

PREDIKSI DAYA FOTOVOLTAIK BERDASARKAN SUHU DAN RADIASI MATAHARI MENGGUNAKAN RANDOM FOREST REGRESSION

¹Chamdalah, ²Istiyo Winarno, ³Muhammad Taufiqurrohman

^{1,2,3} Program Studi Teknik Elektro, Fakultas Teknik dan Ilmu Kelautan, Universitas Hang Tuah, Surabaya

¹hamdalah.pudimin@gmail.com, ²istiyo.winarno@hangtuah.ac.id, ³taufiqurrohman@hangtuah.ac.id

Abstract—The development of photovoltaic (PV) systems as a renewable energy source has increased the need for accurate power-prediction methods, since PV power output is strongly affected by variations in ambient temperature and solar irradiance. This study aims to develop a prediction model for photovoltaic module power output using the Random Forest Regression (RFR) algorithm. The research data were obtained from the Photovoltaic Geographical Information System (PVGIS) for a research site in Surabaya, Indonesia, covering the year 2023. The input variables used were ambient temperature and solar irradiance, while the output variable was PV module power. The research stages included data preprocessing, splitting the dataset into 80 % training data and 20% testing data, hyperparameter optimization using GridSearchCV, and model evaluation using RMSE and R^2 . The results show an RMSE value of 29.573 W and an R^2 value of 0.940, with an average cross-validation score of 0.917. These results indicate that the Random Forest Regression model is able to provide accurate photovoltaic power predictions and possesses good generalization capability under Indonesia's tropical climate conditions.

Keywords—*photovoltaic; power prediction; temperature; solar irradiance; Random Forest Regression; PVGIS.*

Abstrak—Perkembangan sistem fotovoltaik (PV) sebagai sumber energi terbarukan mendorong perlunya metode prediksi daya yang akurat karena keluaran daya dipengaruhi oleh perubahan suhu dan intensitas radiasi matahari. Penelitian ini bertujuan membangun model prediksi output daya modul fotovoltaik menggunakan algoritma Random Forest Regression (RFR). Data penelitian diperoleh dari Photovoltaic Geographical Information System (PVGIS) berdasarkan lokasi penelitian di Surabaya selama tahun 2023. Variabel input yang digunakan adalah suhu dan intensitas radiasi matahari, sedangkan variabel output berupa daya modul fotovoltaik. Tahapan penelitian meliputi preprocessing, pembagian data menjadi 80% data pelatihan dan 20% data pengujian, optimasi parameter menggunakan GridSearchCV, serta evaluasi model menggunakan RMSE dan R^2 . Hasil penelitian menunjukkan nilai RMSE sebesar 29,573 W dan R^2 sebesar 0,940, dengan nilai rata-rata cross validation sebesar 0,917. Hasil tersebut menunjukkan bahwa model Random Forest Regression mampu memberikan prediksi daya modul fotovoltaik yang akurat dan memiliki kemampuan generalisasi yang baik pada kondisi iklim tropis Indonesia.

Kata Kunci—*fotovoltaik; prediksi daya; suhu; intensitas radiasi matahari; Random Forest Regression; PVGIS.*

PENDAHULUAN

Kebutuhan energi listrik terus meningkat seiring pertumbuhan penduduk, perkembangan teknologi, serta meningkatnya aktivitas industri, komersial, dan rumah tangga, sementara ketergantungan terhadap energi fosil menimbulkan permasalahan keterbatasan cadangan energi dan peningkatan emisi gas rumah kaca [1], [2]. Kondisi tersebut mendorong pengembangan sumber energi terbarukan yang lebih ramah lingkungan, salah satunya energi surya melalui teknologi fotovoltaik (PV) yang mengubah radiasi matahari menjadi energi listrik tanpa emisi operasional [3].

Salah satu tantangan utama pemanfaatan sistem PV adalah keluaran daya yang berfluktuasi karena sangat dipengaruhi kondisi lingkungan. Berbagai penelitian menunjukkan bahwa intensitas radiasi matahari dan temperatur merupakan dua parameter yang paling dominan memengaruhi daya keluaran modul PV [3], [4]. Peningkatan radiasi matahari meningkatkan energi yang diterima modul, sedangkan peningkatan temperatur modul cenderung menurunkan tegangan keluaran dan efisiensi konversi energi [5], [6]. Oleh karena itu, diperlukan metode yang mampu memprediksi keluaran daya PV secara akurat dan andal untuk mendukung pengelolaan sistem energi, khususnya pada sistem yang terhubung dengan jaringan listrik.

Perkembangan teknologi komputasi mendorong penggunaan machine learning dalam prediksi daya fotovoltaik karena mampu mempelajari hubungan non-linear yang kompleks antara parameter meteorologi dan daya keluaran berdasarkan data historis, dengan akurasi yang umumnya lebih baik dibandingkan metode statistik konvensional [1], [4]. Salah satu algoritma yang banyak digunakan adalah Random Forest Regression (RFR), yaitu metode ensemble learning berbasis sejumlah decision tree yang mampu menangani hubungan non-linear dan mengurangi risiko overfitting. Ahmad et al. menunjukkan bahwa Random Forest menghasilkan performa lebih baik dibandingkan Support Vector Regression pada beberapa kasus prediksi daya PV [7]. Meng dan Song membuktikan Random Forest mampu memprediksi daya PV harian dengan tingkat kesalahan relatif rendah [8], sedangkan Massaoudi et al. mengembangkan model Enhanced Random Forest yang meningkatkan akurasi prediksi daya PV jangka pendek menggunakan parameter cuaca sebagai masukan [9].

Sebagian besar penelitian terdahulu menggunakan banyak variabel meteorologi seperti kelembapan udara, kecepatan angin, tekanan atmosfer, dan tutupan awan, yang meningkatkan kompleksitas model, kebutuhan komputasi, serta kesulitan memperoleh data secara konsisten. Padahal, beberapa penelitian menunjukkan bahwa intensitas radiasi matahari dan temperatur merupakan dua variabel dengan korelasi paling kuat terhadap daya keluaran PV [3], [4], [7]. Oleh karena itu, penelitian ini membangun model prediksi yang lebih sederhana dengan hanya menggunakan dua parameter utama tersebut, menggunakan data historis dari Photovoltaic Geographical Information System (PVGIS) yang dikembangkan oleh Joint Research Centre European Commission [10] untuk lokasi Surabaya, Jawa Timur.

Berdasarkan uraian tersebut, penelitian ini berjudul “Prediksi Output Daya Modul Fotovoltaik Berdasarkan Data Suhu dan Intensitas Radiasi Matahari Menggunakan Algoritma Random Forest Regression (RFR)”. Kebaruan penelitian terletak pada penggunaan dua variabel meteorologi utama untuk membangun model prediksi daya PV pada kondisi iklim tropis Indonesia, dengan tujuan (1) membangun model prediksi daya keluaran modul fotovoltaik berdasarkan suhu dan intensitas radiasi matahari, dan (2) mengevaluasi kinerja model menggunakan Root Mean Square Error (RMSE) dan Coefficient of Determination (R^2).

METODE PENELITIAN

Data Penelitian

Data penelitian merupakan data historis yang diperoleh dari Photovoltaic Geographical Information System (PVGIS) yang dikembangkan oleh Joint Research Centre European Commission, berupa platform data satelit dan model meteorologi yang bersifat open access [10]. Data diperoleh dengan memasukkan koordinat lokasi penelitian pada wilayah Kota Surabaya sebagaimana ditunjukkan pada Tabel 1.

Tabel 1. Sumber Data PVGIS

Parameter	Nilai
Latitude	-7,246°
Longitude	112,738°
Lokasi	Surabaya, Jawa Timur, Indonesia
Periode data	Januari 2023 – Desember 2023
Interval data	Per jam (hourly data)

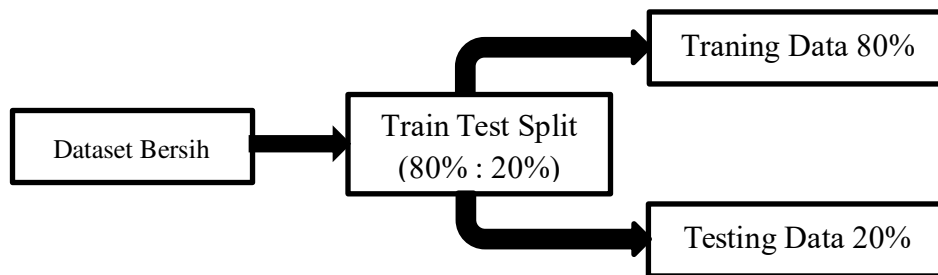
Data yang digunakan terdiri atas variabel masukan (input) berupa temperatur lingkungan dan intensitas radiasi matahari, serta variabel keluaran (output) berupa daya modul fotovoltaik, sebagaimana ditunjukkan pada Tabel 2. Hubungan antar variabel dapat dituliskan sebagai $P = f(T, G)$, dengan P adalah daya keluaran modul PV (W), T adalah temperatur lingkungan ($^{\circ}\text{C}$), dan G adalah intensitas radiasi matahari (W/m^2).

Tabel 2. Variabel Penelitian

No	Variabel	Jenis	Satuan
1	Temperatur lingkungan	Input	°C
2	Intensitas radiasi matahari	Input	W/m ²
3	Daya keluaran modul PV	Output	W

Preprocessing dan Pembagian Data

Tahap preprocessing bertujuan meningkatkan kualitas dataset sebelum digunakan pada proses pelatihan model, meliputi penanganan nilai kosong (missing value), pembersihan data tidak valid atau duplikat, konversi tipe data ke format numerik, serta penyesuaian format waktu dari UTC ke WIB. Setelah proses pembersihan selesai, dataset dipisahkan menjadi data pelatihan (training data) dan data pengujian (testing data) dengan proporsi 80:20 sebagaimana ditunjukkan pada Gambar 1, yang merupakan pendekatan umum dalam pengembangan model prediksi berbasis machine learning [8].



Gambar 1. Ilustrasi Pembagian Data Pelatihan dan Pengujian

Pembangunan Model Random Forest Regression

Random Forest Regression merupakan metode ensemble learning yang dibangun dari sejumlah decision tree yang dilatih menggunakan subset data berbeda melalui metode bootstrap aggregating (bagging). Hasil prediksi seluruh pohon keputusan kemudian digabungkan menggunakan metode rata-rata (averaging) untuk menghasilkan prediksi akhir yang lebih stabil dan akurat, sehingga mampu menangani hubungan non-linear antara variabel suhu, radiasi matahari, dan daya keluaran fotovoltaik [7], [9]. Untuk memperoleh performa model yang optimal, penelitian ini menggunakan metode GridSearchCV dengan 5-fold cross validation guna mencari kombinasi hyperparameter terbaik, dengan kandidat nilai sebagaimana ditunjukkan pada Tabel 3.

Tabel 3. Kandidat Parameter pada Proses Optimasi GridSearchCV

Parameter	Nilai yang Diuji
n_estimators	100, 200, 300
max_depth	10, 15, 20, None
min_samples_split	2, 5, 10
min_samples_leaf	1, 2, 4
max_features	sqrt, log2

Evaluasi Kinerja Model

Evaluasi kinerja model dilakukan dengan membandingkan hasil prediksi terhadap data aktual pada data pengujian menggunakan Root Mean Square Error (RMSE) dan Coefficient of Determination (R^2), sebagaimana Persamaan (1) dan (2), serta 5-fold cross validation untuk menguji kemampuan generalisasi model.

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

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

HASIL DAN PEMBAHASAN

Hasil Optimasi Parameter Menggunakan GridSearchCV

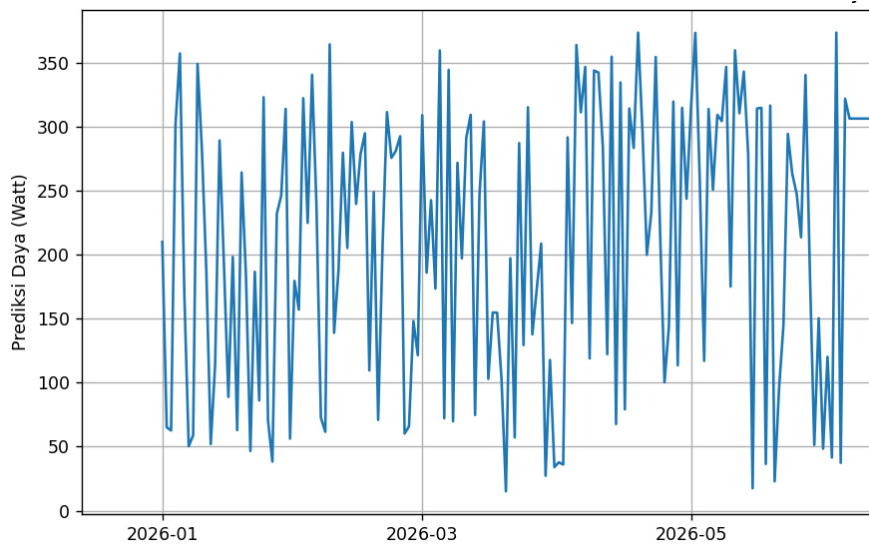
Berdasarkan hasil optimasi menggunakan GridSearchCV, diperoleh kombinasi parameter terbaik sebagaimana ditunjukkan pada Tabel 4. Parameter `n_estimators` menentukan jumlah decision tree yang dibangun; semakin banyak pohon, semakin stabil hasil prediksi meskipun waktu komputasi meningkat. Parameter `max_depth` mengatur kedalaman maksimum tiap pohon sehingga model dapat mempelajari pola kompleks tanpa risiko overfitting yang berlebihan, sedangkan `min_samples_split`, `min_samples_leaf`, dan `max_features` berperan menjaga stabilitas struktur pohon dan meningkatkan keberagaman antar pohon (tree diversity).

Tabel 4. Parameter Terbaik Random Forest Regression

Parameter	Nilai Terbaik
<code>n_estimators</code>	200
<code>max_depth</code>	20
<code>min_samples_split</code>	2
<code>min_samples_leaf</code>	1
<code>max_features</code>	sqrt

Prediksi Daya Modul Fotovoltaik Tahun 2026

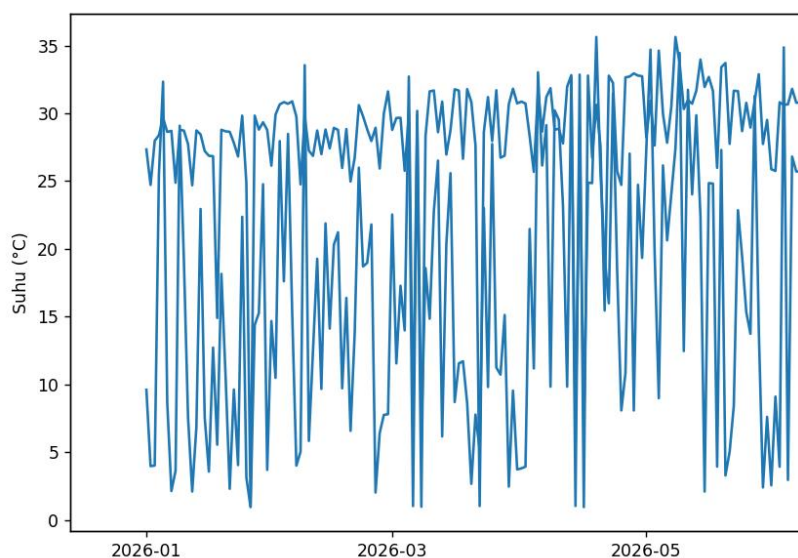
Model terbaik yang diperoleh digunakan untuk memprediksi daya keluaran modul fotovoltaik pada tahun 2026 berdasarkan data suhu dan intensitas radiasi matahari. Sebagaimana ditunjukkan pada Gambar 2, pada awal periode prediksi terlihat variasi daya yang cukup besar dengan nilai berkisar 20–370 W, mengikuti perubahan kombinasi temperatur dan radiasi. Pada pertengahan hingga akhir tahun, nilai prediksi menjadi relatif konstan di sekitar 306 W karena kombinasi variabel masukan pada periode tersebut cenderung sama. Kondisi ini merupakan karakteristik Random Forest Regression yang tidak melakukan ekstrapolasi seperti regresi linear, melainkan menghasilkan prediksi berdasarkan pola yang telah dipelajari pada data pelatihan.



Gambar 2. Prediksi Daya Modul Fotovoltaik Tahun 2026

Analisis Variasi Suhu dan Radiasi Matahari

Gambar 3 memperlihatkan variasi temperatur udara dan intensitas radiasi matahari yang digunakan sebagai variabel masukan model. Temperatur berada pada kisaran sekitar 26–31°C, sedangkan intensitas radiasi matahari berkisar 600–760 W/m² pada sebagian besar periode prediksi. Pada bagian awal data masih terdapat fluktuasi yang cukup tinggi sehingga menghasilkan variasi daya keluaran, namun setelah pertengahan tahun nilai kedua variabel menjadi relatif konstan sehingga prediksi daya juga relatif tetap. Kombinasi kedua variabel tersebut menjadi faktor utama yang menentukan besar kecilnya daya keluaran hasil prediksi [5], [6].

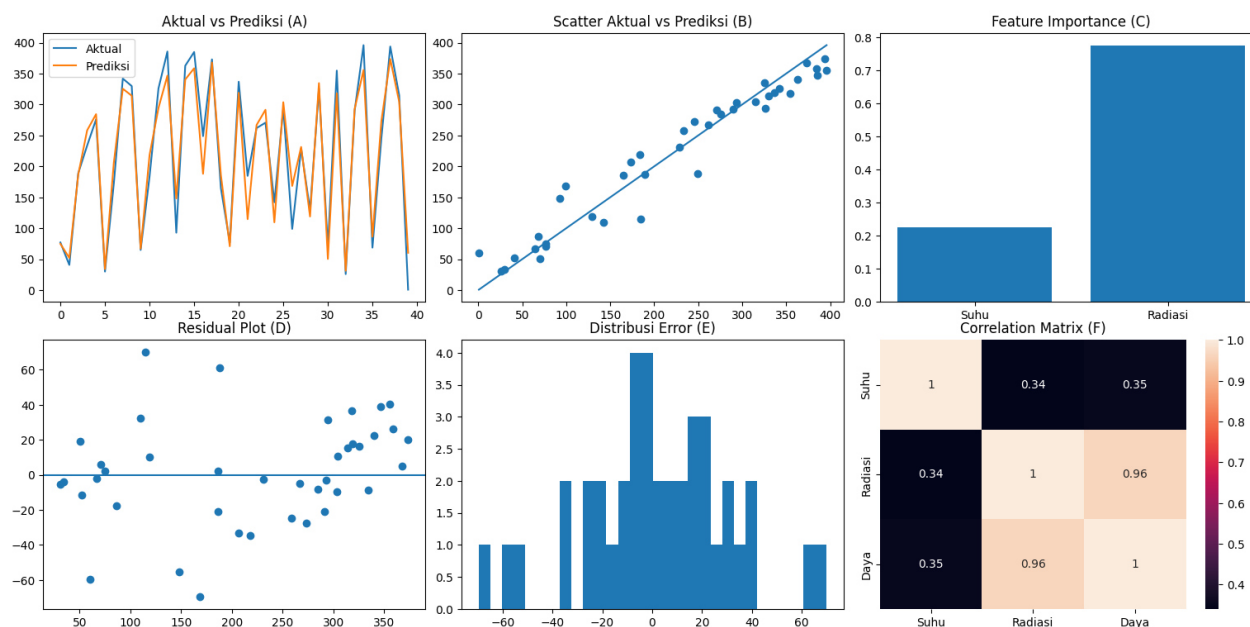


Gambar 3. Variasi Suhu dan Intensitas Radiasi Matahari

Analisis Hasil Prediksi Model

Analisis lebih lanjut dilakukan menggunakan beberapa bentuk visualisasi, yaitu grafik perbandingan data aktual dan hasil prediksi, scatter plot, feature importance, residual plot, dan distribusi error, sebagaimana ditunjukkan pada Gambar 4. Pola hasil prediksi mengikuti pola data aktual dengan cukup baik dan sebagian besar titik pada scatter plot

berada di sekitar garis diagonal ideal, menunjukkan tingkat kesesuaian yang tinggi terhadap data aktual. Residual tersebar secara acak di sekitar garis nol tanpa membentuk pola tertentu, mengindikasikan model tidak mengalami bias sistematis, dan sebagian besar error berada di sekitar nilai nol pada histogram distribusi error.



Gambar 4. (A) Perbandingan Data Aktual dan Prediksi, (B) Scatter Plot, (C) Feature Importance, (D) Residual Plot, (E) Distribusi Error, (F) Correlation Matrix

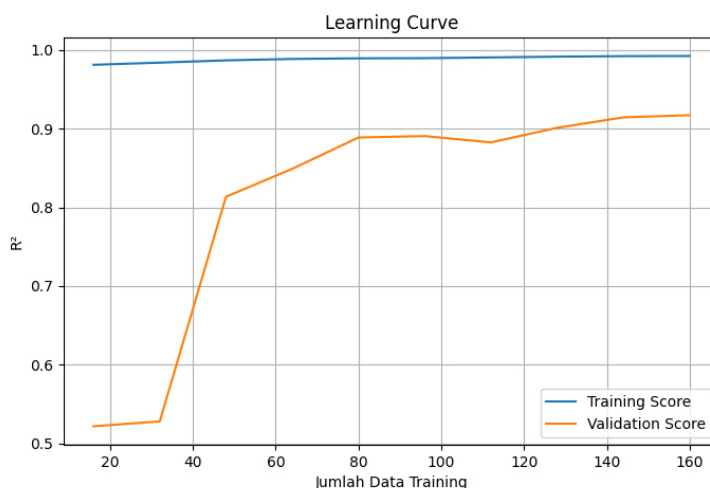
Analisis feature importance menunjukkan bahwa variabel intensitas radiasi matahari memiliki nilai kontribusi sekitar 0,78, sedangkan variabel temperatur sekitar 0,22, sehingga radiasi matahari merupakan variabel yang paling dominan memengaruhi daya keluaran modul PV. Hasil ini diperkuat oleh correlation matrix pada Tabel 5, yang menunjukkan korelasi sangat kuat antara radiasi dan daya (0,96), sedangkan korelasi temperatur terhadap daya jauh lebih rendah (0,35).

Tabel 5. Nilai Correlation Matrix Antar Variabel

Hubungan Variabel	Nilai Korelasi
Suhu – Radiasi	0,34
Suhu – Daya	0,35
Radiasi – Daya	0,96

Learning Curve dan Evaluasi Akhir Model

Learning curve pada Gambar 5 menunjukkan hubungan antara jumlah data pelatihan dengan nilai R^2 pada data pelatihan (training score) dan data validasi (validation score). Training score berada pada kisaran 0,98–0,99 dan cenderung stabil, sedangkan validation score meningkat signifikan dari sekitar 0,52 menjadi sekitar 0,91 seiring bertambahnya jumlah data pelatihan. Selisih akhir antara training score dan validation score relatif kecil (sekitar 0,07–0,08), menunjukkan bahwa model tidak mengalami overfitting yang berlebihan dan memiliki kemampuan generalisasi yang baik.



Gambar 5. Learning Curve Model Random Forest Regression

Hasil evaluasi akhir model pada data pengujian ditunjukkan pada Tabel 6. Nilai RMSE sebesar 29,573 W menunjukkan rata-rata kesalahan prediksi yang relatif kecil dibandingkan rentang daya pada dataset, sedangkan nilai R^2 sebesar 0,940 menunjukkan bahwa model mampu menjelaskan sekitar 94% variasi daya keluaran modul fotovoltaik berdasarkan dua variabel masukan yang digunakan. Nilai rata-rata cross validation R^2 sebesar 0,917 memperkuat bahwa model memiliki kemampuan generalisasi yang baik terhadap data yang belum pernah digunakan pada proses pelatihan.

Tabel 6. Hasil Evaluasi Kinerja Model

Parameter Evaluasi	Nilai
RMSE	29,573 W
R^2	0,940
Rata-rata Cross Validation R^2	0,917

KESIMPULAN

Penelitian ini berhasil membangun model prediksi daya keluaran modul fotovoltaik menggunakan algoritma Random Forest Regression (RFR) dengan memanfaatkan dua variabel masukan utama, yaitu suhu dan intensitas radiasi matahari, melalui tahapan pengumpulan data dari PVGIS, preprocessing, pembagian dataset, pelatihan model, optimasi parameter menggunakan GridSearchCV, serta evaluasi hasil prediksi.

Hasil evaluasi menunjukkan bahwa model Random Forest Regression mampu menghasilkan performa prediksi yang baik terhadap daya keluaran modul fotovoltaik, dengan nilai RMSE sebesar 29,573 W, R^2 sebesar 0,940, serta rata-rata cross validation R^2 sebesar 0,917. Nilai R^2 yang mendekati 1 menunjukkan bahwa model mampu menjelaskan sekitar 94% variasi data daya keluaran modul fotovoltaik, sedangkan nilai cross validation menunjukkan bahwa model memiliki kemampuan generalisasi yang baik pada kondisi iklim tropis Indonesia.

DAFTAR PUSTAKA

[1] A. Alcañiz, D. Grzebyk, H. Ziar, and O. Isabella, "Trends and gaps in photovoltaic power forecasting with machine learning," *Energy Reports*, vol. 9, pp. 447–471, 2023, doi: 10.1016/j.egy.2022.11.208.

-
- [2] J. Gaboitaolelwe, A. M. Zungeru, A. Yahya, C. K. Lebekwe, D. N. Vinod, and A. O. Salau, "Machine Learning Based Solar Photovoltaic Power Forecasting: A Review and Comparison," *IEEE Access*, vol. 11, pp. 40820–40845, 2023, doi: 10.1109/ACCESS.2023.3270041.
- [3] Z. Ramly, M. Z. Ahmad, and M. H. Mohd Noor, "Review on the Application of Photovoltaic Forecasting Using Machine Learning for Very Short- to Long-Term Forecasting," *Sustainability*, vol. 15, no. 4, 2942, 2023, doi: 10.3390/su15042942.
- [4] J. Tian, R. Ooka, and D. Lee, "Multi-scale solar radiation and photovoltaic power forecasting with machine learning algorithms in urban environment: A state-of-the-art review," *Journal of Cleaner Production*, vol. 426, 139040, 2023, doi: 10.1016/j.jclepro.2023.139040.
- [5] A. Hasan et al., "Effects of different environmental and operational factors on the PV performance: A comprehensive review," *Energy Science & Engineering*, 2022, doi: 10.1002/ese3.1043.
- [6] M. Sharaf, M. S. Yousef, and A. S. Huzayyin, "Review of cooling techniques used to enhance the efficiency of photovoltaic power systems," *Environmental Science and Pollution Research*, vol. 29, pp. 26131–26159, 2022, doi: 10.1007/s11356-022-18719-9.
- [7] M. S. Parthiban and P. Ponnambalam, "An Enhancement of the Solar Panel Efficiency: A Comprehensive Review," *Frontiers in Energy Research*, vol. 10, 937155, 2022, doi: 10.3389/fenrg.2022.937155.
- [8] S. Kumar Pathak et al., "A detailed review on the performance of photovoltaic/thermal system using various cooling methods," *Sustainable Energy Technologies and Assessments*, vol. 51, 101844, 2022, doi: 10.1016/j.seta.2021.101844.
- [9] M. Massaoudi, S. S. Refaat, I. Chihi, H. Abu-Rub, and F. S. Oueslati, "Enhanced random forest model for robust short-term photovoltaic power forecasting using weather measurements," *Energies*, vol. 14, no. 13, 3992, 2021.
- [10] European Commission, Joint Research Centre, "Photovoltaic Geographical Information System (PVGIS)," 2025.
- [11] A. de Oliveira Santos, T. AlSkaif, G. C. Barroso, and P. C. M. de Carvalho, "Photovoltaic power estimation and forecast models integrating physics and machine learning: A review on hybrid techniques," *Solar Energy*, vol. 284, 113044, 2024, doi: 10.1016/j.solener.2024.113044.
- [12] A. Kumar, A. K. Dubey, I. Segovia Ramírez, A. Muñoz del Río, and F. P. García Márquez, "Artificial Intelligence Techniques for the Photovoltaic System: A Systematic Review and Analysis for Evaluation and Benchmarking," *Archives of Computational Methods in Engineering*, 2024, doi: 10.1007/s11831-024-10125-3.
- [13] M. Pegalajar et al., "Towards energy efficiency: A comprehensive review of deep learning-based photovoltaic power forecasting strategies," *Heliyon*, vol. 10, no. 13, e33419, 2024, doi: 10.1016/j.heliyon.2024.e33419.
- [14] S. Du et al., "Deep Learning Models for PV Power Forecasting: Review," *Energies*, vol. 17, no. 16, 3973, 2024, doi: 10.3390/en17163973.
- [15] J. Tian, R. Ooka, and D. Lee, "Multi-scale solar radiation and photovoltaic power forecasting with machine learning algorithms in urban environment: A state-of-the-art review," *Journal of Cleaner Production*, vol. 426, 139040, 2023, doi: 10.1016/j.jclepro.2023.139040.