



Prediksi dan Optimasi Performa Operasional PLTS berbasis Data LV dan MV dengan Random Forest

Adhitiya Dwi Arkan¹, Harry Dhika², Iis Dewi Lestari³

¹²³Fakultas Teknik dan Ilmu Komputer, Teknik Informatika, Universitas Indrapratar PGRI, Jakarta, Indonesia

Email: ¹adityduwi@gmail.com, ²dhikatr@yahoo.com, ³lisdewi_lestari@yahoo.co.id

Abstrak—Penelitian ini mengembangkan model prediksi performa Pembangkit Listrik Tenaga Surya (PLTS) menggunakan algoritma Random Forest dengan optimasi Bayesian. Dataset berasal dari data teknis Low Voltage (LV) dan Medium Voltage (MV) PLTS periode September 2024 milik PT Pertamina Power Indonesia. Tahapan meliputi pra-pemrosesan, seleksi fitur, pelatihan, dan evaluasi model menggunakan MAE, RMSE, dan R^2 . Hasil menunjukkan Random Forest memberikan performa terbaik dengan MAE 8,09 kW, RMSE 31,16 kW, dan R^2 0,985, mengungguli LightGBM dan XGBoost pada dataset yang sama. Model diimplementasikan ke dalam sistem web berbasis Django untuk prediksi real-time, dilengkapi interpretasi hasil, visualisasi, serta ekspor laporan PDF dan Excel. Sistem ini mendukung pemantauan kinerja PLTS yang efisien dan pengambilan keputusan berbasis data.

Kata Kunci: PLTS; Random Forest; Prediksi Performa; *Machine Learning*;

Abstract—This study develops a performance prediction model for Solar Power Plants (PLTS) using the Random Forest algorithm optimized with Bayesian optimization. The dataset is derived from Low Voltage (LV) and Medium Voltage (MV) technical data of PLTS for the period of September 2024, provided by PT Pertamina Power Indonesia. The research stages include data preprocessing, feature selection, model training, and evaluation using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). The results show that Random Forest achieved the best performance with an MAE of 8.09 kW, RMSE of 31.16 kW, and R^2 of 0.985, outperforming LightGBM and XGBoost on the same dataset. The model is implemented into a web-based system using the Django framework for real-time prediction, equipped with result interpretation, data visualization, and export features for PDF and Excel reports. This system supports efficient PLTS performance monitoring and data-driven decision-making.

Keywords: Solar Power Plant (PLTS); Random Forest; Performance Prediction; *Machine Learning*;

1. PENDAHULUAN

Kebutuhan energi listrik Indonesia terus meningkat dan mendorong akselerasi pemanfaatan Energi Baru dan Terbarukan (EBT). Berbagai kajian menunjukkan potensi surya yang sangat besar di Indonesia, baik dari sisi teknis maupun sosio-ekonomi, sehingga PLTS menjadi kandidat utama untuk perluasan bauran energi nasional (IRENA, 2023; Kementerian ESDM et al., 2023).

Meski demikian, performa keluaran PLTS dipengaruhi dinamika operasional dan kondisi sistem (mis. parameter LV/MV seperti *power factor*, *reactive power*, dan daya aktif per fasa), sehingga dibutuhkan pendekatan prediksi yang akurat dan andal.

Pendekatan *Machine Learning* (ML) banyak digunakan untuk memodelkan hubungan non-linier dan ketergantungan multivariabel pada pembangkitan surya. Tinjauan literatur menunjukkan ragam algoritma yang dipakai mulai dari model pohon hingga *gradient boosting* serta pentingnya tahapan seleksi fitur dan pra-pemrosesan yang cermat (Amiri et al., 2024; Mahesh, 2020; Panda et al., 2022). Dalam konteks ini, Random Forest menonjol karena robust terhadap *outliers*, mampu menangkap interaksi *non-linier*, dan relatif stabil pada data operasional (Salman et al., 2024).

Optimasi hiperparameter berperan penting untuk memaksimalkan kinerja model. Praktik terbaik merekomendasikan strategi eksplorasi-eksploitasi seperti *Bayesian optimization* atau skema berbasis bandit untuk efisiensi pencarian (Bischl et al., 2023). Untuk pembandingan, keluarga *gradient boosting decision tree* seperti XGBoost dan LightGBM kerap dilaporkan kompetitif pada tugas regresi energi implementasi pada penelitian ini memanfaatkan ekosistem Python/Scikit-learn untuk *pipeline* pemodelan dan evaluasi.

Penelitian ini mengusulkan model prediksi performa PLTS berbasis Random Forest dengan optimasi Bayesian pada data teknis LV/MV milik PT Pertamina Power Indonesia. Kontribusi utama



mencakup: (1) *pipeline* ringkas lima fitur operasional, (2) penalaan hiperparameter yang sistematis, (3) evaluasi terstruktur terhadap metrik MAE, RMSE, dan R^2 , serta (4) integrasi ke sistem web berbasis Django untuk prediksi real-time, interpretasi hasil, visualisasi, dan ekspor laporan.

2. METODE

2.1 Data dan Alur Penelitian

Dataset yang digunakan merupakan data teknis Low Voltage (LV) dan Medium Voltage (MV) dari PLTS milik PT Pertamina Power Indonesia periode September 2024. Variabel input meliputi *power factor* (PF), *reactive power* (Q), *phase A active power* (P_a), *phase B active power* (P_b), dan *Positive active energy* (E⁺), ringkasan fitur disajikan pada Tabel 1.

Alur penelitian mengikuti tahapan: (1) pengumpulan dan konsolidasi data, (2) pra-pemrosesan (penanganan nilai hilang/*outlier* dan penyesuaian satuan/waktu), (3) pemilihan lima fitur operasional, (4) penyusunan *pipeline*, (5) pelatihan dan penalaan model, serta (6) evaluasi dan analisis. Seluruh pemrosesan dilakukan pada ekosistem Python, menggunakan NumPy untuk komputasi array, pandas untuk manajemen data, dan scikit-learn untuk pemodelan (Harris et al., 2020).

Tabel 1. Fitur Input dan Satuannya

Fitur (ID/EN)	Simbol	Satuan	Deskripsi singkat
<i>Power factor</i> (*Power factor*)	PF	-	Rasio daya aktif terhadap daya semu (tak berdimensi).
<i>Reactive power</i> (*Reactive power*)	Q	var	Komponen daya reaktif yang tidak menghasilkan kerja.
Phase A Active Power (*Phase A*)	P _a	W	Daya aktif pada fasa A (pengukuran LV/MV).
Phase B Active Power (*Phase B*)	P _b	W	Daya aktif pada fasa B (pengukuran LV/MV).
<i>Positive active energy</i> (*Positive active energy*)	E ⁺	kWh	Akumulasi energi aktif positif yang disuplai ke beban.

2.2 Pra-Pemrosesan

Pra-pemrosesan mencakup: (1) validasi tipe/satuan, (2) pembersihan nilai hilang dengan strategi domain-driven, (3) penanganan *outlier* berbasis aturan teknis, (4) pembentukan urutan fitur yang konsisten saat inferensi untuk mencegah feature drift, dan (5) pemisahan data untuk validasi. Karena model pohon keputusan tidak sensitif terhadap skala, normalisasi tidak diwajibkan, namun konsistensi satuan tetap dijaga.

2.3 Pemodelan dan Penalaan Hiperparameter

Model utama adalah RandomForestRegressor karena ketahanannya terhadap non-linearitas dan variansi data operasional (Salman et al., 2024). Penalaan hiperparameter dilakukan dengan pendekatan Bayesian optimization untuk menyeimbangkan eksplorasi-eksploitasi ruang parameter secara efisien (Bischl et al., 2023). Ruang pencarian meliputi *n_estimators*, *max_depth*, *min_samples_split*, *min_samples_leaf*, dan *max_features*. Sebagai pembanding, model gradient boosting berbasis pohon XGBoost dan LightGBM turut dilatihkan pada konfigurasi yang wajar untuk regresi.



2.4 Skema Validasi dan Metrik Evaluasi

Validasi menggunakan skema k-fold (5-fold) pada data gabungan dan, bila relevan, GroupKFold/leave-one-site-out untuk menilai generalisasi lintas-lokasi. Kinerja diukur dengan *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan koefisien determinasi (R^2). Definisi dan praktik pelaporan metrik merujuk pada pedoman evaluasi terkini dalam ilmu data/biomedis dan implementasi scikit-learn (Miller et al., 2024). Seluruh eksperimen direplikasi dengan *random_state* yang tetap untuk menjaga keterulangan.

2.5 Implementasi Sistem

Model terlatih diintegrasikan ke aplikasi web berbasis Django untuk prediksi near real-time. Antarmuka menyediakan formulir input, visualisasi ringkas, dan ekspor laporan ke PDF (melalui ReportLab) serta Excel. Beban model diminimalkan lewat caching muatan model di memori dan validasi input sisi server.

2.6 Ketersediaan Data dan Etika

Data operasional PLTS yang digunakan bersifat internal dan tidak untuk distribusi publik. Seluruh hasil yang ditampilkan bersifat agregat dan tidak memuat informasi identitas personel maupun aset sensitif. Kode inferensi dan konfigurasi model dapat dibagikan atas permintaan yang sesuai kebijakan perusahaan.

3. ANALISA DAN PEMBAHASAN

3.1 Korelasi Fitur dan Alasan Pemilihan *Active Power*

Algoritma Random Forest merupakan metode *ensemble learning* berbasis Decision Tree yang bekerja dengan membentuk banyak pohon keputusan secara acak, kemudian menggabungkan prediksi dari masing-masing pohon untuk menghasilkan hasil akhir yang lebih stabil dan akurat.

Pada penelitian ini, target yang diprediksi adalah *active_power* (daya aktif, W) karena P (*Active Power*) merupakan keluaran listrik riil yang disalurkan ke beban dan menjadi indikator utama performa PLTS pada sisi AC. Pemilihan target AC juga sejalan dengan konteks operasi grid-connected dan kebutuhan pemantauan kinerja pada level LV/MV.

Justifikasi teknis Secara fisis, P berkaitan langsung dengan PF, Q, serta komponen daya per fasa karena itu, menggunakan *active_power* sebagai target memberi dasar kausal yang kuat. Selain itu, P dipakai luas sebagai tolok ukur efisiensi/efektivitas operasional, sehingga hasil model mudah diterjemahkan ke keputusan lapangan.

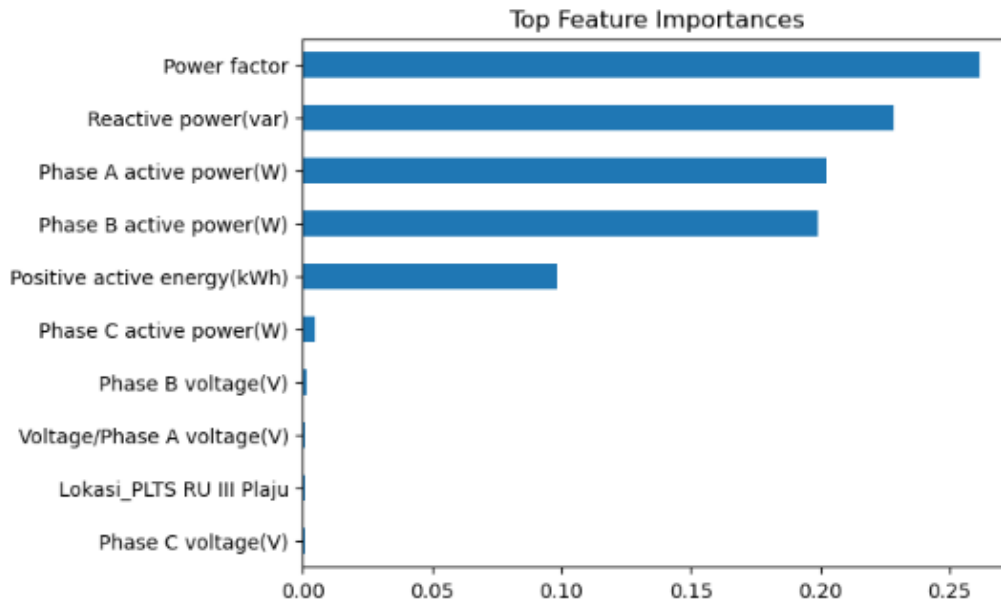
Nilai *active_power* merepresentasikan energi listrik nyata (*real power*) yang disuplai ke beban, sehingga sangat relevan digunakan sebagai tolok ukur efektivitas dan efisiensi operasional PLTS. Pemilihan *active_power* sebagai variabel target diperkuat dengan hasil korelasi numerik antar fitur, di mana *active_power* menunjukkan hubungan kuat terhadap beberapa parameter input teknis ditunjukkan pada Tabel 2.

Tabel 2. Koefisien Korelasi Antar Variabel

No.	Variable	r (Pearson)
1	<i>Active Power</i> (W)	1.00
2	Phase A <i>Active Power</i> (W)	0.67
3	Phase B <i>Active Power</i> (W)	0.66
4	<i>Reactive power</i> (var)	0.48
5	<i>Positive active energy</i> (kWh)	0.44
6	<i>Power factor</i> (PF)	0.13

3.2 Seleksi Fitur dan *Feature Importance*

Seleksi fitur dilakukan dengan pendekatan *model-driven* menggunakan Random Forest untuk menilai kontribusi relatif tiap fitur terhadap *active power* dengan mempertimbangkan interaksi dan non-linearitas. *Feature Importance* dihitung melalui *impurity-based importance* dan diverifikasi dengan *permutation importance* pada skema 5-fold CV untuk menghindari bias ditunjukkan pada Gambar 2.



Gambar 1. Diagram batang *Feature Importance* hasil pemodelan Random Forest

Tahapan meliputi: (a) pelatihan awal dengan seluruh fitur kandidat, (b) ekstraksi dan perankingan *importance* rata-rata lintas lipatan, (c) pemilihan lima fitur teratas yang stabil: *Power factor* (PF), *Reactive power* (var), *Phase A Active Power* (W), *Phase B Active Power* (W), dan *Positive active energy* (kWh).

3.3 Analisis Feature Engineering

Evaluasi *feature engineering* (rasio P/Q, indikator *phase unbalance*, agregasi waktu, interaksi sederhana) menunjukkan peningkatan akurasi tidak signifikan terhadap model dasar. Berdasarkan keseimbangan akurasi kompleksitas, penelitian ini mempertahankan lima fitur mentah sebagai set final ditunjukkan pada Tabel 3.

Tabel 3. Perbandinga Fitur 5-Fold CV dan FE

Varian.	MAE(kW)	RMSE(kW))	R ²
Base (5 Fitur)	0.964	3.320	0.999
FE Baru	0.960	3.315	0.999

3.4 Penalaan Hiperparameter dan Konfigurasi Terbaik

Pelatihan model RandomForestRegressor dilakukan dalam dua tahap pada skema *k-fold* CV (*k*=5) untuk mencegah *data leakage*, dengan *hold-out* 80:20 sebagai uji akhir. Tahap awal (*baseline*, *n_estimators*=300) digunakan untuk verifikasi *pipeline*.

Tahap kedua adalah optimisasi hiperparameter berbasis *Bayesian optimization* (Optuna, objective: minimisasi RMSE rata-rata CV) pada parameter *n_estimators*, *max_depth*, *min_samples_split*, *min_samples_leaf*, dan *max_features*. Konfigurasi terbaik kemudian di-refit pada seluruh data latih untuk menghasilkan model final ditampilkan pada Tabel 4.

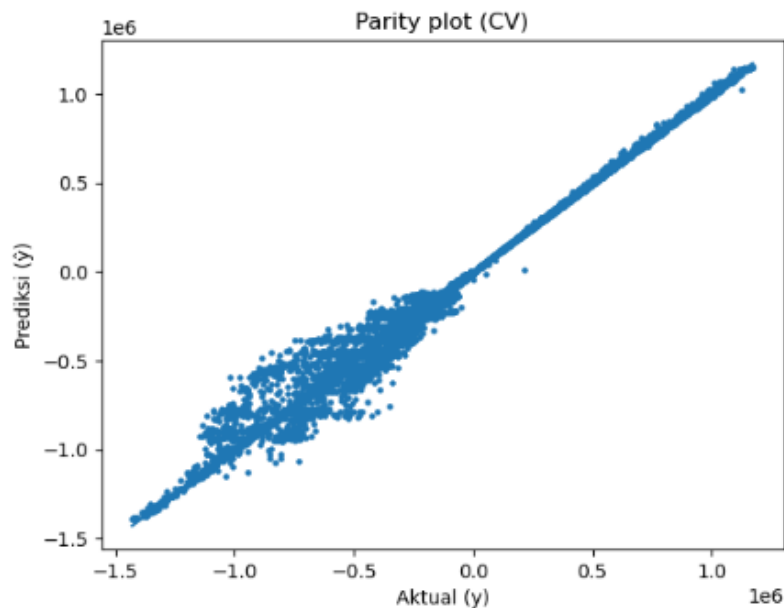
Tabel 4. Konfigurasi Parameter Random Forest Terbaik

No.	Parameter	Nilai
1	max_depth	10
2	max_features	3
3	max_samples_leaf	1
4	min_samples_split	12
5	n_estimators	1200
6	random_state	42

3.5 Visualisasi Galat (Error)

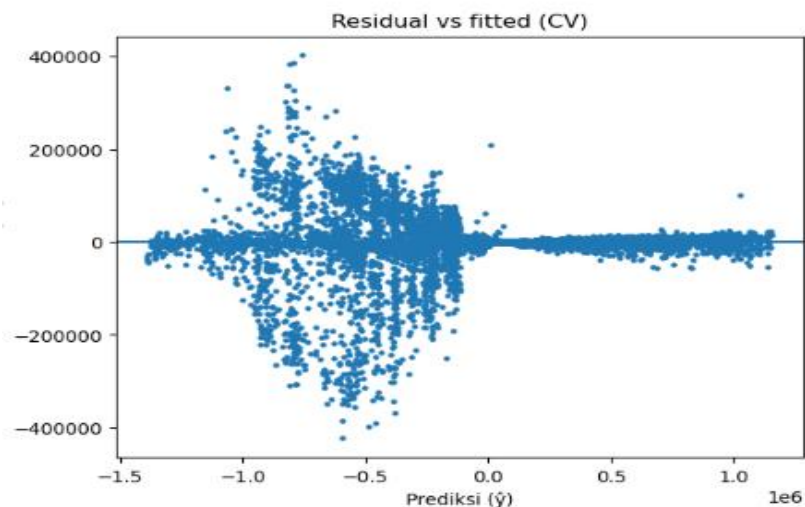
Untuk memastikan bahwa model tidak mengalami overfitting, dilakukan analisis distribusi error pada data training. Visualisasi distribusi ini dianalisis sebaran galat (residual) pada Gambar 2, Gambar 3, Gambar 4.

- a. Titik prediksi ($\hat{y}$) terhadap nilai aktual (y) pada validasi silang $k=5$ menunjukkan sebaran yang menempel pada garis identitas 45° . Ini mengindikasikan bias yang sangat kecil (kemiringan ~ 1 , intersep ~ 0) dan akurasi tinggi pada data yang tidak ikut dilatih. Penyebaran yang sedikit melebar di magnitudo besar masih wajar untuk data operasional PLTS.



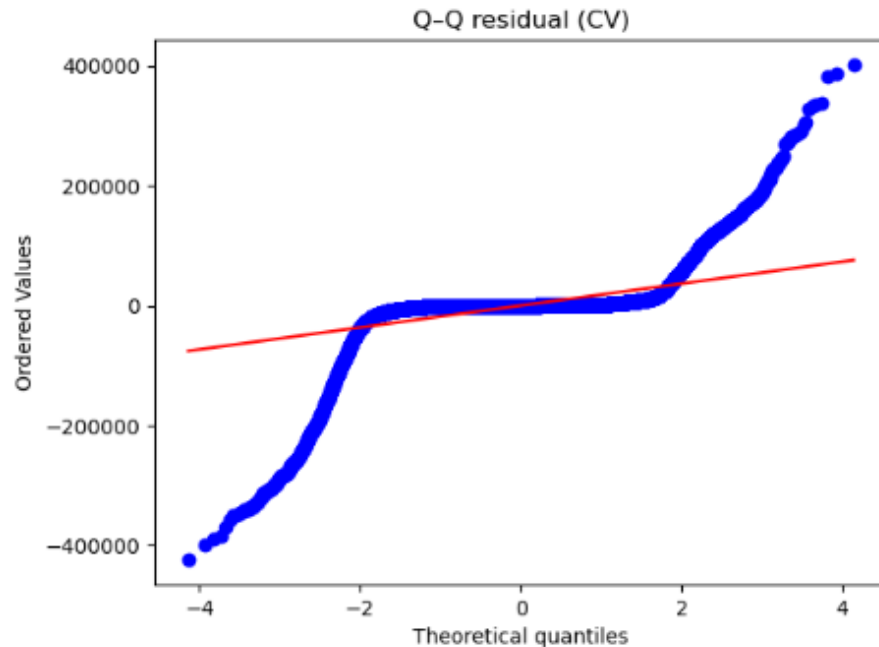
Gambar 2. Parity plot (Out-of-Fold, $k=5$)

- b. sebaran residual ($y - \hat{y}$) terhadap nilai prediksi $\hat{y}$ pada validasi silang menunjukkan titik-titik yang terpusat di sekitar nol tanpa pola melengkung yang jelas. Terlihat pelebaran residual pada magnitudo tertentu (heteroskedastisitas ringan), namun tidak membentuk pola sistematis sehingga bias model tidak terindikasi. Hal ini konsisten dengan *parity plot* Gambar 2, dan metrik evaluasi.



Gambar 3. Residual vs Fitted (Out-of-Fold, $k=5$)

- c. titik tidak mengikuti garis lurus (pola S dan ekor menyimpang), menandakan residual tidak normal dan terdapat heavy tails. Kondisi ini lazim pada data operasional PLTS dan tidak mengganggu validitas evaluasi karena model yang digunakan bersifat non-parametrik serta penilaian kinerja berbasis MAE/RMSE/R².



Gambar 4. Q-Q plot residual (Out-of-Fold, k=5)

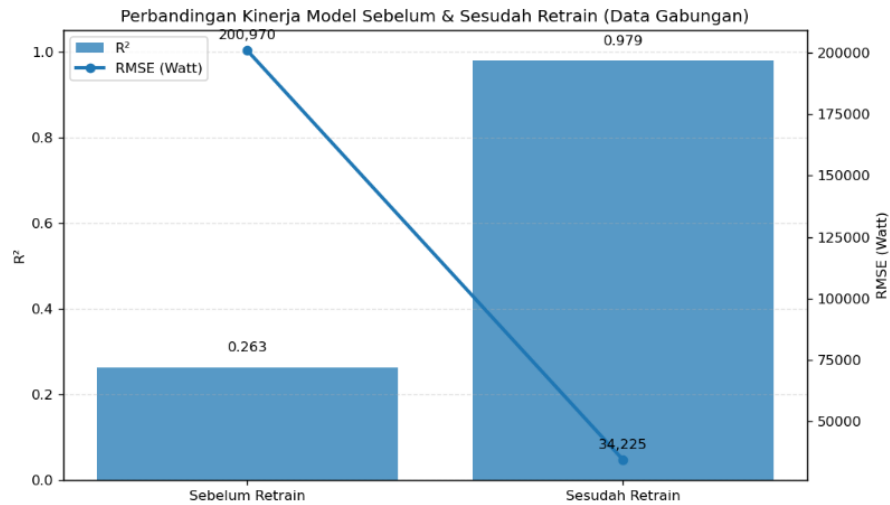
Berdasarkan tiga visual utama, parity plot Gambar 2, residual vs fitted Gambar 3, dan histogram residual Gambar 4, model Random Forest (5 fitur) menunjukkan bias yang sangat kecil (titik mengikuti garis 45°) dan sebaran galat yang terpusat di sekitar 0 tanpa pola sistematis.

Terdapat heteroskedastisitas ringan pada beban tinggi residual sedikit melebar namun masih dalam batas wajar dan tidak mengindikasikan kesalahan pemodelan. (Q-Q plot, bila ditampilkan, mengonfirmasi residual tidak persis normal, yang tidak menjadi prasyarat bagi model berbasis pohon dan evaluasi MAE/RMSE). Secara keseluruhan, visualisasi error mendukung validitas hasil evaluasi (CV/hold-out) dan menunjukkan bahwa model stabil serta layak digunakan untuk prediksi performa PLTS pada data yang tidak dilatih.

3.6 Validasi Generalisasi (LOSO / lintas-lokasi)

Validasi LOSO menunjukkan Random Forest mempertahankan R² tinggi dan MAE/RMSE (kW) yang wajar pada sebagian besar site. Pada site dengan kinerja lebih rendah, dugaan penyebabnya adalah perbedaan profil operasi/*quality* data.

Secara keseluruhan, rata-rata lintas-site mengindikasikan model tetap robust saat diterapkan pada lokasi yang tidak pernah dilihat saat pelatihan. berikut adalah grafik perbandingan dimana hasil dari gabungan dengan data plts baru dimana ada fase sebelum dan sesudah train pada Gambar 5.



Gambar 5. Perbandingan Kinerja Model Sebelum & Sesudah Retrain (Data Luar)

3.7 Evaluasi Akhir dan Perbandingan Model

Evaluasi utama dilakukan dengan 5-fold cross-validation dan hold-out 20%, disertai validasi generalisasi pada data luar serta diagnostik residual (parity plot, residual vs fitted, histogram/Q-Q). Random Forest dengan 5 fitur terpilih menunjukkan kinerja terbaik (R^2 tertinggi, RMSE/MAE terendah) dan unggul tipis namun konsisten dibanding LightGBM dan XGBoost hasil pada Tabel 5.

Tabel 5. Perbandingan Performa Metode *Machine Learning*

Model	RMSE (kW)	MAE (kW)	R^2 Score
Random Forest	31.16	8.09	0.985
LightGBM	31.98	8.15	0.984
XGBoost	32.93	9.03	0.983

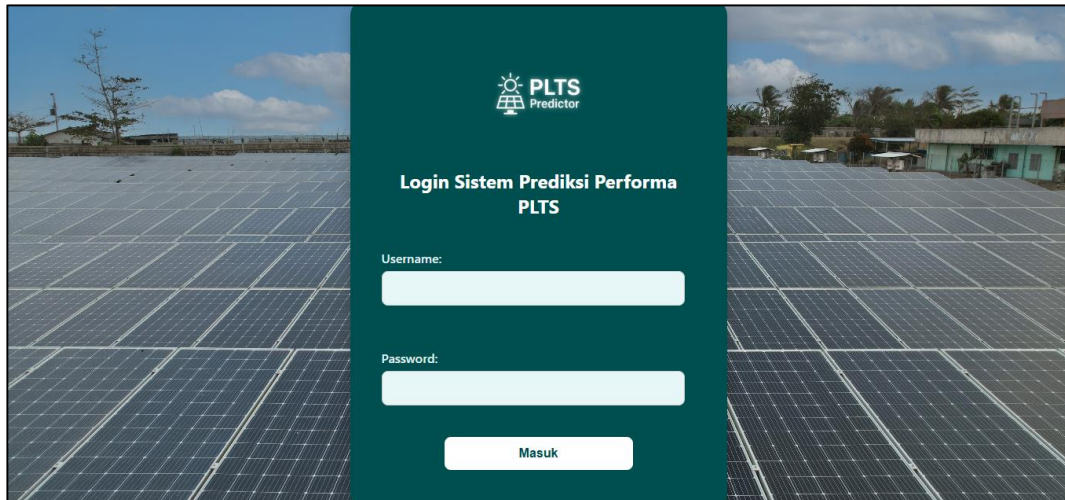
Proses *hyperparameter tuning* dan *retrain* pada data gabungan meningkatkan performa, sementara feature engineering tambahan memberi perbaikan marginal, sehingga set 5 fitur berbasis relasi fisis P-Q-S-PF dipertahankan. Model final siap diimplementasikan di antarmuka web, dengan peluang pengembangan seperti penambahan variabel meteorologi, kalibrasi per-site, *retraining* berkala, dan *quantile prediction*.

3.8 Implementasi Model pada Aplikasi Web

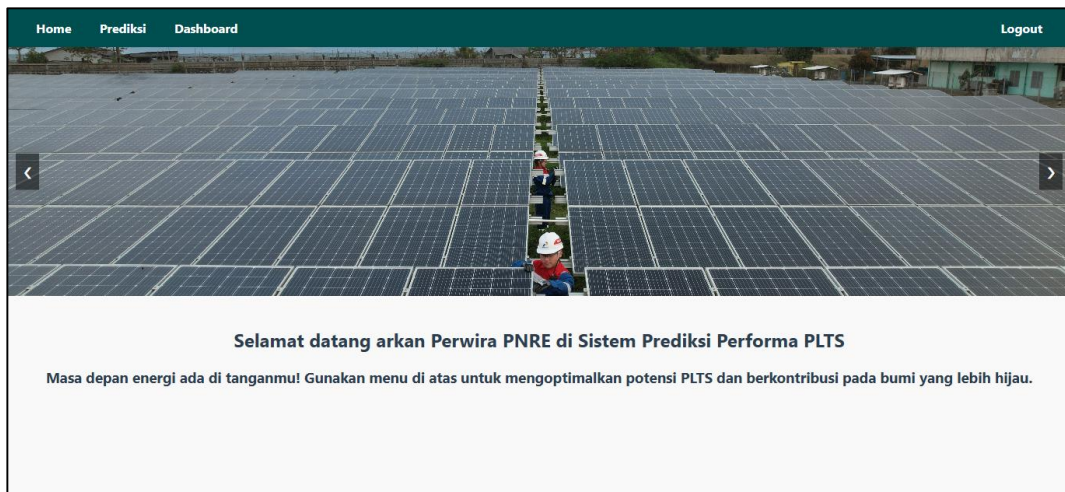
Model *Random Forest* hasil pelatihan diintegrasikan ke dalam aplikasi berbasis web menggunakan framework Django. Model disimpan dalam format .pkl (*pickle*) dan dimuat pada saat proses prediksi dijalankan. Sistem dirancang agar pengguna dapat memasukkan data teknis PLTS melalui *form input* pada halaman prediksi, meliputi lima fitur utama: *Power factor* (PF), *Reactive power* (var), *Phase A Active Power* (W), *Phase B Active Power* (W), dan *Positive active energy* (kWh) pada Gambar 6, Gambar 7, Gambar 8, Gambar 9.



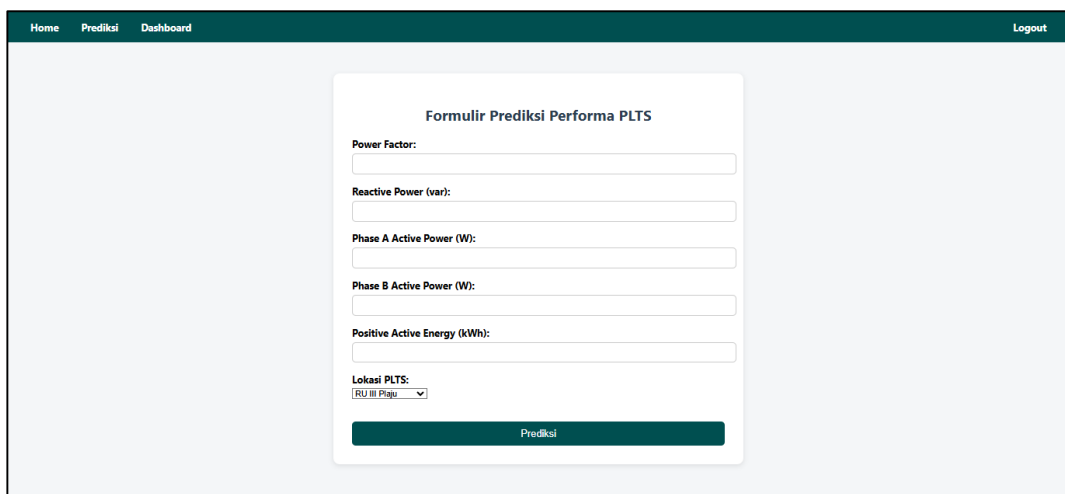
JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 3, No. 7, Desember 2025
ISSN 3025-0919 (media online)
Hal 1596-1604



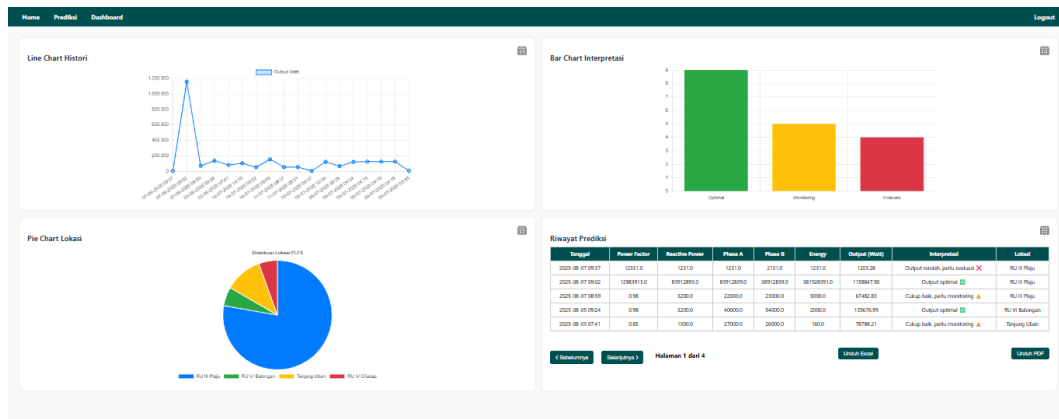
Gambar 6. Tampilan Layar Login



Gambar 7. Tampilan Layar Home



Gambar 8. Tampilan Layar Prediksi



Gambar 9. Tampilan Layar Dashboard

4. KESIMPULAN

Penelitian ini menghasilkan model prediksi performa PLTS berbasis Random Forest Regressor dengan lima fitur utama: *Power factor*, *Reactive power*, *Phase A Active Power*, *Phase B Active Power*, dan *Positive active energy*. Model menunjukkan kinerja tinggi ($R^2 \approx 0,985$) dan stabil pada uji cross-validation, hold-out, serta validasi data luar. Integrasi model ke aplikasi web berbasis Django memudahkan pengguna melakukan prediksi dan memantau kinerja PLTS secara praktis. Hasil ini menunjukkan bahwa Random Forest efektif dan layak digunakan sebagai sistem pendukung keputusan operasional PLTS..

REFERENCES

- Amiri, A. F., Chouder, A., Oudira, H., Silvestre, S., & Kichou, S. (2024). *Improving Photovoltaic Power Prediction : Insights through Computational Modeling and Feature Selection*. *Improving Photovoltaic Power Prediction : Insights through Computational Modeling and Feature Selection*.
- Bischl, B., Richter, J., Becker, M., Binder, M., & Pielok, T. (2023). *Hyperparameter optimization : Foundations , algorithms , best practices , and open challenges*. November 2022, 1–43. <https://doi.org/10.1002/widm.1484>
- Harris, C. R., Millman, K. J., Walt, S. J. Van Der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., & Kerkwijk, M. H. Van. (2020). Array programming with NumPy. *Nature*, 585(February), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- IRENA. (2023). *IRENA Socio-economic footprint of the energy transition: Indonesia*. www.irena.org
- Kementerian ESDM, Wahyu Kencono, A., Dwinugroho, M., Satra Baruna, E., & Ajiwihanto, N. (2023). *Handbook Of Energy & Economic Statistics Of Indonesia 2023*. 73.
- Maresh, B. (2020). *Machine Learning Algorithms - A Review*. October. <https://doi.org/10.21275/ART20203995>
- Miller, C., Portlock, T., Nyaga, D. M., & Sullivan, J. M. O. (2024). *A review of model evaluation metrics for machine learning in genetics and genomics*. September, 1–13. <https://doi.org/10.3389/fbinf.2024.1457619>
- Panda, S., Dhaka, R. K., Panda, B., Pradhan, A., Jena, C., & Nanda, L. (2022). A review on application of Machine Learning in Solar Energy Photovoltaic Generation Prediction. *Proceedings of the International Conference on Electronics and Renewable Systems, ICEARS 2022, Icears*, 1180–1184. <https://doi.org/10.1109/ICEARS53579.2022.9752404>
- Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random Forest Algorithm Overview. *Babylonian Journal of Machine Learning*, 2024, 69–79. <https://doi.org/10.58496/bjml/2024/007>