rainfall predictions generated using a multiple linear regression (MLR) model with air temperature and relative humidity as predictor variables. In general, this model demonstrates more accurate and consistent performance compared to single-variable models, with relatively small deviations from the observed data. In January, the predicted value (459 mm) is close to the actual data (487 mm), and a similar pattern is observed in November (275 mm compared to 294 mm), reflecting the model's ability to follow seasonal rainfall patterns. Although some discrepancies remain in certain months, such as March and July, the overall prediction trend aligns well with the observed rainfall variability, increasing during the wet season and decreasing during the dry season. This indicates that combining air temperature and humidity enables the model to capture climate dynamics more representatively, resulting in more realistic rainfall estimates for tropical regions such as Batam City.

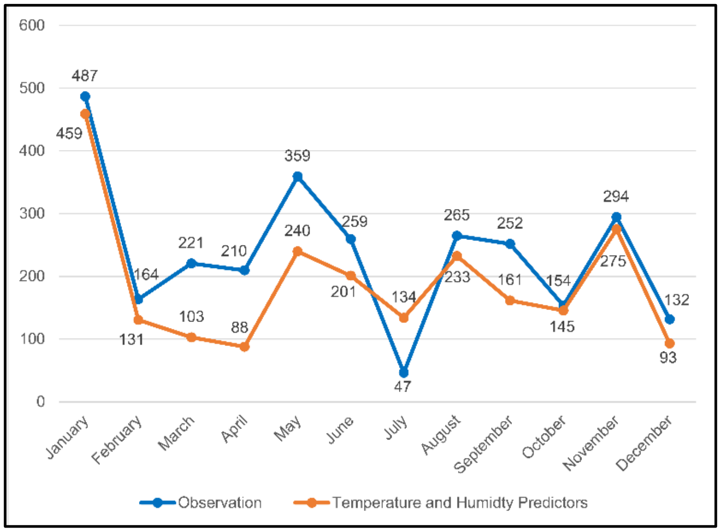

Figure 5. Comparison of predicted total monthly rainfall using air temperature and humidity as predictors with observed data at Hang Nadim Meteorological Station in 2024

Statistically, the performance of each regression model was evaluated using the Root Mean Square Error (RMSE) and the Pearson correlation coefficient (r). The temperature-based SLR model produced an RMSE of approximately ± 138 mm with a correlation coefficient of $r = 0.737$, indicating a moderately strong relationship but relatively large prediction deviations. The humidity-based SLR model showed improved performance, with an RMSE of approximately ± 85 mm and $r = 0.759$, reflecting a stronger and more stable relationship. Among all models, the MLR model combining temperature and humidity achieved the best performance, yielding the lowest RMSE (± 74 mm) and the highest correlation coefficient ($r = 0.855$). Based on the correlation interpretation presented in Table 1, this result indicates a very strong relationship between the predicted and observed rainfall values.

The superior performance of the combined regression model can be explained from a climatological perspective. Temperature influences the evaporation process, while humidity reflects the atmosphere's capacity to retain water vapor before it condenses into precipitation. When these two variables are considered simultaneously, they provide complementary information about atmospheric conditions, thereby significantly enhancing predictive capability for rainfall. These findings are consistent with previous studies conducted in several regions of Indonesia, including Pontianak and Sleman [14], [15], which reported that regression models incorporating multiple predictors yield more accurate rainfall predictions than models based on a single predictor.

Nevertheless, studies from Mojokerto, Southern Bali, and Deli Serdang reported different levels of influence of temperature and humidity on rainfall [5], [13], [18], suggesting that the effectiveness of these predictors depends on local climatic characteristics.

This regional variability highlights the importance of evaluating predictor performance within specific climatic settings. In the case of Batam, the superior performance of the MLR model indicates that temperature and relative humidity jointly provide a more effective representation of rainfall variability in a tropical maritime environment. To further support this interpretation, the residual plots in Figure 6 were examined to assess the statistical adequacy of the competing regression models. The random distribution of residuals around the zero reference line, without discernible systematic patterns, indicates that the assumptions of linearity and homoscedasticity are reasonably satisfied. In addition, the more compact residual distribution of the MLR model indicates smaller prediction errors and greater stability, further supporting its superior predictive performance.

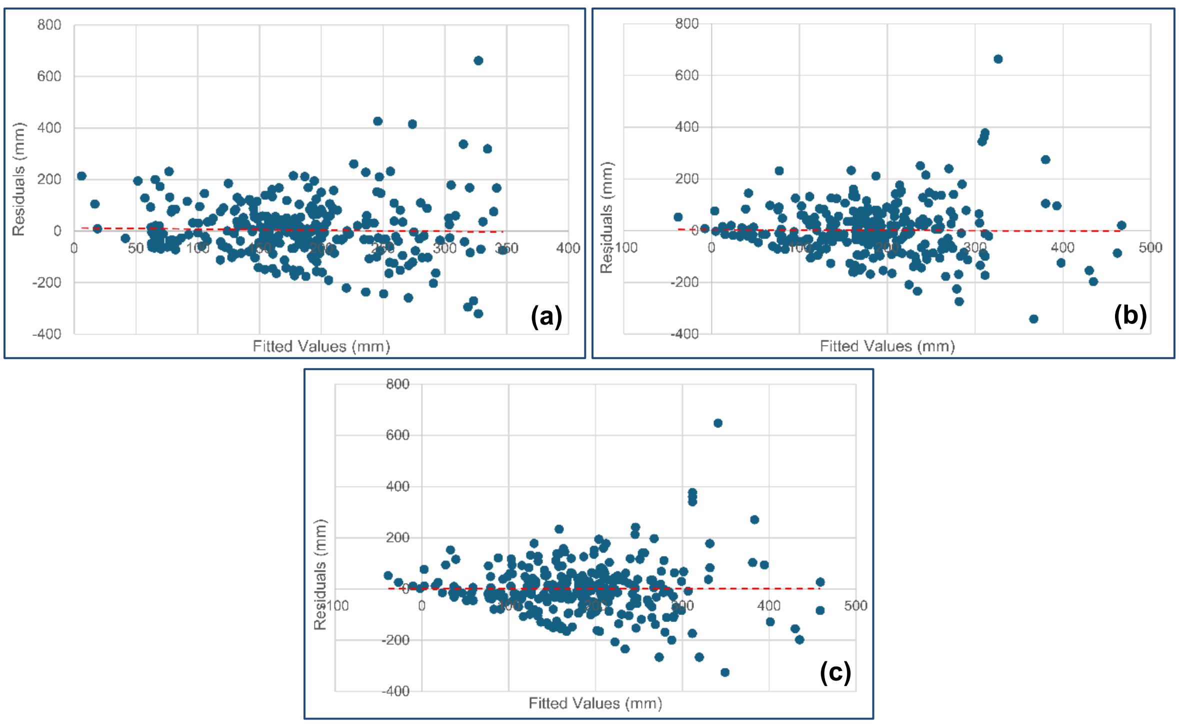

Figure 6. Residual plots for monthly rainfall prediction at Hang Nadim Meteorological Station (2000–2024): simple linear regression models with (a) air temperature and (b) relative humidity as predictors, and (c) a multiple linear regression model incorporating both predictors.

To formally ensure the validity of the parametric inferences, the normality of the model residuals was evaluated using a combination of the Shapiro-Wilk test and a visual inspection via a master normal Q-Q plot (Figure 7). Statistically, the monthly Shapiro-Wilk test results revealed that several months perfectly satisfy the strict normality assumption ($p > 0.05$), such as April ($W = 0.987, p = 0.9858$), May ($W = 0.966, p = 0.567$), June ($W = 0.962, p = 0.470$), and October ($W = 0.949, p = 0.259$). Although marginal deviations ($p < 0.05$) are formally observed in highly dynamic months like December due to the inherent high variability and extreme localized convective rainfall in Batam, the overall structural adequacy of the model remains robust.

This is strongly confirmed by the master Q-Q plot in Figure 7, where the joint residual data points across the entire multi-year period tightly cluster and closely align along the 45-degree linear reference line. This collective visual alignment demonstrates that the regression residuals are symmetrically distributed without major systematic distortions, thereby confirming that the regression coefficients, t-test significance, and subsequent climate interpretations in this study are statistically stable, reliable, and unbiased. With all fundamental classical linear regression assumptions thoroughly validated and satisfied, the detailed statistical outputs presented in Table 4 can now be confidently interpreted, structurally aligning with the empirical finding that relative humidity exerts a far more dominant influence on monthly rainfall variability compared to air temperature.

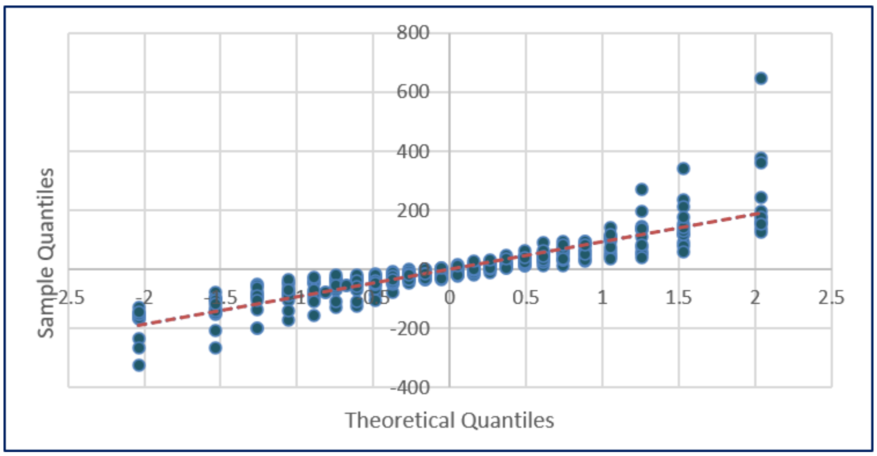

Figure 7. Normal Q-Q plot of the multiple linear regression model residuals across all monthly periods.

Furthermore, rainfall dynamics in Batam are influenced not only by local conditions but also by global climate phenomena such as El Niño and Indian Ocean Dipole (IOD). A study conducted in Batam found that fluctuations in sea surface temperature and wind speed associated with the IOD significantly influence precipitation patterns in the region [31]. During the positive phase of the IOD, reduced moisture transport from the ocean suppresses atmospheric moisture availability, resulting in decreased rainfall. Conversely, during negative or neutral IOD phases, atmospheric humidity increases, enhancing the likelihood of rainfall. This is consistent with the observed monthly data during the wet season.

Nevertheless, extreme climate events such as El Niño and strong IOD episodes may contribute to deviations between observed and predicted rainfall during certain years. These large-scale climate anomalies can alter regional atmospheric circulation, moisture transport, and rainfall distribution patterns beyond the variability captured by local temperature and relative humidity alone, thereby affecting the strength of the regression relationships. In addition, lower predictive performance during transitional seasons may be associated with local convective processes that are not fully represented by the selected predictor variables. During these periods, rainfall formation is often influenced by short-term atmospheric instability, cloud development, and localized weather systems, introducing additional variability that cannot be entirely captured by the regression model. Consequently, prediction errors may increase under anomalous climatic conditions and during months characterized by highly dynamic atmospheric processes.

From a practical perspective, the validity of this predictive model is not only relevant in a climatological context but also provides potential benefits for economic sectors such as tourism. Batam City, as a coastal area, heavily relies on outdoor tourism activities. Therefore, monthly weather predictions may support the development of promotional strategies and risk mitigation planning. In general, weather conditions can influence tourism activity, particularly in coastal destinations where many activities depend on favorable outdoor conditions. In this context, the integration of climatological data into tourism management may help policymakers develop more adaptive, data-driven promotional strategies.

CONCLUSION

Analysis of monthly rainfall data for Batam City in 2024 indicates that the multiple linear regression (MLR) model using air temperature and relative humidity as predictors provides more accurate rainfall predictions than the simple linear regression (SLR) model. This model shows a strong agreement with observational data, particularly during months with high rainfall, such as January and November. Consistent with this finding, it produced the lowest RMSE value (± 74 mm) and the highest Pearson correlation coefficient ($r = 0.855$). In contrast, the model based solely on air temperature yields the lowest predictive accuracy, as evidenced by the highest RMSE (± 138 mm) and the lowest Pearson correlation coefficient ($r = 0.737$). These results indicate greater deviations between the predicted and observed rainfall values across several months. Meanwhile, the model based on relative humidity as a single predictor yields more consistent results, particularly in January, November, and December. This model achieves a lower RMSE ($\pm 85\ mm$) and a higher Pearson correlation coefficient ($r = 0.759$), indicating a more stable contribution of humidity to rainfall variability than air temperature.

Overall, this study indicates that air temperature and relative humidity predictors can effectively represent the dynamics of monthly rainfall in coastal areas such as Batam City. These findings provide useful information to support climate adaptation planning and strengthen disaster risk reduction strategies in tropical coastal regions such as Batam. Nevertheless, the model is limited by its reliance on only two predictor variables and by the use of linear regression, which may not fully capture the nonlinear nature of climate–rainfall relationships. Therefore, the model's accuracy can be further improved by incorporating additional climatological variables, such as wind direction and speed, sea surface temperature, or air pressure. Furthermore, integrating machine learning or hybrid forecasting approaches may better capture complex climate dynamics and improve rainfall prediction accuracy.

ACKNOWLEDGMENTS

The authors would like to thank BMKG Hang Nadim Station for providing the climatological data and STMKG for supporting this research.

REFERENCE

[1] Y. Gao, Q. Xiao, and Z. Fang, "The Impact of Rainfall on Water, Energy, Industry and Economic Growth — Based on Empirical Data from 29 Provinces in China," *Sustainability*, vol. 17, no. 1, pp. 1–24, 2025, doi: 10.3390/su17010040.

[2] T. Yu, Q. Ran, H. Pan, J. Li, J. Pan, and S. Ye, "The impacts of rainfall and soil moisture to fl ood hazards in a humid mountainous catchment : a modeling investigation," *Front. Earth Sci.*, vol. 11, pp. 1–12, 2023, doi: 10.3389/feart.2023.1285766.

[3] L. Zhu, Z. Meng, F. Zhang, and P. M. Markowski, "The influence of sea- and land-breeze circulations on the diurnal variability in precipitation over a tropical island," *Atmos. Chem. Phys.*, pp. 13213–13232, 2017, doi: 10.5194/acp-17-13213-2017.

[4] D. T. Mihailovic and S. Malinovic-Milicevic, "A Novel Approach to Understanding the Complexity of Precipitation," *Atmosphere (Basel).*, vol. 16, no. 6, pp. 1–17, 2025, doi: 10.3390/atmos16060706.

[5] M. D. Purnama and M. E. Mustafidah, "Relationship Between Temperature and Humidity on Rainfall : A Multiple Linear Regression Analysis," *J. Eng. Math. Comput. Sci.*, vol. 6, no. 2, pp. 151–156, 2024, doi: 10.21512/emacsjournal.v6i2.11466.

[6] H. Sama, R. H. Gumolung, and Y. Christian, "Ensemble-Based Machine Learning for Improving Local Weather Prediction Accuracy in Batam , Indonesia," *J. Innov. Technol. Polbeng Ser. Informatics*, vol. 11, no. 1, pp. 128–138, 2026, doi: 10.35314/9p53qe06.

[7] W. Anurogo *et al.*, "Cellular Automata – Artificial Neural Network Approach to Dynamically Model Past and Future Surface Temperature Changes : A Case of a Rapidly Urbanizing Island Area , Indonesia," *Land*, vol. 14, no. 8, pp. 1–20, 2025, doi: 10.3390/land14081656.

[8] I. Anggraeny and I. K. Ayu, "Development of Indonesian Free Trade and Port Zone : Analysis of Historical in Batam Island," *J. Law, Policy Glob.*, vol. 99, pp. 19–24, 2020, doi: 10.7176/JLPG/99-03.

[9] M. Hulu, D. M. Lemy, A. Pramezwary, J. Juliana, and Y. D. Timba, "Batam City Coastal Tourism Management : Economic Opportunities for Micro Small and Medium Enterprises in Tourism," *GATR J. Bus. Econ. Rev.*, vol. 8, no. 1, pp. 31–39, 2023, doi: 10.35609/jber.2023.8.1(1).

[10] H. G. T. Olya and H. Alipour, "Risk assessment of precipitation and the tourism climate index," *Tour. Manag.*, vol. 50, pp. 73–80, 2015, doi: 10.1016/j.tourman.2015.01.010.

[11] D. Touloumidis, M. Madas, and V. Zeimpekis, "Weather-Related Disruptions in Transportation and Logistics : A Systematic Literature Review and a Policy Implementation Roadmap," *Logistics*, vol. 9, no. q, pp. 1–28, 2025, doi: 10.3390/logistics9010032.

[12] Y. Kassem, H. Gökçekuş, H. Çamur, and E. Esenel, "Application of artificial neural network, multiple linear regression, and response surface regression models in the estimation of monthly rainfall in Northern Cyprus," *Desalin. Water Treat.*, vol. 215, pp. 328–346, 2021, doi: 10.5004/dwt.2021.26525.

[13] I. F. Daba, K. Dewi, M. R. A. S. Nia, and D. Akbar, "Monthly Rainfall Prediction in South Bali Using Multiple Regression Based on Air Temperature and Humidity Indicators," *Bul. GAW Bariri*, vol. 6, no. 2, pp. 108–117, 2025, doi: 10.31172/bgb.v6i2.165.

[14] A. Fadholi, "Persamaan Regresi Prediksi Hujan Bulanan di Pontianak dengan Prediktor Suhu

dan Kelembapan Udara," *Forum Ilm. Indonusa*, vol. 10, no. 2, pp. 230–241, 2013, [Online]. Available: https://ejurnal.esaunggul.ac.id/index.php/Formil/article/view/843

[15] W. J. Suwondo and K. Ramadina, "Utilization of Air Humidity and Surface Air Temperature Data in Regression Equations for Rainfall Forecast Simulation in the Gamping Area, Sleman, Yogyakarta (Yogyakarta Class 1 Geophysical Station)," *Sunan Kalijaga J. Phys.*, vol. 6, no. 1, pp. 11–18, 2024, doi: 10.14421/physics.v6i1.5136.

[16] M. Sulistiyono, A. Sidauruk, B. Satria, R. Wardhana, and U. A. Yogyakarta, "Rainfall Prediction Using Multiple Linear Regression," *J. Ilmu Pengetah. dan Teknol. Komput. Nusa Mandiri*, vol. 9, no. 1, pp. 17–22, 2023, doi: 10.33480/jitk.v9i1.4203.RAINFALL.

[17] A. Dawamudin and E. Sofia, "Analysis of The Impact of Climate Variables on Banjarbaru City Rainfall," *Civ. Eng. Int. J. Lambung Mangkurat Creat. From South Kalimantan*, vol. 8, no. 1, pp. 29–38, 2024, doi: 10.20527/crc.v8i1.13374.

[18] I. J. A. Saragih, I. Rumahorbo, R. Yudistira, and D. Sucahyono, "Prediksi Curah Hujanan Bulanan di Deli Serdang Menggunakan Persamaan Regresi Dengan Prediktor Data Suhu dan Kelembapan Udara," *J. Meteorol. Klimatologi dan Geofis.*, vol. 7, no. 2, pp. 6–14, 2020.

[19] N. H. R. Nabila, Y. Fitri, and P. R. Arum, "Analisis Faktor-Faktor Yang Mempengaruhi Indeks Pembangunan Manusia Berdasarkan Kabupaten / Kota di Jawa Tengah," *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 16, no. 1, pp. 424–433, 2023, doi: 10.36456/jstat.vol16.no1.a7079.

[20] S. Lestari, "Analisis Algoritma Regresi Linear Sederhana dalam Memprediksi Tingkat Penjualan Album KPOP," *INSOLOGI J. Sains dan Teknol.*, vol. 2, no. 1, pp. 199–209, 2023, doi: 10.55123/insologi.v2i1.1692.

[21] N. Nurhaswinda *et al.*, "Analisis regresi linier sederhana dan penerapannya," *J. Cahaya Nusant.*, vol. 1, no. 2, pp. 69–78, 2025, [Online]. Available: https://jurnal.cahayapublikasi.com/index.php/jcn/article/view/29

[22] F. Y. Risdiana, "Analisis Jaringan Kantor dan Pertumbuhan Tenaga Kerja Terhadap Total Aset Unit Usaha Syariah Tahun 2019," *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 13, no. 1, pp. 1–7, 2020, doi: 10.36456/jstat.vol13.no1.a3265.

[23] M. Abdullah and S. Said, "Performance Evaluation of Machine Learning Regression Models for Rainfall Prediction," *Iran. J. Sci. Technol. Trans. Civ. Eng.*, vol. 49, pp. 4231–4250, 2024, doi: 10.1007/s40996-024-01691-4.

[24] D. Y. Faidah, H. Kuswanto, and S. Sutikno, "Improving the Accuracy of Rainfall Prediction Using Bias-Corrected NMME Outputs : A Case Study of Surabaya City , Indonesia," *Sci. World J.*, vol. 2022, no. 1, pp. 1–9, 2022, doi: 10.1155/2022/9779829.

[25] P. P. Simanjuntak, "Perbandingan Kinerja Hasil Luaran Model Jaringan Syaraf Tiruan dan SARIMA Untuk Prediksi Awal Musim Hujan Kota Pangkalpinang," *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 16, no. 1, pp. 407–423, 2023, doi: 10.36456/jstat.vol16.no1.a6137.

[26] R. K. Putri, M. Athoillah, and A. Haqiqiyah, "Analisis Faktor Yang Mempengaruhi Ketepatan Kelulusan Mahasiswa Dengan Algoritma Regresi Linear," Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika, vol. 5, no. 2, pp. 671–680, 2024.

[27] N. H. Muzaki, E. Febriati, and Y. D. Haryanto, "Pemanfaatan Satelit Himawari-8 Untuk Estimasi Curah Hujan dengan Metode Convective Stratiform Technique (CST) dan Modified Convective Stratiform Technique (mCST) di Wilayah Ekuatorial dan Monsunal (Studi Kasus

Sulawesi 2020)," *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 14, no. 2, pp. 67–73, 2021, doi: 10.36456/jstat.vol14.no2.a3860.

[28] H. Kim, "Statistical notes for clinical researchers : simple linear regression 3 – residual analysis," *Restor. Dent. Endod.*, vol. 44, no. 1, pp. 1–8, 2019, doi: 10.5395/rde.2019.44.e11.

[29] U. Hidayaturrahman, B. Basirun, D. S. Ayundasari, and M. A. Zulkarnain, "Application of Dummy Regression to Estimate the Income of the Working Population in East Lombok," *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 18, no. 1, pp. 789–797, 2025, doi: 10.36456/jstat.vol18.no1.a9808.

[30] D. S. Hastuti and S. Yulianto, "Factors Affecting the Resilience Index Food in Papua Province and West Papua Province Using a Spatial Model Approach," *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 17, no. 1, pp. 635–644, 2024, doi: 10.36456/jstat.vol17.no1.a9087.

[31] G. Surya *et al.*, "Karakteristik Suhu Permukaan Laut dan Kecepatan Angin di Perairan Batam Hubungannya dengan Indian Ocean Dipole (IOD)," *Din. Marit.*, vol. 6, no. 1, pp. 1–6, 2017, [Online]. Available: https://ojs.umrah.ac.id/index.php/dinamikamaritim/article/view/184