

Pemanfaatan Regresi Linier Berganda dalam Memprediksi Pola Curah Hujan di Kabupaten Karawang

Rosa Andriani^{1*}, Gayatri Giani Rengga²

^{1,2} Program Studi Matematika, Universitas Terbuka, Tangerang Selatan, Indonesia

* Corresponding Author. E-mail: rosa.andriani@ecampus.ut.ac.id

ARTICLE INFO

Article history:

Received: August 20th, 2025

Revised: September 28th, 2025

Accepted: October 2nd, 2025

Available: online October 30th, 2025

Kata Kunci:

curah hujan, prediksi, regresi linier berganda

Keywords:

multiple linear regression, prediction, rainfall



ABSTRAK

Keberadaan objek air di suatu wilayah dilansir mempengaruhi proses terjadinya curah hujan. Secara geografis, kondisi hidrologi Kabupaten Karawang relatif unik karena dialiri dan sekaligus dikelilingi banyak objek air. Beberapa objek air yang cukup masif di area Kabupaten Karawang adalah Laut Jawa di batas utara, Waduk Jatiluhur di area tenggara, Sungai Citarum di batas barat, serta Sungai Cilamaya di batas timur. Temuan tentang kondisi hidrologi Kabupaten Karawang tersebut mendorong peneliti untuk mensintesis model yang dapat memprediksi curah hujan di Kabupaten Karawang. Pemodelan akan berguna dalam berbagai aspek, seperti dalam mitigasi bencana, pembangunan drainase, atau perencanaan pertanian. Pemodelan dilakukan dengan metode regresi linier berganda. Sumber set data adalah data pada periode 2018-2020. Variabel independen yang digunakan dalam menentukan model adalah suhu, kecepatan angin, kelembapan dan lama penyinaran. Pada tahap awal data dieksplorasi dengan menganalisis nilai deskriptif, kemudian dilakukan uji asumsi klasik. Selanjutnya digunakan analisis regresi linier berganda untuk

menghasilkan model curah hujan. Model yang dihasilkan lalu diukur kinerjanya melalui RMSE (*Root Mean Square Error*) yang dinormalisasi dan *R-squared*. Hasil menunjukkan nilai RMSE yang dinormalisasi sebesar 0,091 dan *R-squared* sebesar 87,1% yang mengindikasikan model teruji prediktif dan akurat.

ABSTRACT

The presence of water bodies in a region is reported to influence the process of rainfall formation. Geographically, the hydrological condition of Karawang Regency is relatively unique, as the area is traversed and surrounded by numerous water bodies. Some of the major water bodies in Karawang Regency include the Java Sea to the north, Jatiluhur Reservoir to the southeast, the Citarum River to the west, and the Cilamaya River to the east. These hydrological characteristics of Karawang Regency encouraged researchers to develop a model capable of predicting rainfall in the region. Such modeling is beneficial in various aspects, including disaster mitigation, drainage development, and agricultural planning. The modeling process employed the multiple linear regression method, using data from the 2018–2020 period. The independent variables used to construct the model were temperature, wind speed, humidity, and sunshine duration. In the initial stage, the data were explored through descriptive statistical analysis, followed by classical assumption testing. Subsequently, a multiple linear regression analysis was conducted to produce the rainfall prediction model. The performance of the resulting model was evaluated using Normalized Root Mean Square Error (RMSE) and R-squared. The results showed a normalized RMSE value of 0.091 and an R-squared value of 87.1%, indicating that the developed model demonstrates strong predictive capability and high accuracy.

PENDAHULUAN

Terdapat 514 Kabupaten/Kota di Indonesia, salah satunya adalah Kabupaten Karawang yang terletak di provinsi Jawa Barat (BPS, 2024, p. 7). Berdasarkan katalog BPS Kabupaten Karawang (2023), wilayah ini merupakan dataran rendah. Sebagian besar wilayahnya relatif rata dengan variansi 0-5 m di atas permukaan laut. Secara geografis, kondisi hidrologi Kabupaten Karawang yang merupakan dataran rendah cukup unik karena sekaligus dialiri, dilintasi dan dikelilingi oleh objek air yang jumlahnya cukup signifikan. Bagian utara Kabupaten Karawang berbatasan langsung dengan Laut Jawa. Area tenggara Kabupaten Karawang berdekatan dengan Waduk Jatiluhur yang merupakan waduk terbesar di Indonesia dengan potensi debit air 12,9 miliar m³/tahun (Mulyono & Fauzi, 2020, p. 3). Di bagian baratnya mengalir Daerah Aliran Sungai atau DAS Citarum yang berbatasan dengan Kabupaten Bekasi, sedangkan di bagian timur mengalir DAS Cilamaya yang berbatasan dengan Kabupaten Subang. Selain itu, berdasarkan RKPD Kabupaten Karawang (2024) terdapat 24 objek air lainnya yang mengalir dan melintasi Kabupaten Karawang.

Ada 24 objek air yang mengalir dan mengelilingi Kabupaten Karawang. Dari 15 sungai yang mengalir di wilayah Kabupaten Karawang, 14 di antaranya berhulu di wilayah lain. Dari 21 DAS yang berlokasi di Jawa Barat dan bermuara di Laut Jawa, 3 DAS melintasi Kabupaten Karawang, artinya wilayah ini menampung 14,28% lalu lintas air Jawa Barat yang bermuara di Laut Jawa (RKPD Provinsi Jawa Barat, 2023). Angka tersebut cukup signifikan, mengingat luas Kabupaten Karawang hanya sekitar 3,73% dari luas Provinsi Jawa Barat dan karakteristiknya sebagai dataran rendah yang minim memuat hulu sungai (BPS Kabupaten Karawang, 2022).

Keberadaan objek air pada suatu wilayah dilansir dapat memengaruhi proses presipitasi (Casallas-Garcia, et al., 2023, p. 1). Presipitasi yang terjadi di wilayah tropis turun dalam bentuk curah hujan (*rainfall*) (Tallar, 2023, p. 27). Sebagaimana dijelaskan oleh Asdak (seperti dikutip dalam Sanusi & Side, 2016), "*presipitasi adalah peristiwa klimatik yang bersifat alamiah yaitu perubahan bentuk uap air di atmosfer menjadi curah hujan sebagai akibat proses kondensasi*" (p. 15). Lebih lanjut Tallar (2023) mengemukakan:

"Curah hujan didefinisikan sebagai jumlah hujan terkumpul di wilayah daratan dengan mengabaikan penguapan, rembesan, atau aliran lainnya. Arti dari hujan sebesar 1 mm adalah volume hujan tertampung setinggi 1 mm dan menginfiltrasi area seluas satu meter persegi. Sedangkan intensitas curah hujan adalah jumlah curah hujan yang turun pada durasi waktu tertentu dan biasanya dinyatakan dalam satuan mm/jam, mm/hari, atau mm/tahun". (p. 41-42)

Berdasarkan beberapa literatur diketahui bahwa intensitas curah hujan dipengaruhi berbagai faktor. Menurut Tallar (2023) ada empat faktor yang memengaruhi intensitas curah hujan, yaitu kelembapan udara, jumlah butir-butir air yang dipengaruhi proses kondensasi, suhu atau temperatur udara, dan pembentukan gumpalan awan. Menurut Mulyani, et al. (2019) ada empat faktor untuk memprediksi curah hujan di Kabupaten Majalengka, yaitu suhu, kelembapan, lama penyinaran, dan kecepatan angin. Luthfiarta, et al. (2020) dalam analisa prakiraan cuaca di Semarang mengemukakan terdapat empat faktor yang dijadikan parameter prediksi, adapun diantaranya adalah suhu, kelembapan, tekanan udara, dan kecepatan angin. Dalam studi kasus di Provinsi Yogyakarta yang dilakukan oleh Normelia, et al. (2022) dua faktor dijadikan parameter, yaitu suhu dan lama penyinaran. Sedangkan dalam studi kasus yang dilakukan oleh Maulita & Nurdin (2023) di Kabupaten Aceh Utara, terdapat empat faktor yang dijadikan parameter prediksi curah hujan, yaitu temperatur, kelembapan, lama penerangan matahari, dan kecepatan angin. Purnama & Mustafidah (2024) dalam studi kasus di Kabupaten Mojokerto memprediksi ada dua faktor yang memengaruhi curah hujan, yaitu suhu dan kelembapan. Fiola, et al. (2024) menggunakan tujuh variabel untuk membuat model curah hujan di Provinsi Lampung, yaitu hari hujan, kecepatan angin, kelembapan, tekanan udara, lama penyinaran matahari, suhu, dan suhu minimum.

Temuan tersebut menunjukkan bahwa intensitas curah hujan di suatu wilayah dipengaruhi multi faktor. Selain itu, terlihat bahwa beberapa wilayah telah dijadikan lokasi studi kasus untuk pembuatan model curah hujan. Namun, dalam kasus pemodelan curah hujan, Kabupaten Karawang dengan kondisi hidrologi yang unik belum banyak diteliti secara holistik. Pemodelan curah hujan di suatu wilayah bermanfaat dalam berbagai aspek, seperti mitigasi bencana alam (Stefanny P. S & Suharsono, 2025), perencanaan pertanian (Padilah & Adam, 2019), dan perencanaan pembangunan saluran drainase (Sanusi & Side, 2016, p. 2). Penelitian ini bertujuan untuk melakukan pemodelan curah hujan di Kabupaten Karawang agar masyarakat dapat mendayagunakan model sesuai dengan kebutuhan.

Pemodelan yang mempertimbangkan multi faktor dan parameter terkait dapat disintesis melalui beberapa metode, salah satunya adalah analisis regresi linier berganda (Reza, 2024, p. 188). Seperti dijelaskan oleh Indartini & Mutmainah (2024):

"Analisis regresi linier berganda adalah hubungan secara linier antara dua atau lebih variabel independen (X_1, X_2, \dots, X_n) dengan variabel dependen (Y). Analisis ini dimaksudkan untuk mengetahui arah hubungan antara variabel independen dengan variabel dependen. Apakah masing-masing variabel independen berhubungan positif atau negatif, apakah nilai variabel independen mengalami kenaikan atau penurunan, dan berapa prediksi nilai variabel dependen." (p. 39)

Dalam fokus ini, variabel dependen (Y) adalah curah hujan, sedangkan faktor-faktor yang memengaruhi adalah variabel independen (X_1, X_2, \dots, X_n). Analisis regresi linier berganda memiliki kelebihan ketika dipakai meneliti tentang hubungan antara variabel dependen dan variabel independen. Seperti yang dikemukakan oleh Kartiningrum, et al. (2022):

"Kelebihan regresi linier berganda (*multiple linear regression*) adalah: 1) dapat menentukan karakteristik dan kekuatan hubungan antara variabel independen dan dependen; 2) dapat menentukan formula hitung untuk menjelaskan variabel dependen; 3) dapat menjelaskan kuantitas dan kualitas hubungan antara beberapa variabel independen; 4) dapat menentukan variabel independen yang signifikan dan tidak untuk menjelaskan variabel dependen; 5) dapat menentukan model matematika paling baik untuk menjelaskan hubungan antara variabel dependen dan independen; 6) dapat membandingkan beberapa hubungan regresi yang diperoleh; 7) dapat menentukan efek interaksi dari dua atau lebih variabel independen yang signifikan terhadap variabel dependen; 8) dapat memprediksi satu variabel dependen dengan menggunakan n buah variabel independen; 9) dapat mengetahui besarnya hubungan masing-masing variabel independen dengan variabel dependen melalui pengontrolan variabel independen yang lain; 10) dapat mengetahui variabel independen yang paling dominan." (p. 39)

Meskipun memiliki banyak kelebihan, analisis ini tetap memiliki syarat agar bisa digunakan. Indartini & Mutmainah (2024, p. 9) memaparkan salah satu syarat agar bisa menggunakan persamaan regresi linier berganda adalah terpenuhinya uji asumsi klasik. Uji ini dilakukan untuk memperoleh nilai pemeriksa yang tidak bias dan efisien atau disebut BLUE (*Best Linear Unbiased Estimator*). Anjelijani, et al. (2024) juga memaparkan bahwa uji asumsi merupakan tahapan yang sangat penting untuk diintegrasikan ke dalam proses analisis regresi linier. Dengan demikian, dalam proses pemodelan curah hujan terdapat tahapan uji asumsi klasik sebelum analisis regresi linier berganda.

Untuk mensintesis model, ditetapkan empat faktor yang akan dijadikan parameter prediksi curah hujan. Penetapan jumlah faktor tersebut merujuk pada literatur dan penelitian sebelumnya oleh Tallar (2023), Mulyani, et al. (2019), Luthfiarta, et al. (2020), dan Maulita & Nurdin (2023). Adapun faktor yang dipilih menjadi variabel independen adalah suhu, kecepatan angin, kelembapan, dan lama penyinaran. Dataset dikumpulkan melalui weatherandclimate.com dan rekam data BPS Kabupaten Karawang.

METODE

Penelitian ini dilaksanakan pada April-Juni 2025. Dataset yang dipakai merupakan data bulanan pada periode 2018-2020. Adapun variabel yang digunakan dalam pembuatan model regresi linier berganda adalah sebagai berikut:

- Curah hujan: sebagai variabel dependen (Y) didefinisikan sebagai rata-rata hujan bulanan di Kabupaten Karawang. Contoh: 2,5 mm.
- Suhu: sebagai variabel independen (X_1) didefinisikan sebagai rata-rata suhu bulanan di Kabupaten Karawang. Contoh: 27°C.
- Kecepatan angin: sebagai variabel independen (X_2) didefinisikan sebagai rata-rata kecepatan angin bulanan di Kabupaten Karawang. Contoh: 3 km/jam.
- Kelembapan: sebagai variabel independen (X_3) didefinisikan sebagai rata-rata kelembapan bulanan di Kabupaten Karawang. Contoh: 65%.
- Lama penyinaran: sebagai variabel independen (X_4) didefinisikan sebagai rata-rata lama penyinaran bulanan di Kabupaten Karawang. Contoh: 12,25 jam.

Data yang dikumpulkan kemudian diproses dengan EDA atau *Exploratory Data Analysis*. Kurniawati (2025) menjelaskan EDA digunakan untuk mengkaji struktur, pola, dan karakteristik data. Dalam kasus ini teknik EDA yang digunakan adalah analisis statistik deskriptif. Analisis statistik deskriptif dilakukan dengan bantuan Microsoft Excel, meliputi nilai rerata, median, dan standar deviasi. Hasil EDA kemudian digunakan dalam uji asumsi klasik sebelum memasuki tahapan analisis regresi linier berganda. Proses pengujian dan analisis menggunakan program SPSS.

Uji asumsi klasik dilaksanakan untuk memperoleh nilai pemeriksa yang tidak bias dan efisien atau BLUE (*Best Linear Unbiased Estimator*). Uji ini terdiri dari beberapa tahapan, yaitu uji normalitas, uji homoskedastisitas, uji autokorelasi, dan uji multikolinieritas. Masing-masing tahapan mempunyai kaidah penetapan. Berikut adalah penjelasan lebih rinci mengenai tahapan pada uji asumsi klasik:

a. Uji Normalitas

Analisis regresi linier berganda dapat dipakai jika nilai dari residual telah terdistribusi secara normal. Normalitas dapat diukur dengan melakukan *Shapiro-Wilk Test*. Untuk *Shapiro-Wilk Test*, kaidah penetapannya adalah jika nilai $p > 0,05$ maka diasumsikan data berdistribusi normal (Ghozali, 2018; Maulita & Nurdin, 2023).

H_0 : Data residual berdistribusi normal ($p > 0,05$)

H_1 : Data residual tidak berdistribusi normal dan nilai ($p < 0,05$)

$$W = \frac{(\sum a_i x_i)^2}{\sum (x_i - \bar{x})^2}$$

Dimana: W = koefisien Shapiro-Wilk; a_i = koefisien dari tabel Shapiro-Wilk; x_i = data sampel ke- i ; \bar{x} = rata-rata sampel.

b. Uji Heteroskedastisitas

Uji ini bertujuan untuk memastikan tidak terjadi korelasi antara variabel residual dengan masing-masing variabel independen. Uji dilakukan dengan menganalisis grafik *scatterplot*. Kaidah penetapan adalah jika pada grafik *scatterplot* tidak ada pola yang jelas dan titik menyebar di atas dan bawah angka 0 pada sumbu Y , maka tidak terjadi heteroskedastisitas atau variabel yang dipakai bersifat homoskedastik (Indartini & Mutmainah, 2024).

H_0 : Data terindikasi heteroskedastisitas jika ada pola tertentu pada *scatterplot*

H_1 : Data tidak terindikasi heteroskedastisitas jika titik menyebar acak tanpa pola pada *scatterplot*

c. Uji Autokorelasi

Data dengan indikasi autokorelasi menyebabkan koefisien korelasi menjadi kurang akurat. Uji autokorelasi dilakukan dengan *Runs Test*. Kaidah penetapan adalah jika nilai $p > 0,05$ maka nilai residual menyebar secara acak. Dengan demikian tidak terjadi autokorelasi dalam variabel-variabel yang diteliti (Sihabudin, et al., 2021)

H_0 : Tidak terjadi autokorelasi pada variabel-variabel yang diteliti ($p > 0,05$)

H_1 : Terjadi autokorelasi pada variabel-variabel yang diteliti ($p < 0,05$)

$$Z = \frac{(R - \mu R)}{\sigma R}$$

Dimana: R = jumlah run dalam sampel; μR = rata-rata jumlah run yang diharapkan; σR = standar deviasi jumlah run yang diharapkan.

d. Uji Multikolinieritas

Gejala multikolinieritas dihindari dalam model regresi linier berganda agar variabel-variabel yang menjelaskan model tidak berkorelasi satu sama lain. Hal tersebut vital agar pengaruh masing-masing variabel terhadap model dapat diamati secara terpisah. Gejala multikolinieritas dideteksi dengan memperhatikan nilai VIF (*Variance Inflation Factor*) dan nilai toleransi (*Tolerance*). Apabila nilai VIF < 10 , kemudian nilai *tolerance* $> 0,1$ maka diasumsikan tidak terjadi gejala multikolinieritas (Indartini & Mutmainah, 2024; Mardiatmoko, 2020).

H_0 : Tidak terjadi multikolinieritas ($VIF < 10$; *tolerance* $> 0,1$)

H_1 : Terjadi multikolinieritas ($VIF \geq 10$; *tolerance* $\leq 0,1$)

$$VIF = \frac{1}{(1 - R^2)}$$

Dimana: R^2 = koefisien determinasi.

Setelah melalui proses uji asumsi klasik, apabila ada kriteria yang tidak terpenuhi maka harus dilakukan kalibrasi data. Indartini & Mutmainah (2024) menjelaskan kalibrasi data dapat dilakukan dengan transformasi data atau dengan memperbesar ukuran sampel atau mengeliminasi variabel independen yang dipakai dan menggantinya dengan variabel yang hilang (*omitted variable*). Sebaliknya, apabila seluruh kaidah penetapan terpenuhi maka data terkategori BLUE. Data kemudian dapat dianalisis dengan regresi linier berganda untuk memperoleh model curah hujan di Kabupaten Karawang.

Berikut adalah model regresi linier berganda untuk empat variabel independen sebagaimana diuraikan oleh Reza (2024):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

Model kemudian diukur kinerjanya dengan menghitung RMSE (*Root Mean Square Error*) dan *R-Squared*.

$$RMSE = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n}} \quad \text{dan}$$

$$normalized\ RMSE = \frac{RMSE}{Y_{max} - Y_{min}}$$

Dengan (Y_i) merupakan nilai aktual curah hujan, (\hat{Y}_i) merupakan nilai prediksi model curah hujan, dan (n) adalah ukuran sampel. Model teruji akurat apabila nilai *normalized RMSE* mendekati 0 (Maulita & Nurdin, 2023).

$$R^2 = \frac{\sum_i (Y_i - \bar{Y}) - \sum_i (\hat{Y}_i - \bar{Y})^2}{\sum_i (Y_i - \bar{Y})^2}$$

Dengan (Y_i) merupakan nilai aktual curah hujan dan (\hat{Y}_i) merupakan nilai prediksi model curah hujan. *R-squared* atau koefisien determinasi diukur untuk mengetahui variasi variabel independen, dimana $0 \leq R^2 \leq 1$ (Normelia, et al., 2022).

HASIL DAN PEMBAHASAN

Dataset curah hujan sebagai variabel dependen dikumpulkan melalui rekam data BPS Kabupaten Karawang, sedangkan dataset variabel independen, yang mencakup suhu, kelembapan, kecepatan angin, dan lama penyinaran dikumpulkan melalui weatherandclimate.com. Dataset disajikan secara ringkas pada Tabel 1.

Tabel 1. Dataset iklim bulanan Kabupaten Karawang

No	Tahun	Bulan	Suhu	Kelembapan	Kecepatan Angin	Lama Penyinaran	Curah Hujan
1	2018	Januari	24,25	75,42	5,81	12,44	2,35
2		Februari	24,38	78,64	3,89	12,31	5,42
3		Maret	24,92	76,65	3,39	12,14	5,25
4		April	25,79	74,00	3,63	11,96	2,06
5		Mei	25,90	71,97	4,06	11,83	1,28
...
...
...
34	2020	Oktober	23,79	83,03	4,97	12,24	29,35
35		November	24,99	79,77	4,57	12,40	18,66
36		Desember	24,05	81,55	4,78	12,48	23,93

Data yang diperoleh kemudian dieksplorasi untuk memberikan gambaran umum mengenai sebaran data. Menggunakan Microsoft Excel diperoleh rerata, median, standar deviasi, dan varians sebagaimana dicantumkan pada Tabel 2.

Tabel 2. Nilai Statistik Deskriptif

	Suhu	Kelembapan	Kecepatan Angin	Lama Penyinaran	Curah Hujan
Rerata	24,59	76,57	4,72	12,11	12,35
Standar Deviasi	1,02	6,34	0,98	0,25	12,63
Varians	1,05	40,22	0,95	0,06	159,48
Median	24,41	77,64	4,50	12,10	5,93

Berdasarkan Tabel 2. variabel independen cenderung memiliki nilai rerata yang cukup dekat dengan nilai median, sedangkan untuk variabel dependen nilai reratanya memiliki jarak yang cukup jauh dengan nilai median. Hal tersebut dipengaruhi oleh nilai standar deviasi. Standar deviasi menggambarkan sebaran data dan jarak setiap titik data dari nilai rerata (Ananda, R., & Fadhli, M., 2018).

Curah hujan sebagai variabel dependen memiliki standar deviasi paling besar di antara yang lain, ini berarti setiap titik data terpisah cukup jauh dari nilai reratanya sehingga memengaruhi pula jarak nilai tengah ke rerata. Hal tersebut mengimplikasikan nilai rerata curah hujan tidak representatif untuk keseluruhan distribusi. Setelah mendapati nilai rerata curah hujan tidak representatif, kemudian dilakukan uji asumsi klasik. Pengujian ini bertujuan untuk mengetahui secara detail apakah data terkategori BLUE.

Uji Asumsi Klasik

Uji Normalitas

Hasil pengecekan normalitas dengan *Shapiro-Wilk Test* disajikan pada Gambar 1. Diperoleh nilai $p = 0,375 > 0,05$ sehingga memenuhi kaidah penetapan dan distribusi dinyatakan normal.

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	.093	36	.200 [*]	.968	36	.375

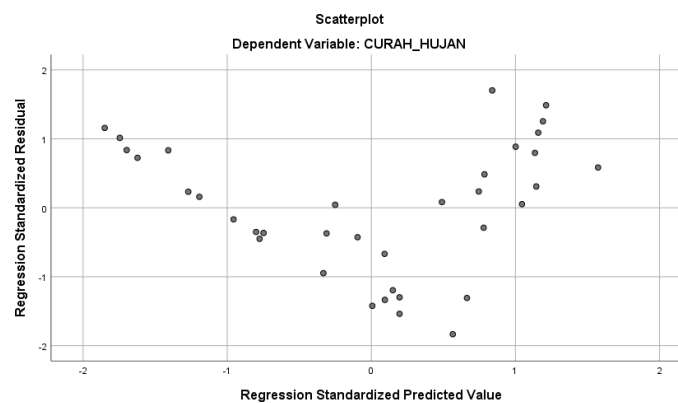
*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Gambar 1. Hasil *Shapiro-Wilk Test*

Uji Homoskedastisitas

Gejala heteroskedastisitas dapat dideteksi apabila *scatterplot* menunjukkan pola tertentu. Hal tersebut dihindari agar data terkategori BLUE. Hasil pengecekan *scatterplot* disajikan pada **Gambar 2**.



Gambar 2. Hasil pengecekan *scatterplot*

Dari hasil yang diperoleh, teramati ada pola yang muncul yaitu kerucut terbalik. Hal ini menunjukkan bahwa terdapat gejala heteroskedastisitas pada variabel curah hujan. Sehingga data tidak terkategori BLUE.

Uji Autokorelasi

Hasil pengecekan uji autokorelasi ditunjukkan pada **Gambar 3**. Diperoleh nilai $p = 0,01 < 0,05$ yang berarti residual tidak menyebar secara acak. Karena tidak memenuhi kaidah penetapan maka terindikasi ada autokorelasi antar variabel. Hal ini menunjukkan bahwa data tidak terkategori BLUE di dua uji, yakni uji homoskedastisitas dan uji autokorelasi.

Runs Test

	Unstandardized Residual
Test Value ^a	.48111
Cases < Test Value	18
Cases >= Test Value	18
Total Cases	36
Number of Runs	11
Z	-2.536
Asymp. Sig. (2-tailed)	.011

a. Median

Gambar 3. Hasil *Runs Test*

Uji Multikolinieritas

Hasil pengecekan uji multikolinieritas ditunjukkan pada **Gambar 4**.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-239.023	84.426		-2.831	.008		
	SUHU	.565	3.078	.046	.184	.856	.144	6.925
	KELEMBAPAN	1.568	.499	.787	3.144	.004	.143	6.985
	KEC_ANGIN	-1.165	1.677	-.090	-.695	.492	.535	1.870
	LAMA_PENYINARAN	10.148	6.704	.204	1.514	.140	.494	2.024

a. Dependent Variable: CURAH_HUJAN

Gambar 4. Hasil Uji Multikolinieritas

Teramati untuk variabel suhu, nilai VIF = 6,925 < 10 dan nilai *tolerance* = 0,144 > 0,1. Untuk variabel kelembapan, VIF = 6,985 < 10 dan nilai *tolerance* = 0,143 > 0,1. Untuk variabel kecepatan angin, VIF = 1,870 < 10 dan nilai *tolerance* = 0,535 > 0,1. Untuk variabel lama penyinaran, VIF = 2,024 < 10 dan nilai *tolerance* = 0,494 > 0,1. Semua variabel independen terindikasi bebas dari gejala multikolinieritas.

Kalibrasi Data

Dari uji asumsi klasik yang dilakukan ditemukan bahwa data gagal memenuhi homoskedastisitas dan terindikasi ada autokorelasi. Dengan demikian, data disimpulkan tidak terkategori BLUE, sehingga langkah selanjutnya adalah kalibrasi data. Peneliti memutuskan untuk mentransformasi data menjadi bentuk logaritma basis 10. Transformasi dilakukan pada seluruh variabel, baik dependen maupun independen. Dataset yang telah ditransformasi disajikan secara ringkas pada Tabel 3.

Tabel 3. Dataset yang ditransformasi ke dalam logaritma basis 10

No	Tahun	Bulan	Suhu	Kelembapan	Kecepatan Angin	Lama Penyinaran	Curah Hujan
1	2018	Januari	1,38	1,88	0,76	1,09	0,37
2		Februari	1,39	1,90	0,59	1,09	0,73
3		Maret	1,40	1,88	0,53	1,08	0,72
4		April	1,41	1,87	0,56	1,08	0,31
5		Mei	1,41	1,86	0,61	1,07	0,11
...
...
...
34	2020	Oktober	1,38	1,92	0,70	1,09	1,47
35		November	1,40	1,90	0,66	1,09	1,27
36		Desember	1,38	1,91	0,68	1,10	1,38

Data yang telah ditransformasi kembali melalui proses uji asumsi klasik. Hal ini ditujukan untuk melihat apakah kalibrasi berhasil. Apabila kalibrasi berhasil, maka data akan memenuhi semua kriteria pada uji asumsi klasik dan terkategori BLUE.

1.1.1 Uji Normalitas

Hasil pengecekan normalitas dengan *Shapiro-Wilk Test* disajikan pada Gambar 5. Diperoleh nilai $p = 0,267 > 0,05$ sehingga memenuhi kaidah penetapan dan distribusi dinyatakan normal.

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	.105	36	.200*	.963	36	.267

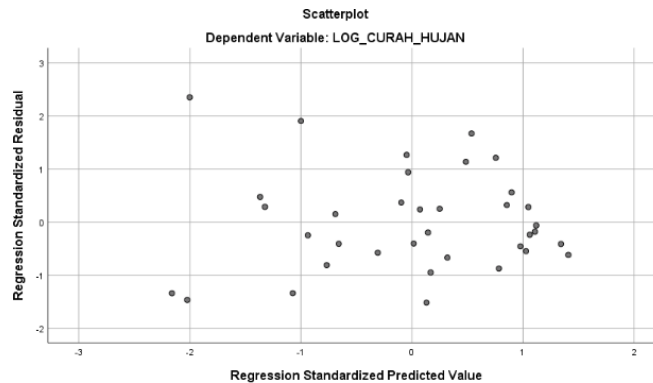
*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Gambar 5. Hasil Shapiro-Wilk Test

Uji Homoskedastisitas

Hasil pengecekan *scatterplot* disajikan pada Gambar 6. Teramati tidak ada pola yang muncul dan data tersebar random. Hal ini menunjukkan bahwa tidak ada gejala heteroskedastisitas pada variabel curah hujan.



Gambar 6. Hasil pengecekan *scatterplot*

Uji Autokorelasi

Hasil pengecekan uji autokorelasi ditunjukkan pada Gambar 7. Diperoleh nilai $p = 0,063 > 0,05$ yang artinya nilai residual menyebar secara acak. Karena memenuhi kaidah penetapan maka dapat diasumsikan tidak ada autokorelasi antar variabel.

Runs Test

	Unstandardized Residual
Test Value ^a	-.05190
Cases < Test Value	18
Cases >= Test Value	18
Total Cases	36
Number of Runs	13
Z	-1.860
Asymp. Sig. (2-tailed)	.063

a. Median

Gambar 7. Hasil *Runs Test*

Uji Multikolinieritas

Hasil pengecekan uji multikolinieritas ditunjukkan pada Gambar 8.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-85.223	11.610		-7.340	.000		
	LOG_SUHU	17.227	6.835	.410	2.520	.017	.139	7.171
	LOG_KELEMBAPAN	25.106	3.319	1.222	7.564	.000	.141	7.089
	LOG_KEC_ANGIN	1.744	.719	.202	2.425	.021	.529	1.891
	LOG_LAMA_PENYINARAN	12.498	7.294	.150	1.714	.097	.478	2.094

a. Dependent Variable: LOG_CURAH_HUJAN

Gambar 8. Hasil Uji Multikolinieritas

Teramati untuk variabel suhu, nilai $VIF = 7,171 < 10$ dan nilai $tolerance = 0,139 > 0,1$. Untuk variabel kelembapan, $VIF = 7,089 < 10$ dan nilai $tolerance = 0,141 > 0,1$. Untuk variabel kecepatan angin, $VIF = 1,891 < 10$ dan nilai $tolerance = 0,529 > 0,1$. Untuk variabel lama penyinaran, $VIF = 2,094 < 10$ dan nilai $tolerance = 0,478 > 0,1$. Semua variabel independen terindikasi bebas dari gejala multikolinieritas.

Dengan demikian, data terkategori BLUE karena memenuhi seluruh kriteria. Sehingga langkah selanjutnya adalah analisis regresi linier berganda.

Analisis Regresi Linier Berganda

Model disintesis berdasarkan hasil hitung tabel *Coefficients* yang ditunjukkan pada Gambar 9. Konstanta dan koefisien variabel dapat dilihat pada kolom *Unstandardized B*. Model persamaan regresi linier berganda diformulasikan sebagai berikut:

$$\log Y = -85,223 + 17,227 \log X_1 + 25,106 \log X_2 + 1,744 \log X_3 + 12,498 \log X_4$$

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-85.223	11.610		-7.340	.000
	LOG_SUHU	17.227	6.835	.410	2.520	.017
	LOG_KELEMBAPAN	25.106	3.319	1.222	7.564	.000
	LOG_KEC_ANGIN	1.744	.719	.202	2.425	.021
	LOG_LAMA_PENYINARAN	12.498	7.294	.150	1.714	.097

a. Dependent Variable: LOG_CURAH_HUJAN

Gambar 9. Hasil hitung koefisien model regresi linier berganda

Pada kolom Sig. terpantau hanya variabel lama penyinaran dengan $p - value = 0,097$ yang nilainya lebih besar dari α (0,05). Ini berarti variabel lama penyinaran dalam lingkup kolektif tidak berpengaruh signifikan terhadap variabel curah hujan. Sedangkan tiga variabel independen lainnya yakni suhu, kelembapan, dan kecepatan angin memiliki $p - value < 0,05$. Yang berarti tiga variabel tersebut berpengaruh signifikan terhadap variabel curah hujan.

Evaluasi Model

Untuk mengukur kinerja model regresi linier berganda maka perlu diperhatikan nilai RMSE dan *R-squared*. Pengecekan *R-squared* dilakukan menggunakan SPSS dan hasilnya disajikan pada Gambar 10.

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics			Sig. F Change
						F Change	df1	df2	
1	.941 ^a	.886	.871	.27143	.886	60.128	4	31	.000

a. Predictors: (Constant), LOG_LAMA_PENYINARAN, LOG_KELEMBAPAN, LOG_KEC_ANGIN, LOG_SUHU

b. Dependent Variable: LOG_CURAH_HUJAN

Gambar 10. Hasil pengecekan *R-squared*

Pada kolom *Adjusted R-square* diperoleh nilai 0,871 atau $0 \leq R^2 = 0,871 \leq 1$ sehingga model terindikasi berkinerja baik. Nilai 0,871 atau sebesar 87,1% menunjukkan seberapa besar variabel independen mampu menjelaskan keragaman terhadap variabel dependen. Sedangkan 12,9% sisanya dipengaruhi oleh faktor-faktor lain di luar variabel independen yang diteliti.

Untuk RMSE, perhitungan dilakukan dengan Microsoft Excel. Hasil yang diperoleh adalah sebagai berikut:

$$RMSE = \sqrt{\frac{\sum(Y_i - \hat{Y}_i)^2}{n}} = \sqrt{\frac{2,28396}{36}} = 0,252$$

$$normalized\ RMSE = \frac{RMSE}{Y_{max} - Y_{min}} = \frac{0,252}{2,767} = 0,091$$

Nilai *normalized RMSE* sebesar 0,091 memiliki artian bahwa model teruji akurat karena mendekati 0. Dengan demikian, model teruji prediktif dan akurat untuk memprediksi curah hujan di Kabupaten Karawang.

SIMPULAN

Berdasarkan hasil pembahasan, formulasi model regresi linier berganda untuk memprediksi curah hujan di Kabupaten Karawang adalah sebagai berikut:

$$\log Y = -85,223 + 17,227 \log X_1 + 25,106 \log X_2 + 1,744 \log X_3 + 12,498 \log X_4$$

Variabel lama penyinaran dengan $p - value = 0,097$ menjadi satu-satunya variabel yang tidak signifikan terhadap variabel curah hujan dalam konteks kolektif. Meskipun demikian, kemampuan prediksi model terukur baik dengan $R-square$ sebesar 87,1% atau relatif mendekati 100%. Hal ini berarti suhu, kelembapan, kecepatan angin dan lama penyinaran secara bersamaan mampu dengan baik menjabarkan keragaman curah hujan di Kabupaten Karawang. Model juga memiliki akurasi yang terindikasi baik dengan nilai *normalized* RMSE sebesar 0,091 atau mendekati 0. Sehingga dapat disimpulkan model yang dihasilkan prediktif dan akurat.

DAFTAR PUSTAKA

- Anjeliyani, A., Fatimah, F., & Casmat, M. (2024). Pemodelan Regresi Linear untuk Memprediksi Nilai Penjualan di PT Goodiebag Custom Indonesia Tanpa Mempertimbangkan Uji Asumsi. *DESANTA (Indonesian of Interdisciplinary Journal)*, 5(1), 1–13. <https://jurnal.desantapublisher.com/index.php/desanta/article/view/336>
- Ananda, R., & Fadhli, M. (2018). *Statistik Pendidikan: teori dan praktik dalam Pendidikan*. CV. Widya Puspita.
- Casallas-García, A., Hernández-Deckers, D., & Mora-Páez, H. (2023). Understanding convective storms in a tropical, high-altitude location with in-situ meteorological observations and GPS-derived water vapor. *Atmósfera*, 36(2), 225-238. <https://doi.org/10.20937/atm.53051>
- Fiola, E., Yulius, F., Presilia, P., Risani, D., Alvionita, M., & Irawati, F. (2024). Metode Seleksi Variabel dalam Pemodelan Regresi Linear Data Curah Hujan Provinsi Lampung. *PROSIDING SEMINAR NASIONAL SAINS DATA*, 4(1), 351-366. <https://doi.org/10.33005/senada.v4i1.213>
- Ghozali, I. (2018). *Aplikasi Analisis Multivariate dengan Program IBM SPSS 25*. Badan Penerbit Universitas Diponegoro.
- Indartini, M., & Mutmainah, M. (2019). *Analisis Data Kuantitatif: Uji Instrumen, Uji Asumsi Klasik, Uji Korelasi dan Regresi Linier Berganda*. Lakeisha.
- Kartiningrum, E. D., Notobroto, H. B., Otok, B. W., Kumarijati, N. E., & Yuswatiningsih, E. (2022). *Aplikasi Regresi dan Korelasi dalam Analisis Data Hasil Penelitian*. STIKes Majapahit Mojokerto.
- Katalog BPS Kabupaten Karawang Tahun 2023 Tentang Kabupaten Karawang Dalam Angka. <https://karawangkab.go.id/sites/default/files/pdf/Kabupaten%20Karawang%20Dalam%20Angka%202023.pdf>
- Katalog BPS Volume 15 Tahun 2024 Tentang Statistik Potensi Desa Indonesia. <https://www.bps.go.id/id/publication/2024/12/10/2f5217e2d6a695a0830290a7/statistik-potensi-desaindonesia-2024.html>
- Kurniawati, Y. E. (2025, January 21). *What is Exploratory Data Analysis?* <https://sis.binus.ac.id/2025/01/21/39476/>
- Luthfiarta, A., Febriyanto, A., Lestiawan, H., & Wicaksono, W. (2020). Analisa Prakiraan Cuaca dengan Parameter Suhu, Kelembaban, Tekanan Udara, dan Kecepatan Angin Menggunakan Regresi Linear Berganda. *JOINS (Journal of Information System)*, 5(1), 10–17. <https://doi.org/10.33633/joins.v5i1.2760>
- Maulita, M., & Nurdin, N. (2023). PENDEKATAN DATA MINING UNTUK ANALISA CURAH HUJAN MENGGUNAKAN METODE REGRESI LINEAR BERGANDA (STUDI KASUS: KABUPATEN ACEH UTARA). *IDEALIS : InDonEsiA Journal Information System*, 6(2), 99–106. <https://doi.org/10.36080/idealisis.v6i2.3034>
- Mardiatmoko, G. (2020). PENTINGNYA UJI ASUMSI KLASIK PADA ANALISIS REGRESI LINIER BERGANDA (STUDI KASUS PENYUSUNAN PERSAMAAN ALLOMETRIK KENARI MUDA [CANARIUM INDICUM L.]). *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 14(3), 333-342. <https://doi.org/10.30598/barekengvol14iss3pp333-342>
- Mulyani, E. D. S., Septianingrum, I., Nurjanah, N., Rahmawati, R., Nurhasani, S., & Milky R K, K. (2019). Prediksi Curah Hujan di Kabupaten Majalengka dengan Menggunakan Algoritma Regresi. *E-JURNAL JUSITI : Jurnal Sistem Informasi Dan Teknologi Informasi*, 8(1), 67–77. <https://doi.org/10.36774/jusiti.v8i1.602>
- Mulyono, J. & Fauzi, C. (2020). *Setengah Abad HCV Bendungan Ir. H. Djuanda*. Kementerian Pekerjaan Umum.
- Normelia, R., Dewi, T., Prihana, E., & Widodo, E. (2022). Penerapan metode regresi linier berganda untuk memperkirakan curah hujan (studi kasus: Stasiun Geofisika Sleman). *Jurnal Ilmiah Matematika*, 9(1), 8–18. <https://doi.org/10.26555/konvergensi.v9i1.23054>

- Padilah, T. N. & Adam, R. I. (2019). ANALISIS REGRESI LINIER BERGANDA DALAM ESTIMASI PRODUKTIVITAS TANAMAN PADI DI KABUPATEN KARAWANG. *FIBONACCI: Jurnal Pendidikan Matematika dan Matematika*, 5(2), 117-128. <https://doi.org/10.24853/fbc.5.2.117-128>
- Purnama, M. D. & Mustafidah, M. E. (2024). Relationship Between Temperature and Humidity on Rainfall: A Multiple Linear Regression Analysis. *EMACS: Engineering, MAThematics and Computer Science*, 6(2), 151-156. <https://doi.org/10.21512/emacsjournal.v6i2.11466>
- Reza, W. (2024). *Analisis Regresi "Pendekatan Praktis dan Sistematis"*. Eureka Media Aksara.
- RKPD (Rencana Kerja Pemerintah Daerah) Kabupaten Karawang Tahun 2024. <https://www.karawangkab.go.id/sites/default/files/pdf/RKPD%20Kabupaten%20Karawang%20Tahun%202024.pdf>
- RKPD (Rencana Kerja Pemerintah Daerah) Provinsi Jawa Barat Tahun 2023. https://bappeda.jabarprov.go.id/wp-content/uploads/2023/05/RKPD-2023-V.13-Okt-22_PK21.23-WIB.pdf
- Sanusi, W. & Side, S. (2016). *Statistika untuk Pemodelan Data Curah Hujan*. Badan Penerbit UNM.
- Sihabudin, S., Wibowo, D., Mulyono, S., Kusuma, J. W., Arofah, I., Ningsi, B. A., Saputra, E., Purwasih, R., & Syaharuddin, S. (2021). *Ekonometrika Dasar: Teori dan Praktik Berbasis SPSS*. CV. Pena Persada.
- Stefanny P. S. C., & Suharsono, T. N. (2025). PREDIKSI BANJIR DI KABUPATEN KARAWANG BERDASARKAN CURAH HUJAN DENGAN METODE ARTIFICIAL NEURAL NETWORK. *Journal of Data Analytics, Information, and Computer Science*, 2(1), 10–19. <https://doi.org/10.70248/jdaics.v2i1.1404>
- Tallar, R. Y. (2023). *Dasar-Dasar Hidrologi Terapan*. Ideas Publishing.