

PENDEKATAN DATA MINING UNTUK ANALISA CURAH HUJAN MENGUNAKAN METODE REGRESI LINEAR BERGANDA (STUDI KASUS: KABUPATEN ACEH UTARA)

Maya Maulita^{1*}, Nurdin²

^{1,2} Magister Teknologi Informasi, Fakultas Teknik, Universitas Malikussaleh, Aceh, Indonesia

Email: ^{1*}mayamaulita12@gmail.com, ²nurdin@unimal.ac.id

(*: coresponding author)

Abstrak-Cuaca menjadi salah satu faktor penentu dalam melakukan suatu aktivitas. Adanya prakiraan cuaca yang memanfaatkan teknologi terkini akan menghasilkan prakiraan cuaca yang tepat dan hasilnya akan mempengaruhi berbagai sektor seperti bahari, pertanian, transportasi dan lainnya. Untuk memperkirakan cuaca dapat digunakan salah satu teknik dalam *data mining* yaitu regresi linear berganda. Adanya penelitian ini akibat cuaca yang tidak menentu di Kabupaten Aceh Utara, dengan dilakukannya analisis untuk mengetahui perubahan dari pola hujan, maka di waktu yang akan datang masyarakat dapat melakukan persiapan guna menyambut musim hujan kedepannya. Dalam melakukan penelitian ini terdapat empat parameter atau variabel yang digunakan antara lain temperatur atau suhu, kelembapan, lama penerangan matahari dan kecepatan angin. Data yang digunakan dihimpun dari *website* BMKG yang bersumber dari Stasiun Meteorologi Malikussaleh untuk Kabupaten Aceh Utara dengan rentang waktu Januari 2020-Januari 2023. Berdasarkan hasil yang didapatkan dari analisis menggunakan metode regresi linear berganda mengindikasikan parameter atau variabel kelembapan berpengaruh positif serta signifikan pada curah hujan, variabel temperature tidak berpengaruh pada curah hujan, variabel lama penerangan matahari tidak berpengaruh pada curah hujan dan variabel kecepatan angin berpengaruh negatif. Pada uji F didapatkan hasil bahwa temperature, kelembapan, lama penerangan matahari serta kecepatan angin secara bersamaan berpengaruh secara positif serta signifikan pada curah hujan. Nilai $F_{hitung} = 13.531 > F_{tabel} 3.849$ dengan nilai signifikan = 0,000. Nilai *Root Mean Square Error* (RMSE) sebesar 0.009.

Kata Kunci: Curah Hujan, *Data Mining*, Kabupaten Aceh Utara, Regresi Linear Berganda

Abstract- *Weather is one of the determining factors when engaging in activities. Having weather forecasts that utilize the latest technology will result in accurate predictions, and the outcomes will impact various sectors such as maritime, agriculture, transportation, and others. One technique used in data mining to estimate weather is multiple linear regression. This research is due to the unpredictable weather in North Aceh Regency, by conducting an analysis to determine changes in rainfall patterns, so that in the future, people can make preparations to welcome the upcoming rainy season. This research used four parameters, namely temperature, humidity, duration of sunlight, and wind speed. The data used was collected from the BMKG website, sourced from the Malikussaleh Meteorological Station for North Aceh Regency, covering the period from January 2020 to January 2023. Based from the results obtained from the analysis using multiple linear regression, the humidity parameter or variable shows a positive and significant influence on rainfall. The temperature variable does not affect rainfall. The duration of sunlight variable does not affect rainfall either, and the wind speed variable has a negative impact. Results of the F-test indicate the temperature, humidity, duration of sunlight, and wind speed collectively have a positive and significant influence on rainfall. The calculated F value is $13.531 > F_{table} 3.849$, with a significance value of 0.000. The Root Mean Square Error (RMSE) value is 0.009.*

Keywords: *Rainfall, Data Mining, North Aceh Regency, Multiple Linear Regression*

1. PENDAHULUAN

Cuaca merupakan kondisi dari atmosfer pada jangka waktu tertentu yang karakteristiknya berubah dari waktu ke waktu. Cuaca menjadi salah satu faktor penentu dalam melakukan suatu aktivitas dan cuaca mempengaruhi berbagai aspek kehidupan lainnya [1]. Pengamatan cuaca dapat dilakukan dalam jangka waktu yang singkat dikarenakan cuaca sering berubah secara tidak menentu [2] dan dalam melakukan pengamatan terhadap cuaca, manusia memiliki keterbatasan wilayah pengamatan [3]. Indonesia merupakan wilayah yang memiliki iklim tropis dan curah hujan merupakan unsur cuaca yang sangat tinggi keragamannya [4]. Di Indonesia sendiri, terdapat 407 pola dari iklim yang terbagi atas 2 jenis pola yaitu 342 pola yang termasuk dalam zona musim (ZOM) dan 65 pola lainnya yang termasuk ke dalam zona non-musim (Non ZOM) [5]. Terdapat beberapa faktor cuaca antara lain kecepatan angin, arah angin, suhu, kelembapan udara dan curah hujan [6]. Karakteristik keragaman curah hujan dipengaruhi oleh faktor topografis, orografis serta geografis dan akibat faktor tersebut, struktur dari distribusi curah hujan tidak merata diantara suatu wilayah dengan wilayah lainnya [7].

Prakiraan cuaca sudah dilakukan sejak dulu, namun biasanya hanya berdasarkan pengamatan pola kejadian, namun peramalan seperti ini terbukti tidak dapat diandalkan [8]. Prakiraan cuaca adalah proses memprediksi status atmosfer untuk wilayah atau wilayah tertentu lokasi dengan memanfaatkan teknologi terkini [9]. Ketepatan dalam memprediksi hujan akan berpengaruh di banyak sektor seperti kelautan, pertanian, transportasi, penanggulangan bencana dan lainnya [10], maka diperlukan metode dalam memprediksi curah hujan, salah satunya dengan *data*

mining. *Data mining* merupakan sebuah proses, dimana proses tersebut menggunakan satu teknik atau lebih dalam melakukan pembelajaran komputer untuk melakukan analisis serta pengotomatisan pengetahuan [11].

Dalam *data mining*, terdapat beberapa metode seperti metode regresi linear berganda, *decision tree*, *machine learning*, *neural network* ataupun *clustering*. Terdapat beberapa perbedaan antara metode tersebut, salah satunya berdasarkan fokus analisis, metode regresi linear berfokus pada pemodelan hubungan linier antara variabel dependen dan variabel independen, ini cocok untuk memahami pengaruh relatif variabel independen terhadap variabel dependen sedangkan metode *decision tree* dan *clustering*, tidak secara khusus memodelkan hubungan linier dan dapat mempelajari pola yang lebih kompleks atau non linear dalam data.

Berdasarkan skalabilitas, metode regresi linear berganda cocok untuk dataset dengan jumlah variabel relatif kecil dan dimensi yang terbatas. Namun metode lain seperti metode *machine learning* dan jaringan saraf tiruan (*neural networks*) dapat lebih efektif dan skalabel dalam menangani data dengan dimensi yang lebih tinggi dan jumlah variabel yang besar.

Metode regresi linear berganda umumnya digunakan untuk tujuan prediksi, yaitu memprediksi nilai variabel dependen berdasarkan variabel independen. Sementara itu, metode *data mining* lainnya cenderung lebih berfokus pada penemuan pengetahuan baru atau pola yang tersembunyi dalam data, seperti asosiasi, ancaman, atau *outlier*.

Analisis regresi merupakan salah satu dari sekian banyak teknik dalam *data mining* yang berbasis statistik, dimana dilakukan pemodelan data berdasarkan konsep persamaan garis linier [3]. Analisis regresi terbagi atas 3 metode yaitu metode regresi linear, metode regresi linear berganda dan metode regresi logistik [12] dan analisis regresi linear berganda merupakan metode yang dikembangkan dari analisis regresi linear sederhana [13]. Analisis regresi linear berganda digunakan sebagai sebuah metode yang dapat melakukan analisis serta pengambilan sebuah kesimpulan yang sangat penting mengenai keterkaitan ketergantungan suatu variabel dengan variabel lainnya [14]. Dalam algoritma regresi linear berganda terdapat lebih dari satu variabel bebas ($X_1, X_2, X_3, \dots, X_n$) dan satu variabel terikat (Y). Algoritma regresi linear berganda memiliki keunggulan dalam memproses peramalan dengan hasil yang efektif dan efisien [15]. Dengan menggunakan banyak variabel dalam memprediksi cuaca, maka berefek pada ketepatan serta kecepatan dalam memprediksi cuaca [16][17].

Untuk melakukan analisis dengan menggunakan regresi linear berganda digunakan aplikasi *Statistical Product and Service Solutions* atau yang dikenal sebagai SPSS. SPSS menjadi sebuah *software* yang dapat melakukan berbagai proses pengolahan data khususnya statistik dengan efisien serta memiliki *output* yang cukup akurat [18].

Adanya penelitian ini akibat cuaca yang tidak menentu di Kabupaten Aceh Utara. Perubahan cuaca yang mendadak, berefek pada banyaknya kegiatan masyarakat yang terganggu [19]. Dengan dilakukannya analisis untuk mengetahui perubahan dari pola hujan, maka di waktu yang akan datang masyarakat dapat melakukan persiapan guna menyambut musim hujan kedepannya [20]. Banyak penelitian lainnya yang menggunakan metode regresi linear berganda untuk memprediksi curah hujan, antara lain penelitian yang dilakukan oleh [2], [12], [3], dan [21]. Namun berdasarkan identifikasi yang telah dilakukan belum ditemukan penelitian terkait di Kabupaten Aceh Utara.

Penulis melakukan penelitian ini berdasarkan beberapa referensi jurnal lainnya. Salah satu referensi penelitian diambil berdasarkan penelitian [2]. Penelitian ini menggunakan beberapa variabel yaitu suhu (X_1), lamanya pencahayaan (X_2) serta curah hujan (Y). Data yang digunakan berupa data sekunder dari bulan April-Mei 2021. Didapatkan hasil penelitian berupa variabel yang berpengaruh terhadap curah hujan adalah lama penerangan (X_2) dengan kemampuan model untuk menjelaskan variabel lama penerangan (X_2) terhadap curah hujan sebesar 11.36%, sedangkan suhu (X_1) berpengaruh negatif. Sisanya 88.64% dijelaskan oleh faktor lainnya yang tidak terdapat pada penelitian ini.

Berdasarkan penelitian [12], pada penelitian ini terdapat 2 perbandingan metode yaitu metode regresi linear sederhana dan regresi linear berganda dengan variabel yang digunakan yaitu suhu (X_1) dan kelembapan udara (X_2) serta perhitungan dilakukan dengan aplikasi Microsoft Excel. Penelitian ini menghasilkan nilai RMSE pada variabel suhu (X_1) yaitu 77,42mm dan pada variabel kelembapan udara (X_2) sebesar 77,13mm. Prediksi curah hujan dengan variabel suhu menggunakan metode regresi linear sederhana menghasilkan nilai yang lebih baik daripada perhitungan nilai korelasi menggunakan variabel suhu dan kelembapan udara yang menggunakan metode regresi linear berganda.

Berdasarkan penelitian [3], pada penelitian ini digunakan 4 variabel yaitu tekanan, kelembapan, suhu dan kecepatan angin dengan menggunakan data dari tahun 2015-2017. Prediksi yang dilakukan dengan analisis regresi linear berganda didapatkan hasil R_2 yaitu 25,5%. Maka 4 variabel bebas yang digunakan dapat menjelaskan sebesar 25,5% pada variabel terikat curah hujan. Sisanya sebesar 74,5% dijelaskan oleh variabel lainnya yang tidak terdapat pada penelitian ini.

Berdasarkan penelitian [21], untuk penelitian ini digunakan variabel bebas yaitu data suhu permukaan lautan di Indonesia serta variabel terikat yaitu data curah hujan. Penelitian ini menggunakan metode penelitian regresi komponen utama. Hasil yang didapatkan pada penelitian ini adalah validasi terhadap pola dari curah hujan menunjukkan nilai RMSE yaitu 81,92% dan validasi untuk sifat hujan dimana menggunakan tabel kontingensi menjelaskan kebenaran prediksi senilai 42%.

2. METODE PENELITIAN

2.1 Regresi Linear Berganda

Secara umum bentuk persamaan dari Regresi Linear Berganda yaitu [22]:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_nX_n \quad (1)$$

Dimana:

Y : Parameter atau Variabel dependen (terikat)

a : Bilangan konstanta

X : Parameter atau Variabel independen (bebas)

b : Koefisien regresi untuk setiap variabel independen

Dalam melakukan analisis maka dilakukan beberapa tahapan pengujian meliputi:

1. Uji Multikolinearitas

Multikolinearitas adalah kondisi terjadinya hubungan yang linear secara sempurna diantara variabel bebas didalam model regresi. Model regresi yang terjadi multikolinearitas biasanya terdapat fungsi linear yang cukup sempurna pada variabel bebas pada garis linear. Hal lainnya yang dapat diamati terjadinya multikolinearitas adalah dengan memperhatikan VIF *value* dan nilai *tolerance*. Apabila VIF *value* <10, kemudian nilai *tolerance* >0,1 dapat diasumsikan tidak ada multikolinearitas [23].

2. Uji Heterokedastisitas

Heterokedastisitas adalah kondisi yang tidak memiliki kesamaan varian pada residual di semua model regresi yang diamati. Pengujian dilakukan dengan uji *glejser* yang meregresikan variabel bebas pada *absolute residual value*. Apabila terdapat nilai dari signifikansi variabel terikat dengan *absolute residual value* >0,05 dapat diasumsikan tidak ada heterokedastisitas.

3. Uji Normalitas

Dalam menentukan nilai dari residual telah terdistribusi secara normal ataupun tidak, maka dilakukan sebuah pengujian yaitu uji normalitas. Cara untuk mengetahuinya dapat dilihat dari pola tersebarnya data pada grafik *Normal P Plot*. Jika plot tersebar di sekitar *line* serta terlihat mengikuti *diagonal line* maka dapat diasumsikan model dari regresi tersebut normal dan bisa digunakan untuk melakukan prediksi variabel bebas maupun sebaliknya.

4. Uji t

Untuk menentukan apakah parameter atau variabel bebas secara fragmentaris berpengaruh pada parameter atau variabel terikat, maka dilakukan sebuah pengujian yaitu uji t. Hipotesisnya adalah:

a. H_0 dapat diterima jika nilai dari probabilitas memiliki taraf signifikan yang lebih dari 0,05 (Sig. > $\alpha_{0,05}$).

b. H_0 dapat ditolak jika nilai dari probabilitas memiliki taraf signifikan yang kurang dari 0,05 (Sig. < $\alpha_{0,05}$).

5. Uji F

Dilakukan uji F guna menguji variabel bebas dengan bersamaan untuk melihat berpengaruh atau tidak secara signifikan pada variabel terikat, hipotesisnya adalah:

a. H_0 dapat diterima jika nilai dari probabilitas memiliki taraf signifikan yang lebih dari 0,05 (Sig. > $\alpha_{0,05}$).

b. H_0 dapat ditolak jika nilai dari probabilitas memiliki taraf signifikan yang kurang dari 0,05 (Sig. < $\alpha_{0,05}$).

6. Koefisien Determinasi (R^2)

Nilai dari *R Square* digunakan sebagai presentase kontribusi variabel bebas secara bersamaan terhadap variabel terikat.

7. Root Mean Square Error (RMSE)

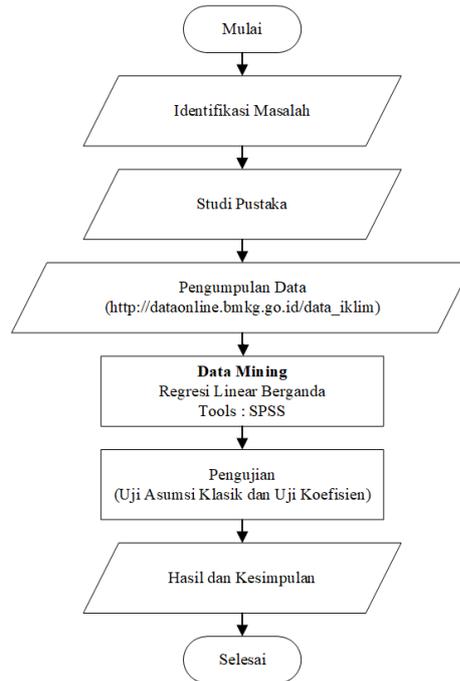
Pengujian ini dilakukan untuk menguji level kesalahan dari sebuah hasil prediksi yang telah dilakukan.

Nilai dari *RMSE* dapat diartikan akurat apabila nilai yang didapat mendekati 0 [24].

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \quad (2)$$

2.2 Tahapan Penelitian

Pada penelitian ini yang alurnya digambarkan pada Gambar 1, terdapat beberapa tahapan yang dilakukan. Dimulai dari identifikasi masalah dan dilakukan studi pustaka. Kemudian dilakukan pengumpulan data berupa data variabel bebas meliputi temperatur (X_1), kelembapan (X_2), lama penerangan matahari (X_3), serta kecepatan angin (X_4) kemudian variabel yang terikat yaitu curah hujan (Y) pada website BMKG. Digunakan sebanyak 1.114 data laporan variabel harian dan setelah seluruh data dihimpun kemudian direkapitulasi pada aplikasi excel, kemudian data tersebut di analisis menggunakan aplikasi SPSS. Setelah hasil analisis didapatkan, kemudian dilakukan tahapan pengujian berupa uji asumsi klasik dan uji koefisien untuk memastikan data sudah sesuai. Dengan didapatnya hasil serta kesimpulan dari analisis, maka penelitian selesai.



Gambar 1. Tahapan Penelitian

3. HASIL DAN PEMBAHASAN

3.1 Uji Asumsi Klasik

3.1.1 Uji Multikolinearitas

Pada tabel 1 dibawah ini, mengindikasikan bahwa setiap variabel $X_1=0,564$, $X_2=0,583$, $X_3=0,928$, $X_4=0,959$ menunjukkan persentase yang cukup baik yaitu dengan nilai toleransi diatas 0,1 atau diatas 10 persen.

Tabel 1. Uji Multikolinearitas

Variabel	Tolerance	VIF
Temperatur (X_1)	0,564	1.772
kelembapan (X_2)	0,583	1.716
lama penerangan matahari (X_3)	0,928	1.078
kecepatan angin (X_4)	0,959	1.042

Kemudian hasil dari analisis menunjukkan VIF *value* dari variabel $X_1=1.772$, $X_2=1.716$, $X_3=1.078$, $X_4=1.042$ yang nilai tersebut <10 . Dapat di asumsikan bahwa, pada model regresi yang telah terbentuk tidak ditemukan multikolinearitas pada setiap variabel yang digunakan.

3.1.2 Uji Heterokedastisitas

Pada tabel 2 dibawah ini, berdasarkan pengujian dari uji *Glejser* yang meregresikan setiap variabel bebas terhadap *absolute residual value* terlihat bahwa nilai dari probabilitas pada seluruh variabel independen adalah Sig lebih dari nilai 0,05. Hasil analisis ini merepresentasikan bahwa variabel regresinya terbebas dari gangguan heterokedastisitas.

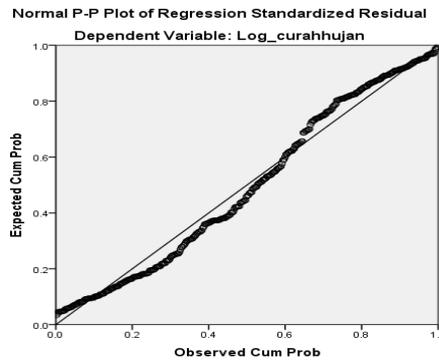
Tabel 2. Uji Heterokedastisitas

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std Error	Beta		
	(Constant)	-6.781	3.203		
Log_temperatur	3.107	1.238	.170	2.509	.213
Log_kelembapan	1.407	.930	.101	1.513	.131
Log_lamapeneranganmatahari	.096	.048	.106	2.000	.146
Log_kecepatanangin	.121	.116	.054	1.039	.299

a. Dependent Variable: Abs_RES

3.1.3 Uji Normalitas

Pada gambar 2 dibawah ini, menunjukkan bahwa pola dari *plot* variabel terikat curah hujan terlihat normal. Sebaran *plot* tersebut menunjukkan *plot* mengikuti area dari garis diagonal. Dapat disimpulkan model regresi pada uji normalitas ini telah mencukupi asumsi normalitas dan bisa digunakan.



Gambar 2. Uji Normalitas

3.2 Uji Koefisien pada Regresi Linier Berganda

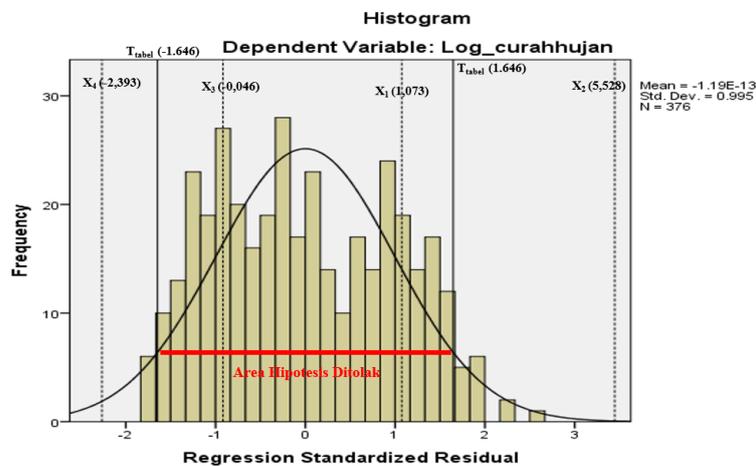
3.2.1 Uji t

Pada tabel 3 dan gambar 3 dibawah ini menunjukkan bahwa nilai variabel temperature (X_1) t_{hitung} sebesar 1.073 < 1.646 yang memiliki nilai signifikansi sebesar 0,284 > 0,05 serta lama penerangan matahari (X_3) t_{hitung} sebesar -0.918 < 1.646 yang memiliki nilai signifikansi sebesar 0,359 > 0,05 maka hipotesis ditolak. Maka temperature dan lama penerangan matahari tidak berpengaruh terhadap curah hujan. Sedangkan variabel kelembapan (X_2) t_{hitung} sebesar 5.528 > 1.646 yang memiliki nilai signifikansi sebesar 0,000 < 0,05. Artinya kelembapan berpengaruh secara positif dan signifikan pada curah hujan. Selain itu, variabel kecepatan angin (X_4) t_{hitung} sebesar -2.393 > 1.646 yang memiliki nilai signifikansi sebesar 0,017 > 0,05. Artinya kecepatan angina berpengaruh negatif dan signifikan pada curah hujan.

Tabel 3. Uji t

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std Error	Beta		
(Constant)	-22.517	6.344		-3.549	.000
Log_temperatur	2.633	2.453	.069	1.073	.284
Log_kelembapan	10.178	1.841	.351	5.528	.000
Log_lamapeneranganmatahari	-.088	.095	-.046	-.918	.359
Log_kecepatanangin	-.551	.230	-.118	-2.393	.017

a. Dependent Variable: curah hujan



Gambar 3. Grafik Hasil Uji t

3.2.2 Uji F (Uji Simultan)

Pada tabel 4 dibawah ini menunjukkan bahwa pengujian hipotesis menyatakan terdapat pengaruh secara simultan atau secara keseluruhan temperature, kelembapan, lama penerangan matahari dan kecepatan angin terhadap curah hujan yang dilihat dari hasil uji F. Nilai $F_{hitung} = 13.531 > F_{tabel} 3.849$ dengan nilai signifikan = 0,000.

Tabel 4. Uji F

ANOVA ^a					
Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	15.368	4	3.842	13.531	.000 ^b
Residual	105.340	371	.284		
Total	120.708	375			

3.2.3 Koefisien Determinasi

Pada tabel 5 dibawah ini menunjukkan nilai *adjusted* R² yaitu 0.118 yang artinya bahwa 11.8% variasi nilai curah hujan ditentukan oleh dua variabel yang berpengaruh yaitu kelembapan dan kecepatan angin.

Tabel 5. Koefisiens Determinasi

Model Summary ^b				
R	R Square	Adjusted R Square	Std. Error of the Estimate	
0.357 ^a	0.127	0.118	0.53286	

3.2.4 Analisis Regresi Linear Berganda

Tabel 6. Regresi Berganda

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std Error	Beta		
	(Constant)	-22.517	6.344		
Log_temperatur	2.633	2.453	.069	1.073	.284
Log_kelembapan	10.178	1.841	.351	5.528	.000
Log_lamapeneranganmatahari	-.088	.095	-.046	-.918	.359
Log_kecepatanangin	-.551	.230	-.118	-2.393	.017

a. Dependent Variable: curah hujan

Berdasarkan pada tabel 6 nilai yang ditunjukkan sebagai berikut:

$$Y = \beta X_1 + \beta X_2 + \beta X_3 + \beta X_4$$

$$Y = 2.633X_1 + 10.178X_2 - 0.088X_3 - 0.551X_4$$

1. Koefisien regresi temperatur sebesar 2.633. Jika variabel bebas lain nilainya tetap, kemudian temperatur mengalami kenaikan sebesar 1, maka curah hujan akan meningkat 2.633 satuan.
2. Koefisien regresi kelembapan sebesar 10.178. Jika variabel bebas lain nilainya tetap, kemudian kelembapan mengalami kenaikan sebesar 1%, maka curah hujan akan meningkat 10.178 satuan.
3. Koefisien regresi lama penerangan matahari memiliki nilai -0.088. Jika variabel bebas lain nilainya tetap, kemudian lama penerangan matahari mengalami kenaikan sebesar 1%, maka curah hujan akan menurun 0.088 satuan.
4. Koefisien regresi kecepatan angin sebesar -0.551. Jika variabel bebas lain nilainya tetap, kemudian kecepatan angin mengalami kenaikan 1%, maka curah hujan akan menurun 0.551 satuan.

3.2.5 Ukuran Error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}}$$

$$RMSE = \sqrt{\frac{(22.517 - 12.172)^2}{1114}}$$

$$RMSE = \sqrt{\frac{(22.517 - 12.172)^2}{1114}} = 0.009$$

Berdasarkan hasil persamaan tersebut dapat disimpulkan bahwa RMSE atau perhitungan standard error dalam penelitian ini yaitu 0.009. Sehingga error yang terjadi tergolong kecil karena berada dalam nilai 0,0-1,0.

4. KESIMPULAN

Berdasarkan analisis pada regresi linier berganda didapatkan persamaan yaitu $Y = 2.633X_1 + 10.178X_2 - 0.088X_3 - 0.551X_4$. Variabel temperature (X_1) tidak berpengaruh terhadap curah hujan karena $t_{hitung} 1.073 < 1.646$ dan signifikansi $0,284 > 0,05$. Variabel kelembapan (X_2) berpengaruh secara positif dan signifikan pada curah hujan karena $5.528 > 1.646$ yang memiliki nilai signifikansi sebesar $0,000 < 0,05$. Variabel lama penerangan matahari (X_3) tidak berpengaruh terhadap curah hujan karena t_{hitung} sebesar $-0.918 < 1.646$ dengan signifikansi $0,359 > 0,05$. Variabel kecepatan angin (X_4) berpengaruh negatif dan signifikan pada curah hujan karena t_{hitung} sebesar $-2.393 > 1.646$ dengan signifikansi sebesar $0,017 > 0,05$. Pada pengujian uji F bahwa temperature, kelembapan, lama penerangan matahari dan kecepatan angin secara bersamaan berpengaruh secara positif dan signifikan pada curah hujan karena $F_{hitung} = 13.531 > F_{tabel} 3.849$ dengan nilai signifikansi = $0,000$ serta didapatkan nilai koefisien determinasi yaitu 11.8% sehingga dapat disimpulkan bahwa variasi nilai curah hujan ditentukan oleh dua variabel yang berpengaruh yaitu kelembapan dan kecepatan angin.

DAFTAR PUSTAKA

- [1] M. F. Naufal, "Analisis Perbandingan Algoritma SVM, KNN, dan CNN untuk Klasifikasi Citra Cuaca," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 2, p. 311, 2021, doi: 10.25126/jtiik.2021824553.
- [2] R. Normelia, T. Dewi, E. Prihana, and E. Widodo, "Jurnal Ilmiah Matematika curah hujan (studi kasus : Stasiun Geofisika Sleman) Application of Multiple Linear Regression method to estimate rainfall (case study : Sleman Geophysics Station)," vol. 9, no. 1, pp. 8–18, 2022.
- [3] A. Luthfiarta, A. Febriyanto, H. Lestiawan, and W. Wicaksono, "Analisa Prakiraan Cuaca dengan Parameter Suhu, Kelembaban, Tekanan Udara, dan Kecepatan Angin Menggunakan Regresi Linear Berganda," *JOINS (Journal Inf. Syst.*, vol. 5, no. 1, pp. 10–17, 2020, doi: 10.33633/joins.v5i1.2760.
- [4] E. Dewi, S. Mulyani, I. Septianingrum, N. Nurjanah, and R. Rahmawati, "Prediksi Curah Hujan Di Kabupaten Majalengka Dengan Menggunakan Algoritma Regresi," *e-Jurnal JUSITI (Jurnal Sist. Inf. dan Teknol. Informasi)*, vol. 8–1, no. 1, pp. 67–77, 2019, doi: 10.36774/jusiti.v8i1.602.
- [5] M. Y. R. Rangkuti, M. V. Alfansyuri, and W. Gunawan, "Penerapan Algoritma K-Nearest Neighbor (Knn) Dalam Memprediksi Dan Menghitung Tingkat Akurasi Data Cuaca Di Indonesia," *Hexag. J. Tek. dan Sains*, vol. 2, no. 2, pp. 11–16, 2021, doi: 10.36761/hexagon.v2i2.1082.
- [6] S. S. Purwandari, M. Bettiza, and A. Uperiati, "Penerapan Algoritma Apriori Untuk Menemukan Hubungan Antara Faktor Cuaca Dan Curah Hujan (Studi Kasus: Kota Tanjungpinang)," vol. 2, no. 2020, pp. 178–183, 2022.
- [7] M. Yusuf, A. Setyanto, and K. Aryasa, "Analisis Prediksi Curah Hujan Bulanan Wilayah Kota Sorong Menggunakan Metode Multiple Regression," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 6, no. 1, pp. 405–417, 2022.
- [8] N. Anusha, M. Sai Chaithanya, and G. Jithendranath Reddy, "Weather Prediction Using Multi Linear Regression Algorithm," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 590, no. 1, 2019, doi: 10.1088/1757-899X/590/1/012034.
- [9] F. Q. Kareem, A. M. Abdulazeez, and D. A. Hasan, "Predicting Weather Forecasting State Based on Data Mining Classification Algorithms," *Asian J. Res. Comput. Sci.*, no. June, pp. 13–24, 2021, doi: 10.9734/ajrcos/2021/v9i330222.
- [10] N. Anusha, P. Radhika Priyanka, and P. Sai Harini, "Segmentation of multi-temporal images using Gaussian Mixture Model (GMM)," *ARNP J. Eng. Appl. Sci.*, vol. 13, no. 16, pp. 4633–4640, 2018.
- [11] A. M. H. Pardede *et al.*, "Application of Data Mining Prediction of Electricity Deviation Flow Using Metode Backpropagation at PLN Binjai Area," *J. Phys. Conf. Ser.*, vol. 1363, no. 1, 2019, doi: 10.1088/1742-6596/1363/1/012067.
- [12] I. Jhonson, A. Saragih, I. Rumahorbo, and R. Yudistira, "Prediksi Curah Hujan 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.
- [13] O. : Soecahyadi, *Analisa Statistik Dengan Aplikasi Spss*, Edisi Pert., no. April. Jakarta: Universitas Sahid Jakarta, 2019.
- [14] N. Nazeriandy, Y. Syahra, and M. Syaifudin, "Penerapan Data Mining Untuk Memprediksi Penggunaan Daya Listrik Pada PT.PLN (Persero) Rayon Medan Selatan Dengan Menggunakan Metode Regresi Linier Berganda," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 20, no. 1, p. 20, 2021, doi: 10.53513/jis.v20i1.2431.
- [15] N. Nurdin, F. Fajriana, M. Maryana, and A. Zanati, "Information System for Predicting Fisheries Outcomes Using Regression Algorithm Multiple Linear," *J. Informatics Telecommun. Eng.*, vol. 5, no. 2, pp. 247–258, 2022, doi: 10.31289/jite.v5i2.6023.
- [16] P. Dendy, "Data Mining Implementation On Java North Coast Weather Forecast Dataset Using C4.5 Algorithm," *J. Teknol. Pelita Bangsa*, vol. 13, no. 3, pp. 139–148, 2022.
- [17] S. Andriani, D. M. Akhmad, and F. D. Wihartiko, "Pemodelan Monte Carlo Untuk Prediksi Sifat Hujan Harian," *Comput. J. Comput. Sci. Inf. Syst.*, vol. 4, no. 2, p. 124, 2020, doi: 10.24912/computatio.v4i2.9697.
- [18] P. Pemberdayaanmasyarakat, D. Liana, R. Devianti, and F. Munawaroh, "Pengabdian dan Pemberdayaan Masyarakat," pp. 81–87, 2022.
- [19] A. Hot Iman, F. Ready Permana, G. Putro Wardana, R. Kemmy Rachmansyah, and M. Mega Santoni, "Perbandingan Algoritma Klasifikasi Random Forest dan Extreme Gradient Boosting pada Dataset Cuaca Provinsi DKI Jakarta Tahun 2018," *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, pp. 593–601, 2022, [Online]. Available:

- <https://katalog.data.go.id/dataset/data-prakiraan-cuaca-wilayah-provinsi-dki-jakarta-tahun-2018>.
- [20] I. Muhandhis, A. S. Ritonga, and M. H. Murdani, "Implementasi Metode Inferensi Fuzzy Tsukamoto Untuk Memprediksi Curah Hujan Dasarian Di Sumenep," *J. Ilm. Educat. Pendidik. dan Inform.*, vol. 8, no. 1, pp. 01–10, 2021, doi: 10.21107/edutic.v8i1.8907.
- [21] A. Pratiwi *et al.*, "Prediksi Curah Hujan Bulanan Di Wilayah Lombok Barat Menggunakan Principal Component Regression (PCR)," *J. Teor. dan Apl. Fis.*, vol. 08, no. 02, pp. 175–182, 2020, [Online]. Available: <http://iridl.ldeo.columbia.edu/>.
- [22] D. S. O. Panggabean, E. Buulolo, and N. Silalahi, "Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda," *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 1, p. 56, 2020, doi: 10.30865/jurikom.v7i1.1947.
- [23] G. Mardiatmoko, "PENTINGNYA UJI ASUMSI KLASIK PADA ANALISIS REGRESI LINIER BERGANDA (STUDI KASUS PENYUSUNAN PERSAMAAN ALLOMETRIK KENARI MUDA [CANARIUM INDICUM L .]) The Importance of the Classical Assumption Test in Multiple Linear Regression Analysis (A Case Study of ," vol. 14, no. 3, pp. 333–342, 2020.
- [24] W. S. P. Wiega Pratama, Daffa Brilliant Samodra, Hayuda Altita Perkasa, Andhika Hermawan, Favian Bhagaskara Ramadhan, "Prediksi Pemeliharaan Transformator Distribusi Berbasis Artificial Neural Network," in *SNESTIK*, 2023, pp. 112–120, doi: 10.31284/p.snestik.2023.4004.