Original Article

e-ISSN: 2774-2016 - https://journal.itera.ac.id/index.php/indojam/
p-ISSN: 2774-2067

Received 21st July 2026
Accepted 26th October 2026
Published 31st October 2026

Open Access

DOI
https://doi.org/10.35472/indojam.v6i2.3001

Prediksi Banyak Hari Hujan di Kota Jambi Menggunakan Algoritma Random Forest Regression Berbasis Data Iklim

Ferdinand Manurung[a,*], Sherli Yurinanda[a], Sarmada[a]

[a]*Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Jambi*
** Corresponding E-mail: ferdinanmanurung913@gmail.com*

Abstract: *The number of rainy days is an important climatological indicator that influences the agricultural sector, water resource management, and hydrometeorological disaster mitigation. Variations in the number of rainy days are influenced by several climatological factors, including rainfall, air humidity, air temperature, and sunshine duration (LPM), which exhibit complex and nonlinear relationships. This study aims to predict the number of rainy days in Jambi City using the Random Forest Regression method and identify the most influential climatological factors. Monthly secondary data obtained from the Sultan Thaha Class I Meteorological Station Jambi during the 2018 to 2025 period consisted of 96 observations. The research methodology includes data exploration, hyperparameter tuning using Grid Search with 5-Fold Cross Validation, Random Forest Regression model construction, model evaluation using RMSE and R^2, and feature importance analysis. The results show that the model achieved an RMSE value of 2.769 and an R^2 of 0.695 on the testing data, indicating that the model is capable of explaining 69.5% of the variance in the number of rainy days with an average prediction error of approximately three days. Feature importance analysis reveals that rainfall is the most influential factor, followed by sunshine duration (LPM), air humidity, and air temperature. These findings demonstrate the effectiveness of Random Forest Regression in modeling nonlinear climatological relationships and predicting the number of rainy days in Jambi City.*

Keywords: *random forest regression, rainy days, feature important, climatological variables, Jambi City*

Abstrak: Banyak hari hujan merupakan salah satu indikator klimatologi penting yang memengaruhi sektor pertanian, pengelolaan sumber daya air, serta mitigasi bencana hidrometeorologi. Variasi banyak hari hujan dipengaruhi oleh beberapa faktor klimatologi, antara lain curah hujan, kelembapan udara, suhu udara, dan LPM, yang memiliki hubungan kompleks dan nonlinier. Penelitian ini bertujuan untuk memprediksi banyak hari hujan di Kota Jambi menggunakan metode Random Forest Regression serta mengidentifikasi faktor-faktor klimatologi yang paling berpengaruh. Data sekunder bulanan yang diperoleh dari Stasiun Meteorologi Kelas I Sultan Thaha Jambi pada periode 2018 hingga 2025 sebanyak 96 pengamatan. Metodologi penelitian meliputi eksplorasi data, penentuan hyperparameter menggunakan Grid Search dengan 5-Fold Cross Validation, pembangunan model Random Forest Regression, evaluasi model menggunakan RMSE dan R^2, serta analisis feature importance. Hasil penelitian menunjukkan bahwa model memperoleh nilai RMSE sebesar 2,769 dan R^2 sebesar 0,695 pada data testing, yang mengindikasikan bahwa model mampu menjelaskan 69,5% variasi banyak hari hujan dengan rata-rata kesalahan prediksi sekitar tiga hari. Analisis yang dilakukan melalui feature importance menunjukkan bahwa curah hujan merupakan faktor yang paling berpengaruh, diikuti oleh LPM, kelembapan udara, dan suhu udara. Temuan ini menunjukkan efektivitas Random Forest Regression dalam memodelkan hubungan klimatologi yang nonlinier serta memprediksi banyak hari hujan di Kota Jambi.

Kata Kunci: *Random forest regression, Banyak hari hujan, Feature important, Variabel klimatologi, Kota Jambi*

Copyright © 2026 - Indonesian Journal of Applied Mathematics.
Content from this work may be used under the terms of the Creative Commons Attribution-NonCommercial 4.0 International Licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI. Published under licence by Indonesian Journal of Applied Mathematics.

Original Article

Pendahuluan

Indonesia merupakan negara beriklim tropis yang dipengaruhi oleh angin muson dan dinamika atmosfer global sehingga memiliki variasi banyak hari hujan yang cukup tinggi dari waktu ke waktu [1]. Banyak hari hujan merupakan salah satu indikator klimatologi penting karena berpengaruh terhadap sektor pertanian, pengelolaan sumber daya air, serta mitigasi bencana hidrometeorologi [2]. Oleh karena itu, informasi mengenai banyak hari hujan menjadi kebutuhan penting dalam mendukung perencanaan dan pengambilan keputusan yang lebih adaptif terhadap kondisi iklim.

Kota Jambi merupakan salah satu wilayah yang rentan terhadap perubahan pola hujan karena berada di daerah dataran rendah dan dilalui oleh Sungai Batanghari [3]. Beberapa tahun terakhir, peningkatan frekuensi hujan telah menyebabkan berbagai permasalahan, seperti banjir, genangan, gangguan aktivitas ekonomi dan transportasi, serta kerusakan lingkungan [4]. Kondisi tersebut menunjukkan pentingnya upaya prediksi banyak hari hujan yang akurat sebagai dasar dalam pengelolaan sumber daya air dan mitigasi risiko bencana hidrometeorologi di Kota Jambi.

Banyak hari hujan dipengaruhi oleh berbagai faktor klimatologi, seperti curah hujan, kelembapan udara, suhu udara, dan LPM [5]. Namun, hubungan antarvariabel tersebut bersifat kompleks dan *nonlinier* sehingga sulit dijelaskan menggunakan pendekatan statistik konvensional seperti regresi *linier* [6]. Sebagai contoh, peningkatan curah hujan tidak selalu diikuti oleh peningkatan jumlah hari hujan karena adanya interaksi antarvariabel iklim yang memengaruhi proses pembentukan hujan [7]. Permasalahan ini menyebabkan proses prediksi banyak hari hujan menjadi lebih menantang dan membutuhkan metode yang mampu menangkap pola hubungan yang kompleks.

Random Forest Regression merupakan salah satu metode *machine learning* yang mampu memodelkan hubungan *nonlinier* dan interaksi kompleks antarvariabel melalui kombinasi pohon keputusan, *bootstrap sampling*, dan proses *agregasi* prediksi [8]. Selain menghasilkan prediksi yang stabil, metode ini juga mampu mengidentifikasi tingkat kepentingan masing-masing variabel melalui analisis *feature importance*. Beberapa penelitian sebelumnya menunjukkan bahwa *Random Forest Regression* memiliki kinerja yang baik dalam pemodelan variabel iklim dan prediksi hujan [9]. Oleh karena itu, penelitian ini bertujuan untuk memprediksi banyak hari hujan di Kota Jambi menggunakan *Random Forest Regression* serta mengidentifikasi faktor klimatologi yang paling berpengaruh terhadap banyak hari hujan.

Metode

Pra-pemrosesan data

Tahap pra-pemrosesan data diawali dengan pemeriksaan terhadap nilai yang hilang (*missing values*) pada dataset. Apabila ditemukan data yang tidak tersedia, dilakukan proses imputasi menggunakan nilai rata-rata dari variabel yang bersangkutan untuk menjaga kelengkapan dan konsistensi data yang digunakan dalam pemodelan. Selanjutnya, dataset dibagi menjadi dua bagian, yaitu data pelatihan (*training data*) dan data pengujian (*testing data*) dengan perbandingan 80% data pelatihan dan 20% data pengujian [10]. Data pelatihan digunakan untuk membangun model prediksi, sedangkan data pengujian digunakan untuk mengevaluasi kinerja model terhadap data yang belum pernah digunakan selama proses pelatihan.

Hyperparameter Tuning

Hyperparameter merupakan parameter pengendali yang ditentukan sebelum proses pelatihan model dan tidak dipelajari secara langsung dari data. Pada penelitian ini, *hyperparameter* yang dioptimalkan meliputi *n_estimators*, *max_depth*, *min_samples_split*, dan *min_samples_leaf*. Proses pencarian kombinasi parameter terbaik dilakukan menggunakan metode *Grid Search* dengan *5-Fold Cross Validation* pada ruang pencarian *n_estimators* = {100, 200, 300, 400, 500}, *max_depth* = {2, 3, 4}, *min_samples_split* = {2, 5}, dan *min_samples_leaf* = {1, 2} [6].

Original Article

Setiap kombinasi dievaluasi menggunakan nilai *Mean Squared Error* (MSE) dengan rumus:

$$MSE_{fold,k} = \frac{1}{n_k}\sum_{i=1}^{n_k}\left(y_i - \hat{y}_i\right)^2 \quad (1)$$

Keterangan:

$MSE_{fold,k}$: *Nilai mean squared error* pada *fold* ke-k

n_k : Jumlah sampel pada *fold* ke-k

y_i : *variabel independen* dari data ke-i pada *fold* ke-k

$\hat{y}_i$: Nilai prediksi dari model terhadap data ke-i pada *fold* ke-k

$\left(y_i - \hat{y}_i\right)^2$: Selisih kuadrat aktual dan prediksi.

Selanjutnya, semua nilai MSE dari setiap lipatan dihitung rata-ratanya dengan rumus:

$$Mean\ MSE_{CV} = \frac{1}{n_k}\sum_{k=1}^{n_k} MSE_{fold,k} \quad (2)$$

Keterangan:

$Mean\ MSE_{CV}$: Rata-rata nilai *mean squared error* dari seluruh *fold*

n_k : Jumlah *fold*

$MSE_{fold,k}$: nilai MSE pada *fold* ke-k

dan kombinasi dengan rata-rata MSE terendah dipilih sebagai parameter terbaik.

Penerapan *Random Forest Regression*

1. Pembentukan *Subset* Data

   Data *training* diambil secara acak menggunakan metode *bootstrap sampling* (pengambilan sampel dengan pengembalian), di mana setiap subset data digunakan untuk melatih satu pohon keputusan pada model *Random Forest Regression*. Proses ini diulang berkali-kali sesuai jumlah pohon yang ditentukan sehingga menghasilkan model
   yang lebih stabil dan mampu mengurangi variansi prediksi.

2. Pembuatan Pohon Keputusan

   Pohon keputusan merupakan model berbentuk struktur pohon yang terdiri atas simpul akar, simpul internal, dan simpul daun [11]. Adapun langkah-langkah untuk membuat pohon keputusan [6].

a) Menghitung rata-rata variabel y simpul akar pohon ke-b

$$\bar{y}^{(b)} = \frac{1}{n_b}\sum_{i=1}^{n_b} y_i^{(b)} \quad (3)$$

Keterangan:

$\bar{y}^{(b)}$: Rata-rata variabel y pada simpul akar pohon ke-b

$y_i^{(b)}$: Nilai variabel y pada sampel ke-i dari subset *bootstrap* ke-b

n_b : Jumlah sampel dalam *subset bootstrap* ke-b

b) MSE simpul akar pohon ke-b

$$MSE^{(b)} = \frac{1}{n_b}\sum_{i=1}^{n_b}(y_i^{(b)} - \bar{y}^{(b)}) \quad (4)$$

Keterangan:

$MSE^{(b)}$: MSE pada simpul akar pohon ke-b sebelum *split*

$y_i^{(b)}$: Nilai variabel y pada sampel ke-i dari *subset bootstrap* ke-b

n_b : Jumlah sampel dalam subset bootstrap ke-b

$\bar{y}^{(b)}$: Rata-rata variabel y pada simpul akar dipohon

c) Penentuan *Threshold*

$$\theta = \frac{x_{(i)j}^{(b)} + x_{(i+1)j}^{(b)}}{2} \quad (5)$$

Keterangan:

θ : Nilai *threshold* untuk *split*

$x_{(i)j}^{(b)}$: Nilai variabel *independen* ke-j dari sampel ke-i dalam subset bootstrap ke-b

$x_{(i+1)j}^{(b)}$: Nilai variabel *independen* ke-j dari sampel ke-$(i+1)$ dalam subset bootstrap ke-b

j : Variabel *independen* yang sedang dipertimbangkan untuk *split*

d) Total MSE setelah *split*

$$MSE_{split}^{(b,j,\theta)} = \frac{1}{n_L+n_R}\left(\left(n_L\left(\sum_{i=1}^{n_L}\left(y_{L,i}^{(b)} - \bar{y}_L^{(b)}\right)^2\right)\right) + \left(n_R\left(\sum_{i=1}^{n_R}\left(y_{R,i}^{(b)} - \bar{y}_R^{(b)}\right)^2\right)\right)\right) \quad (6)$$

Keterangan:

Copyright © 2026 – Indonesian Journal of Applied Mathematics

Published by: Lembaga Penelitian, Pengabdian Masyarakat, dan Penjaminan Mutu Institut Teknologi Sumatera, Lampung Selatan, Indonesia

Original Article

$MSE_{split}^{(b,j,\theta)}$: MSE setelah *split* pada simpul akar dipohon ke-b menggunakan variabel *independen* ke-j *threshold* θ

n_L : Jumlah sampel *subset* kiri $S_{L,i}^{(b)}$

n : Total sampel

n_R : Jumlah sampel *subset* kanan $S_{R,i}^{(b)}$

$y_{L,i}^{(b)}$: Nilai variabel *dependen* pada sampel ke-i dari *subset* kiri $S_{L,i}^{(b)}$

$\bar{y}_L^{(b)}$: Nilai rata-rata variabel *dependen* pada *subset* kiri

$y_{R,i}^{(b)}$: Nilai variabel *dependen* pada sampel ke-i dari *subset* kanan $S_{R,i}^{(b)}$

$\bar{y}_R^{(b)}$: Nilai rata-rata variabel banyak hari hujan pada subset kanan

e) Information Gain

$$Gain^{(b,j,\theta)} = MSE^{(b)} - MSE_{split}^{(b,j,\theta)} \quad (7)$$

Keterangan:

$Gain^{(b,j,\theta)}$: *Gain* dari *split* pada *node* akar dipohon ke-b menggunakan variabel x ke-j dan *threshold* θ

$MSE^{(b)}$: MSE sebelum *split*

$MSE_{split}^{(b,j,\theta)}$: MSE setelah *split*

3. Pembentukan Model Pohon Keputusan

Model pohon keputusan dibangun melalui proses pemisahan data secara berulang pada setiap node berdasarkan variabel independen dan nilai ambang (*threshold*) tertentu untuk menghasilkan kelompok data yang lebih homogen. Secara matematis, pohon keputusan dapat direpresentasikan sebagai fungsi *piecewise* yang membagi ruang variabel independen ke dalam beberapa region, di mana setiap region berasosiasi dengan satu *leaf node* yang menghasilkan nilai prediksi konstan sesuai karakteristik data pada node tersebut [6]. Berikut bentuk model *random forest regression*:

$$\hat{y}^{(b)} = \begin{cases} c_1^{(b)} & if\ S_{1,1} \wedge S_{1,2} \dots \wedge S_{1,L1} \\ c_2^{(b)} & if\ S_{2,1} \wedge S_{2,2} \dots \wedge S_{1,L2} \\ \vdots & \vdots \\ c_K^{(b)} & if\ S_{K^{(b)},2} \wedge S_{K^{(b)},2} \dots \wedge S_{K^{(b)},L_K{}^{(b)}} \end{cases} \quad (8)$$

Keterangan:

$\hat{y}^{(b)}$: Hasil prediksi dari pohon keputusan ke-b

$c_K^{(b)}$: Nilai prediksi rata-rata hari hujan pada daun ke-K dari pohon ke-b

$K^{(b)}$: Jumlah daun pada pohon ke-b

$L_K{}^{(b)}$: Jumlah *split* yang membentuk jalur menuju daun ke-k dari pohon ke- b

$S_{K^{(b)},L_K{}^{(b)}}$: Kondisi *split* (pemisahan) logis ke-L yang dipenuhi input x agar mencapai daun ke-k pada pohon ke-b

4. *Aggregatting*

Setelah semua variabel input digunakan oleh model dalam menciptakan pohon keputusan, tiap pohon memberikan nilai prediksi berdasarkan aturan pemisahan yang telah ditetapkan. Selanjutnya, semua hasil prediksi dari pohon-pohon tersebut dikumpulkan untuk mendapatkan hasil prediksi yang final. Pada tahap ini, nilai prediksi final ditentukan dengan cara menghitung rata-rata dari semua prediksi pohon yang terkait dengan variabel input yang telah diberikan dengan rumus [8]:

$$\hat{y} = \frac{1}{B}\sum_{i=1}^{B} \hat{y}^{(b)} \quad (9)$$

Keterangan:

$\hat{y}$: Prediksi akhir

$\hat{y}^{(b)}$: Prediksi dari pohon ke-b

B : Jumlah pohon dalam hutan

5. Prediksi Menggunakan Data *Testing*

$$\hat{y}_{test} = \frac{1}{B_{trees}}\sum_{i=1}^{B_{trees}} \hat{y}^{(b)} \quad (10)$$

Keterangan:

$\hat{y}_{test}$: Nilai prediksi akhir untuk sampel data *testing*

B_{trees} : Jumlah total pohon keputusan dalam *random forest*

$\hat{y}^{(b)}$: Prediksi dari pohon keputusan ke-b untuk sampel data *testing*

6. Evaluasi Model

Evaluasi kinerja model Random Forest Regression dilakukan menggunakan metrik *Root Mean Squared*

Error (RMSE) dan koefisien determinasi. RMSE digunakan untuk mengukur rata-rata kesalahan prediksi model terhadap nilai aktual. Sementara itu, koefisien determinasi digunakan untuk mengukur kemampuan model dalam menjelaskan variasi data variabel *dependen* [12]. Berikut formula untuk menghitung RMSE dan Koefisien Determinasi:

a) RMSE

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (11)$$

Keterangan:

y_i : Nilai sebenarnya
$\hat{y}_i$: Nilai yang diprediksi model
n : Total pengamatan

b) Koefisien Determinasi

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (12)$$

Keterangan:

$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$: Jumlah kuadrat dari selisih antara nilai prediksi dan nilai *aktual*
$\sum_{i=1}^{n}(y_i - \bar{y})^2$: Jumlah kuadrat dari selisih antara nilai *aktual* dan rata-rata
y_i : Nilai akual pada data ke-i
$\hat{y}_i$: Nilai prediksi pada data ke-i
$\bar{y}$: Nilai rata-rata *aktual*

Feature Important

Feature Importance digunakan untuk mengukur kontribusi setiap variabel independen terhadap kemampuan prediksi model. Pada *Random Forest Regression*, nilai *feature importance* diperoleh dengan mengakumulasi kontribusi setiap variabel dari seluruh pohon keputusan berdasarkan pengurangan *Mean Squared Error* (MSE) pada setiap pemisahan (*split*) [6]. Nilai tersebut kemudian dirata-ratakan untuk menentukan tingkat kepentingan masing-masing variabel dengan rumus:

$$FI(j) = \frac{n_{t,j}}{n_b}\left(MSE_{before} - MSE_{split}\right) \quad (13)$$

Keterangan:
$FI(j)$: Variabel *independen* ke-j yang dihitung
$n_{t,j}$: Jumlah sampel pada *node* t yang menggunakan j sebagai split
n_b : Total sampel pada *bootstrap* ke-b
MSE_{before} : MSE sebelum *split*
MSE_{split}: MSE sesudah *split*

Hasil dan Diskusi

Pra-pemrosesan Data

Sebelum proses pemodelan, dilakukan pemeriksaan *missing values* untuk memastikan kualitas data yang digunakan. *Missing values* disajikan pada **Tabel 1** di bawah ini.

Tabel 1. *Missing Value*

Variabel	Jumlah missing values
Jumlah Hari Hujan (HH) (y)	0
Curah Hujan (CH) (x_1)	0
Kelembapan Udara (KU) (x_2)	0
Suhu Udara (SU) (x_3)	0
Lama Penyinaran Matahari (LPM) (x_4)	0

Berdasarkan **Tabel 1**, seluruh variabel yang digunakan dalam penelitian menunjukkan 0 *missing values*, sehingga imputasi tidak perlu dilakukan karena data sudah lengkap. Selanjutnya, dataset dibagi menjadi dua bagian, yaitu data pelatihan sebesar 80% (76 pengamatan) dan data pengujian sebesar 20% (20 pengamatan) menggunakan teknik *Random Holdout Split* (pembagian acak) [10]. Pembagian secara acak ini dipilih untuk menguji kemampuan model *Random Forest Regression* dalam memetakan interaksi *nonlinier* antarvariabel iklim secara *independen* pada setiap observasi bulanan tanpa mengasumsikan ketergantungan temporal yang ketat Seluruh proses pengolahan data, pra-pemrosesan, pembentukan model *Random Forest Regression*, hingga evaluasi kinerja model

Copyright © 2026 – Indonesian Journal of Applied Mathematics

Published by: Lembaga Penelitian, Pengabdian Masyarakat, dan Penjaminan Mutu Institut Teknologi Sumatera, Lampung Selatan, Indonesia

Original Article

diimplementasikan menggunakan bahasa pemrograman Python pada *platform Google Colab*.

Hyperparameter Tuning

Sebelum membangun model *Random Forest Regression*, dilakukan *hyperparameter tuning* untuk memperoleh konfigurasi parameter yang optimal. Hyperparameter yang digunakan meliputi *max_depth* = (2, 3, 4), *min_samples_split* = (2, 5), *min_samples_leaf* = (1, 2), dan *n_estimators* = (100, 200, 300, 400, 500), sehingga menghasilkan 60 kombinasi parameter. Evaluasi setiap kombinasi dilakukan menggunakan *Grid Search* dengan *5-Fold Cross Validation* berdasarkan nilai rata-rata *Mean Squared Error* (MSE), dan kombinasi dengan nilai MSE terendah dipilih sebagai model terbaik. Hasil proses tuning disajikan pada **Tabel 2**.

Tabel 2. *Hyperparameter Tuning*

Max depth	*Min Samples Leaf*	*Min Samples Split*	*n Estimator*	*Mean Test MSE*
2	2	2	400	13,9032
3	2	5	500	13,1904
2	2	5	400	13,8320
3	1	2	100	13,7843
4	2	2	400	13,5761
4	1	5	100	13,3196
2	1	2	300	13,9458
2	1	2	400	13,9389
3	1	2	500	13,6542
4	2	5	500	13,2233
4	1	5	200	13,3797
3	1	2	400	13,7478
3	1	2	200	13,7835
3	2	5	100	13,4759
⋮	⋮	⋮	⋮	⋮
3	**1**	**5**	**500**	**13,1217**

Berdasarkan **Tabel 2** di atas, kombinasi *hyperparameter* paling optimal diperoleh pada *max_depth* = 3, *min_samples_leaf* = 1, *min_samples_split* = 5, dan *n_estimators* = 500 dengan nilai *Mean Test* MSE sebesar 13,121716.

Penerapan *Random Forest Regression*

1. Pembentukan Subset Data

Setelah diperoleh kombinasi *hyperparameter* yang optimal, tahap selanjutnya adalah menerapkan teknik *bootstrap sampling* dalam proses pembentukan model *Random Forest Regression*. Berikut adalah hasil *bootstrap sampling* pada **Tabel 3**.

Tabel 3. *Subset Bootstrap Sampling* pertama

CH	KU	LPM	SU	HH	*Bootstrap*
145	84	48	27	24	1
150	83	50	27,1	20	1
386	87	41	27,5	21	1
244	85	63	26,8	16	1
191	85	27	26,3	26	1
265	83	58	26,4	20	1
220	85	47	26,5	21	1
152	83	73	27,1	22	1
239	82	52	27,3	20	1
87	84	61	27,5	10	1
326	86	31	26,9	23	1
161	86	50	27,6	16	1
⋮	⋮	⋮	⋮	⋮	⋮
109	83	45	27,6	22	1

2. Pembuatan Pohon keputusan Pertama

a) Menghitung rata-rata variabel y pada *subset bootstrap* pertama

$$\bar{y}^{(1)} = \frac{1}{76}\sum_{i=1}^{76} y_i^{(1)}$$

$$\bar{y}^{(1)} = \frac{1}{76}(24 + 20 + 21 + 16 + 26 + 20 + 21 + 22 + 20 + 10 + 23 + 16 + 16 + 20 + 5 + 19 + 5 + 22 + 21 + 15 + 17 + 6 + 21 + 18 + 26 + 23 + 24 + 12 + 19 + 24 + 21 + 17 + 24 + 24 + 8 + 19 + 21 + 15 + 19 + 1 + 22 + 21 + 10 + 17 + 13 + 25 + 13 + 1 + 8 + 13 + 24 + 17 + 18 + 19 + 16 + 26 + 18 + 17 + 23 + 21 + 21 + 15 + 17 + 18 + 14 + 13 + 23 + 17 + 12 + 24 + 21 + 21 + 26 + 16 + 2 + 22$$

$$\bar{y}^{(1)} = 17{,}776$$

Hasil perhitungan rata-rata nilai banyak hari hujan pada *subset bootstrap* ke-1 yang digunakan dalam pembentukan *node* akar diperoleh sebesar 17,776 atau 18 hari.

Original Article

b) MSE simpul akar pohon pertama

$$MSE^{(1)} = \frac{1}{n_b}\sum_{i=1}^{n_b}(y_i^{(1)} - \bar{y}^{(1)})$$

$$MSE^{(1)} = \frac{1}{76}[(24-17{,}776)^2 + (20-17{,}776)^2 + (21-17{,}776)^2 + (16-17{,}776)^2 + (26-17{,}776)^2 + (20-17{,}776)^2 + \cdots + (22-17{,}776)^2]$$

$$MSE^{(1)} = 34{,}3052$$

Hasil perhitungan nilai MSE sebelum *split* sebesar 34,3052. Nilai ini yang digunakan untuk menghitung nilai *information gain* dalam mengidentifikasi seberapa besar peningkatan homogenitas yang dihasilkan pada proses *split*.

c) Penentuan *Threshold* pohon pertama

Penentuan *threshold* dilakukan dengan mengurutkan nilai pada setiap variabel independen dari yang terkecil hingga terbesar. Kemudian, setiap pasangan nilai yang berurutan dihitung nilai tengahnya yang akan digunakan sebagai kandidat *threshold*. Hasil berupa nilai *threshold* dari masing-masing variabel *independen* beserta nilai MSE kiri dan MSE kanan disajikan pada **Tabel 4**.

Tabel 4. *Threshold Bootstrap* pertama

Variabel	Threshold	Jumlah Sampel Kiri (n_L)	Jumlah Sampel Kanan (n_R)	MSE Kiri	MSE Kanan
CH	17	2	74	0,0000	27,4201
CH	35	4	72	4,0000	23,1849
CH	43,5	5	71	4,6400	21,2449
CH	55,5	9	67	31,1358	18,8135
CH	66	10	66	30,6900	18,5169
CH	73	11	65	28,1983	17,4674
CH	**79**	**12**	**64**	**26,2430**	**14,8974**
CH	84	14	62	32,2295	15,1540
CH	91	15	61	30,1066	13,8419
CH	102	16	60	30,0625	13,6763
CH	117,5	17	59	36,6020	13,8333
CH	132	18	58	36,8055	13,9262
CH	141,5	19	57	40,0221	14,1502
CH	147,5	21	55	49,9591	14,0350
CH	150,5	28	48	41,1670	10,1666
CH	151,5	29	47	40,5445	10,2471
CH	152,5	31	45	42,5223	10,6054
CH	154	33	43	41,4618	10,9908
CH	156,5	34	42	42,4775	11,1071
⋮	⋮	⋮	⋮	⋮	⋮
SU	27,95	75	1	34,7544	0,0000

Tabel 4 di atas memperlihatkan nilai dari *threshold* pada *subset bootstrap* ke-1 yang dihitung berdasarkan persamaan (5).

d) MSE total setelah split

Pada *bootstrap* ke-1 ($b = 1$), variabel independen ke-1 ($j = 1$), dan nilai *threshold* $\theta = 79$, yang ada pada **Tabel 4** dengan jumlah sampel pada cabang kiri (n_L) sebanyak 12 sampel dengan nilai MSE 26,2430, dan pada cabang kanan (n_R) sebanyak 64 sampel dengan nilai MSE 14,8974.

$$MSE_{split}^{(1.1.79)} = \frac{1}{12+64}(12\,.\,26{,}2430 + 64\,.\,14{,}8974) = 16{,}6888$$

Dari hasil perhitungan di atas, didapat nilai MSE sebesar 16,6888 pada *threshold* Curah Hujan ≤ 79.

e) *Information Gain*

Tahap berikutnya adalah menilai kualitas pemisahan (*split*) dengan menghitung nilai *gain* untuk setiap kandidat *threshold* pada *subset bootstrap* ke-1. Nilai *gain* diperoleh dari selisih antara MSE sebelum dilakukan pemisahan dan MSE setelah pemisahan. Berikut perhitungan *information gain* dilakukan pada *threshold* Curah Hujan ≤ 79 dengan mengacu pada persamaan (7).

$$Gain^{(1.1.79)} = MSE^{(1)} - MSE_{split}^{(1.1.79)} = 34{,}3052 - 16{,}6888 = 17{,}6164$$

Berdasarkan hasil perhitungan, nilai *information gain* pada *threshold* Curah Hujan ≤ 79 diperoleh dari selisih antara MSE sebelum pemisahan sebesar 34,3052 dan

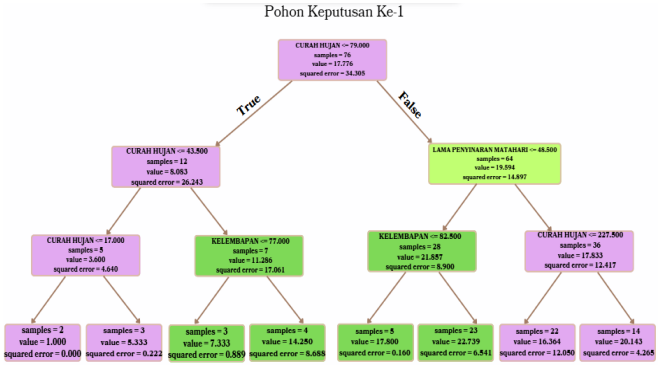
Copyright © 2026 – Indonesian Journal of Applied Mathematics

Published by: Lembaga Penelitian, Pengabdian Masyarakat, dan Penjaminan Mutu Institut Teknologi Sumatera, Lampung Selatan, Indonesia

Original Article

MSE setelah pemisahan sebesar 16,6888, sehingga menghasilkan nilai *gain* sebesar 17,6164.

Pada tahap yang telah dilakukan sebelumnya, yaitu pada langkah a hingga langkah e, proses yang dilakukan bertujuan untuk membentuk *node* akar. Setelah itu, dilanjutkan dengan pembentukan *node* cabang serta *node* daun. Berdasarkan tahapan pembentukan node tersebut, struktur pohon keputusan yang dihasilkan dari *subset bootstrap* pertama dapat ditampilkan secara keseluruhan pada **Gambar 1** di bawah ini.



Gambar 1. Pohon keputusan pertama

Gambar 1 menampilkan pohon keputusan ke-1 yang dibangun menggunakan parameter *max_depth* = 3, *min_samples_split* = 5, dan *min_samples_leaf* = 1. Proses pembentukan pohon dimulai dari *node* akar dengan variabel Curah Hujan sebagai pemisah utama pada *threshold* ≤ 79 mm karena menghasilkan *information gain* terbesar. Pada cabang kiri (Curah Hujan ≤ 79 mm), pemisahan lanjutan dilakukan menggunakan variabel Curah Hujan dan Kelembapan Udara sehingga menghasilkan beberapa *node* akhir dengan jumlah sampel yang relatif sedikit dan variasi yang lebih rendah. Sementara itu, pada cabang kanan (Curah Hujan > 79 mm), pemisahan selanjutnya ditentukan oleh variabel Lama Penyinaran Matahari, kemudian dilanjutkan dengan variabel Kelembapan Udara dan Curah Hujan pada tingkat berikutnya. Proses pemisahan berhenti pada kedalaman pohon ketiga sesuai dengan nilai parameter *max_depth* = 3 dan syarat minimal jumlah data untuk dilakukan pemisahan, yaitu *min_samples_split* = 5.

3. Pembentukan Model pohon keputusan pertama

Secara matematis, model pohon keputusan pertama dapat direpresentasikan dalam bentuk fungsi *piecewise* berdasarkan aturan pemisahan pada setiap node di Gambar 1. Berikut model pohon keputusan pertama pada subset *bootstrap* ke-1 berdasarkan Persamaan (8).

$$\hat{y}^{(1)} = \begin{cases} 1 & jika\ x_1 \le 79 \wedge x_1 \le 43{,}5 \wedge x_1 \le 17 \\ 5{,}333 & jika\ x_1 \le 79 \wedge x_1 \le 43{,}5 \wedge x_1 > 17 \\ 7{,}333 & jika\ x_1 \le 79 \wedge x_1 > 43{,}5 \wedge x_2 \le 77 \\ 14{,}250 & jika\ x_1 \le 79 \wedge x_1 > 43{,}5 \wedge x_2 > 77 \\ 17{,}8 & jika\ x_1 > 79 \wedge x_3 \le 48{,}5 \wedge x_2 \le 82{,}5 \\ 22{,}739 & jika\ x_1 > 79 \wedge x_3 \le 48{,}5 \wedge x_2 > 82{,}5 \\ 16{,}364 & jika\ x_1 > 79 \wedge x_3 > 48{,}5 \wedge x_1 \le 227{,}5 \\ 20{,}143 & jika\ x_1 > 79 \wedge x_3 > 48{,}5 \wedge x_1 > 227{,}5 \end{cases}$$

Berdasarkan fungsi *piecewise* di atas, model *decision tree* pohon keputusan pertama $\hat{y}^{(1)}$ pada kelompok data dengan intensitas curah hujan rendah ($x_1 \le 79$ mm), prediksi ditentukan sepenuhnya oleh fluktuasi curah hujan dan kelembapan. Jika curah hujan sangat rendah ($x_1 \le 17$ mm), model menghasilkan nilai konstan 1 hari, namun nilai tersebut meningkat menjadi 5,333 atau 6 hari, ketika curah hujan berada di rentang $17 < x_1 \le 43{,}5$. Untuk kondisi curah hujan antara > 43,5 mm hingga 79 mm, variabel kelembapan (x_2) menjadi penentu, di mana nilai prediksi adalah 7,333 atau 8 hari untuk $x_2 \le 77\%$ dan melonjak menjadi 14,250 atau 13 hari saat kelembapan melebihi ambang tersebut.

4. *Aggregating* data *training*

Prediksi akhir pada *Random Forest Regression* diperoleh dengan menghitung rata-rata hasil prediksi dari seluruh pohon keputusan yang dibangun menggunakan subset *bootstrap* yang berbeda. Pada sampel pertama data *training* dengan curah hujan 231 mm, kelembapan udara 85%, lama penyinaran matahari 55%, dan suhu udara 27,6°C, pohon keputusan pertama menghasilkan prediksi sebesar 20,143 atau 21 hari hujan berdasarkan aturan pemisahan pada *node* yang melibatkan variabel curah hujan dan lama penyinaran matahari. Nilai tersebut kemudian digabungkan dengan prediksi dari 499 pohon lainnya untuk menghasilkan prediksi akhir model. Berikut prediki sampel pertama dari 500 pohon keputusan.

Tabel 5.Prediksi Hari Hujan Data *Training* Sampel pertama

Model ke-	Prediksi Banyak Hari Hujan Sampel Ke-1 data *training*

1	20,143
2	18,583
3	17,737
4	19,641
5	19,556
6	18,467
7	17,885
8	18,615
9	17,891
10	17,333
⋮	⋮
499	18,885
500	19,5

Setelah seluruh nilai prediksi banyak hari hujan diperoleh, langkah selanjutnya menghitung rata-rata prediksi untuk sampel ke-1 pada 500 pohon keputusan menggunakan persamaan (9).

$$\hat{y} = \frac{1}{B}\sum_{i=1}^{B} \hat{y}^{(b)}$$

$$= \frac{1}{500}(20{,}143 + 18{,}583 + 17{,}737 + 19{,}641 + 19{,}556 + 18{,}467 + 17{,}885 + 18{,}615 + 17{,}891 + 17{,}333 + \cdots + 18{,}885 + 19{,}5)$$

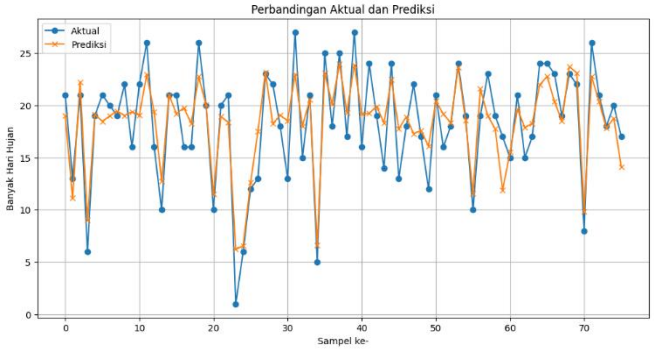
$$= 19{,}002.$$

Setelah dilakukan proses *agregasi* dengan menghitung rata-rata dari seluruh prediksi pohon, diperoleh nilai akhir untuk sampel pertama sebesar 19,002. Ringkasan hasil akhir dari proses prediksi seluruh sampel disajikan pada **Tabel 6** di bawah.

Tabel 6. Prediksi Data Training

CH	KU	LPM	SU	HH Aktual	HH Prediksi
231	85	55	27,6	21	19,002
63	81	69	27,6	13	11,120
386	87	41	27,5	21	22,209
77	74	71	27	6	9,028
258	83	61	27,7	19	19,129
138	83	56	27	21	18,467
239	82	52	27,3	20	18,972
237	84	50	27,2	19	19,448
236	83	52	27,7	22	19,022
244	85	63	26,8	16	19,381
220	85	49	27,9	22	19,049
184	85	24	27,2	26	22,936
227	82	35	26,5	16	19,372
87	84	61	27,5	10	12,741
343	82	51	27	21	20,993
199	83	47	26,9	21	19,189
324	81	45	27,5	16	19,758
⋮	⋮	⋮	⋮	⋮	⋮
81	80	49	26,6	17	14,095

Tabel 6 menyajikan hasil prediksi akhir yang diperoleh dari model *Random Forest Regression* pada data *training*. Selanjutnya, hasil tersebut divisualisasikan dalam bentuk grafik untuk menunjukkan perbandingan antara nilai aktual dan nilai prediksi sebagai berikut.



Gambar 2. Grafik Perbandingan Prediksi Data *Training*

Grafik pada **Gambar 2** di atas menggambarkan hubungan antara nilai aktual dan hasil prediksi banyak hari hujan untuk setiap sampel pada data *training*. Garis biru merepresentasikan data aktual banyak hari hujan, sedangkan garis oranye menggambarkan nilai prediksi banyak hari hujan. Secara umum, pola garis prediksi terlihat mengikuti pola nilai aktual, yang menunjukkan bahwa model mampu menangkap pola data.

5. Evaluasi Hasil Prediksi Data *Training*

Evaluasi model pada data *training* menggunakan perhitungan nilai RMSE dan R^2 menggunakan persamaan (11) dan (12) sebagai berikut.

1. *Root Mean Squared Error* (RMSE)

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

$$= \sqrt{\frac{1}{76}(3.992 + 3.534 + 1.462 + 9.169 + \cdots + 8.439}$$

Copyright © 2026 – Indonesian Journal of Applied Mathematics

Published by: Lembaga Penelitian, Pengabdian Masyarakat, dan Penjaminan Mutu Institut Teknologi Sumatera, Lampung Selatan, Indonesia

Original Article

$= 2.581$

2. Koefisien Determinasi

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$= 1 - \frac{506.207}{2201.197}$$

$$= 0.77$$

Nilai *Root Mean Squared Error* (RMSE) sebesar 2,581 menunjukkan bahwa rata-rata kesalahan prediksi model sebanyak 3 hari hujan. Sementara itu, nilai koefisien determinasi (R^2) sebesar 0,77 mengindikasikan bahwa 77% variasi pada data aktual mampu dijelaskan oleh model *Random Forest Regression* dengan variabel-variabel yang ada.

6. *Aggegating* Data *Testing*

Setelah model dilatih menggunakan data *training*, evaluasi dilakukan pada data *testing* yang tidak terlibat dalam proses *training*. Pada *Random Forest Regression*, setiap pohon keputusan menghasilkan prediksi terhadap sampel data *testing*, kemudian seluruh prediksi tersebut dirata-ratakan untuk memperoleh prediksi akhir. Sebagai contoh, pada sampel pertama data *testing*, pohon keputusan pertama menghasilkan prediksi sebesar 16,364 hari hujan berdasarkan jalur keputusan yang dibentuk oleh variabel curah hujan dan lama penyinaran matahari. Hasil lengkap prediksi seluruh pohon untuk sampel tersebut disajikan pada **Tabel 7.**

Tabel 7. Prediksi Banyak Hari Hujan Sampel ke-1 data *Testing*

Model ke-	Prediksi Banyak Hari Hujan Sampel Ke-1 data *testing*
1	16,364
2	18,583
3	17,737
4	16,385
5	16,250
6	18,467
7	17,885
8	18,615
9	17,891
10	17,333
⋮	⋮
499	18,885
500	19,500

Tabel 7 di atas merupakan hasil prediksi banyak hari hujan dari 500 pohon keputusan terhadap sampel data *testing* ke-1. Setelah seluruh nilai prediksi banyak hari hujan diperoleh, langkah selanjutnya menghitung rata-rata prediksi untuk sampel ke-1 pada 500 pohon keputusan menggunakan persamaan (10).

$$\hat{y}_{test} = \frac{1}{B_{trees}} \sum_{i=1}^{B_{trees}} \hat{y}^{(b)}$$

$$= \frac{1}{500}(16{,}364 + 18{,}583 + 17{,}737 + 16{,}385 + 16{,}250 + 18{,}467 + 17{,}885 + 18{,}615 + 17{,}891 + 17{,}333 + \cdots + 18{,}885 + 19{,}5)$$

$$= 17{,}584$$

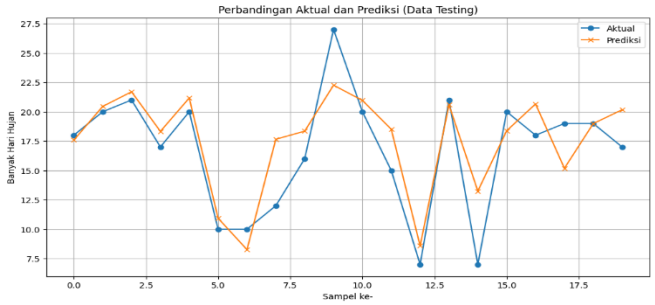
Perhitungan tersebut menggambarkan tahapan dalam menentukan nilai prediksi untuk sampel pertama pada data *testing*, dengan hasil prediksi banyak hari hujan sebesar 17,584 atau 18 hari. Prediksi untuk seluruh sampel data *testing* disajikan secara lengkap pada **Tabel 8** di bawah ini.

Tabel 8. Prediksi Data *Testing*

CH	KU	LPM	SU	HH Aktual	HH Prediksi
194	80	61	27,8	18	17,584
179	85	28	27,4	20	20,451
530	87	79	27	21	21,697
191	81	35	27,6	17	18,327
368	82	50	27,2	20	21,170
56	78	66	28,1	10	10,932
43	78	63	28,3	10	8,265
186	82	67	27,4	12	17,658
135	81	47	27	17	18,344
421	86	56	27	27	22,260
320	86	49	27,7	20	20,963
198	84	64	27,7	15	18,508
21	79	62	27,3	7	8,582
290	86	59	27,3	21	20,590
96	79	79	27,6	7	13,247
232	81	61	28,1	20	18,382
167	84	33	26	18	20,663
95	82	37	27,3	19	15,187
148	83	47	26,6	19	18,992
254	85	79	26,7	17	20,167

Tabel 8 di atas, menampilkan hasil prediksi pada data *testing* yang diperoleh dengan menerapkan model

Random Forest Regression. Grafik yang menggambarkan perbandingan tersebut dapat dilihat pada **Gambar 3** di bawah ini.



Gambar 3. Grafik Perbandingan Prediksi Data *Testing*

Secara umum, pola garis prediksi pada **Gambar 3** di atas mengikuti tren garis aktual, yang menunjukkan bahwa model cukup mampu menangkap hubungan dan pola data yang ada.

7. Evaluasi Hasil Prediksi Data *Testing*

Evaluasi model pada data *testing* menggunakan perhitungan nilai RMSE dan R^2 menggunakan persamaan (11) dan (12) sebagai berikut.

1. *Root Mean Squared Error* (RMSE)

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

$$= \sqrt{\frac{1}{20}(0{,}173 + 0{,}203 + 0{,}486 + 1{,}761 + \cdots + 10{,}030}$$

$$= 2{,}769$$

2. Koefisien Determinasi

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$= 1 - \frac{153{,}336}{503{,}75}$$

$$= 0{,}695$$

Nilai RMSE sebesar 2,769 menunjukkan bahwa rata-rata kesalahan prediksi model berada di kisaran 3 hari hujan dari nilai aktual. Sementara itu, nilai koefisien determinasi (R^2) sebesar 0,695 mengindikasikan bahwa 69,5% variasi banyak hari hujan dapat dijelaskan oleh model. Meskipun model menunjukkan performa yang cukup baik, terdapat keterbatasan terkait ukuran sampel yang digunakan. Penggunaan data sekunder bulanan sebanyak 96 pengamatan tergolong terbatas untuk pemodelan *ensemble* berbasis 500 pohon keputusan, sehingga berisiko kurang representatif dalam mengekstraksi variabilitas iklim jangka panjang. Selain itu, evaluasi model yang mengandalkan pembagian data tunggal memiliki kerentanan terhadap variansi pengambilan sampel (sampling bias) pada data uji [13]. Untuk penelitian selanjutnya, disarankan menggunakan evaluasi *K-Fold Cross Validation* secara menyeluruh pada pengujian akhir serta menambah rentang data historis agar estimasi generalisasi model menjadi lebih stabil dan objektif.

Feature Important

Feature Importance digunakan untuk mengukur kontribusi setiap variabel independen dalam menghasilkan prediksi model berdasarkan penurunan nilai *Mean Squared Error* (MSE) pada proses split di pohon keputusan. Nilai *Feature Importance* diperoleh dari akumulasi kontribusi seluruh pohon dalam model. Sebagai ilustrasi, pada *node* akar pohon pertama, variabel curah hujan dipilih sebagai pemisah karena memberikan penurunan MSE terbesar, sehingga memiliki kontribusi dalam meningkatkan kinerja prediksi model.

$$FI(1) = \frac{n_{t,j}}{n_b}\left(MSE_{before} - MSE_{split}\right)$$

$$= \frac{76}{76}(34{,}3052 - 16{,}688)$$

$$= 17{,}617$$

Hasil perhitungan tersebut menunjukkan bahwa variabel curah hujan memiliki kontribusi sebesar 17,617 dalam menurunkan nilai *galat* pada *node* akar. Adapun kontribusi masing-masing variabel *independen* pada seluruh pohon keputusan model *Random Forest Regression* disajikan secara ringkas pada **Tabel 9,** yang menunjukkan peran relatif setiap variabel dalam proses prediksi model secara keseluruhan.

Tabel 9. *Feature Importance*

Variabel	Importance Total	Persentase
CH	8725,1072	76,97%
KU	1276,3021	7,28%
LPM	825,474	11,26%
SU	507,717	4,48%

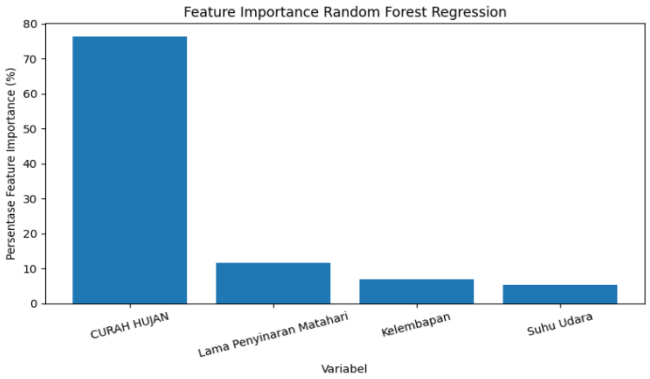
Copyright © 2026 – Indonesian Journal of Applied Mathematics

Published by: Lembaga Penelitian, Pengabdian Masyarakat, dan Penjaminan Mutu Institut Teknologi Sumatera, Lampung Selatan, Indonesia

Original Article

Tabel 9 di atas, hasil *Feature Importance* menunjukkan bahwa Curah Hujan merupakan variabel paling dominan dengan kontribusi sebesar 76,97%, Lama Penyinaran Matahari sebesar 11,26%, di ikuti Kelembapan Udara memberikan kontribusi sebesar 7,28%, dan Suhu Udara sebesar 4,48% yang berperan sebagai faktor pendukung dengan pengaruh yang relatif lebih kecil. Secara konseptual, curah hujan bulanan (mm) dan banyak hari hujan (hari) memiliki keterikatan fisis dan dependen yang erat secara definisi. Hari hujan didefinisikan sebagai hari di mana terukur akumulasi presipitasi melebihi ambang batas tertentu (misalnya ≥1,0 mm). Oleh karena itu, jumlah akumulasi hujan bulanan secara kuantitatif berkorelasi langsung dengan frekuensi kejadian hujan pada bulan tersebut. Secara empiris, bulan dengan total curah hujan yang tinggi hampir selalu diiringi oleh frekuensi hari hujan yang lebih banyak, meskipun hubungan ini tidak bersifat linier sempurna karena adanya pengaruh intensitas hujan (misalnya hujan lebat berdurasi singkat dan hujan ringan yang berlangsung terus-menerus). Perbedaan tingkat kontribusi antarvariabel tersebut dapat diamati dengan lebih jelas melalui visualisasi dalam bentuk diagram batang. Diagram ini memberikan gambaran yang lebih intuitif mengenai besarnya pengaruh relatif masing-masing variabel terhadap model prediksi yang dibangun. Adapun diagram batang yang menampilkan hasil persentase *Feature Importance* pada **Gambar 4**.

Grafik *Feature Importance* di atas menunjukkan bahwa variabel Curah Hujan memiliki pengaruh paling besar dalam model *Random Forest*, dengan nilai *importance* yang jauh lebih tinggi dibandingkan variabel lainnya. Hal ini menandakan bahwa prediksi banyak hari hujan sangat didominasi oleh informasi dari curah hujan.



Gambar 4. Diagram Batang *Feature Important*

Sementara itu, Lama Penyinaran Matahari berada pada posisi kedua dengan kontribusi yang lebih kecil, diikuti oleh Kelembapan dan Suhu Udara yang memiliki pengaruh relatif rendah terhadap model. Ketiga variabel ini hanya berperan sebagai pendukung dalam meningkatkan akurasi prediksi [14].

Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, algoritma *Random Forest Regression* menghasilkan nilai RMSE sebesar 2,769 serta nilai koefisien determinasi (R^2) sebesar 0,695 pada data *testing* dan nilai RMSE sebesar 2,581 serta nilai koefisien determinasi (R^2) sebesar 0,77 pada data *training*, yang menunjukkan bahwa model mampu menangkap sebagian besar variasi dalam data serta menghasilkan prediksi dengan tingkat kesalahan yang rendah. Keakuratan prediksi tersebut menjadi aspek penting dalam mendukung perencanaan dan pengambilan keputusan, khususnya dalam pengelolaan sumber daya air, serta upaya mitigasi bencana *hidrometeorologi* di wilayah Kota Jambi.

Berdasarkan hasil penelitian, faktor iklim yang paling berperan dalam memengaruhi banyaknya hari hujan di Kota Jambi ditentukan melalui nilai *feature importance* dari model *Random Forest Regression*. Hasil yang diperoleh menunjukkan bahwa curah hujan menjadi variabel yang paling berpengaruh sebesar 76,97%, kemudian diikuti oleh LPM sebesar 11,26%, kelembapan udara sebesar 7,28% dan suhu udara sebesar 4,48%. Nilai-nilai tersebut mengindikasikan bahwa curah hujan memiliki kontribusi paling dominan

dalam menjelaskan variasi jumlah hari hujan, sedangkan variabel lainnya tetap memberikan pengaruh meskipun tidak sebesar curah hujan.

Saran

Perlu mengintegrasikan aspek spasial, seperti data lokasi geografis atau peta wilayah, sehingga hasil prediksi tidak hanya bersifat numerik tetapi juga dapat menggambarkan distribusi hujan secara wilayah.

Penelitian selanjutnya disarankan untuk menerapkan metode *Time Series Split* (pembagian berdasarkan urutan waktu historis) atau *Block Cross-Validation* untuk menguji andal tidaknya model dalam skenario peramalan (*forecasting*) masa depan secara temporal.

Konflik Kepentingan

Tidak ada konflik kepentingan yang dinyatakan.

Referensi

[1] E. Aldrian and R. Dwi Susanto, "Identification of three dominant rainfall regions within Indonesia and their relationship to sea surface temperature," *Int. J. Climatol. A J. R. Meteorol. Soc.*, vol. 23, no. 12, pp. 1435–1452, 2003.

[2] E. Surmaini and A. Faqih, "Kejadian iklim ekstrem dan dampaknya terhadap pertanian tanaman pangan di Indonesia," *J. Sumberd. Lahan*, vol. 10, no. 2, 2016.

[3] E. R. Eko, H. Syarifuddin, and J. Jalius, "Analisis kualitas air sungai batanghari berkelanjutan di kota jambi," *J. Pembang. Berkelanjutan*, vol. 1, no. 2, pp. 123–141, 2018.

[4] D. Sijabat, "Kaleidoskop 2025: Daftar Kota Jambi Beberapa Kali Diterjang Banjir, Ribuan Mengungsi, Jatuh Korban," TribunJambi. [Online]. Available: https://jambi.tribunnews.com/news/1184569/kaleidoskop-2025-daftar-kota-jambi-beberapa-kali-diterjang-banjir-ribuan-mengungsi-jatuh-korban

[5] A. Indryani, U. Khaira, and M. F. Putri, "Rainfall Prediction Using Long Short-Term Memory Method (Case Study: Jambi City)," *J. Pepadun*, vol. 6, no. 1, pp. 57–70, 2025.

[6] L. E. O. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001, [Online]. Available: https://link.springer.com/article/10.1023/A:1010933404324

[7] M. Sulistiyono, A. Sidauruk, B. Satria, and R. Wardhana, "Rainfall Prediction Using Multiple Linear Regression Algorithm," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 9, no. 1, pp. 17–22, 2023.

[8] T. Hastie, R. Tibshirani, and J. Friedman, "The elements of statistical learning," 2009, *Springer series in statistics New-York*.

[9] G. A. RAKHMAT and W. MUTOHAR, "Prakiraan Hujan menggunakan Metode Random Forest dan Cross Validation," *MIND (Multimedia Artif. Intell. Netw. Database) J.*, vol. 8, no. 2, pp. 173–187, 2023.

[10] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*, 3rd ed. " O'Reilly Media, Inc.," 2022.

[11] L. Rokach and O. Maimon, "Decision trees," in *Data mining and knowledge discovery handbook*, Springer, 2005, pp. 165–192.

[12] D. Arya, L. Rochmawati, and I. Sonhaji, "Koefisien korelasi (r) dan koefisien determinasi (r2)," *J. Penelit.*, vol. 5, no. 4, pp. 289–296, 2020.

[13] H. Bichri, A. Chergui, and M. Hain, "Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets," vol. 15, no. 2, pp. 331–339, 2024.

[14] J. D. Mendiola-Santibañez *et al.*, "Artificial Intelligence Algorithm Based on Genetics to Predict Responses to Interferon-Beta Treatment in Multiple Sclerosis Patients.," *Bioeng. (Basel, Switzerland)*, vol. 13, no. 5, p. 523, 2026.

Copyright © 2026 – Indonesian Journal of Applied Mathematics

Published by: Lembaga Penelitian, Pengabdian Masyarakat, dan Penjaminan Mutu Institut Teknologi Sumatera, Lampung Selatan, Indonesia