

Model Prediksi Kebutuhan Bahan Bakar Batu Bara Pada Pembangkit Listrik Dengan *eXtreme Gradient Boosting* (XGBoost)

Zulkifli Azis¹, Yaddarabullah², dan Budi Arifitama³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Sains Teknik & Desain, Universitas Trilogi, Jakarta

Abstrak. Prediksi konsumsi batubara yang tidak akurat dapat meningkatkan biaya operasional, emisi karbon, atau mengganggu pasokan listrik. Penelitian ini membangun model prediksi kebutuhan batubara menggunakan algoritma *XGBoost* dengan optimasi hyperparameter *Optuna* dan *GridSearch*. Dari 1000 data dengan sepuluh variabel, dipilih lima fitur utama hasil transformasi dan interaksi. Validasi menggunakan *5-Fold Cross Validation* menunjukkan bahwa *Optuna* memberikan hasil terbaik dengan MAE 0.0912, RMSE 0.3142, dan R^2 0.9641. Pada implementasi, model mampu menghasilkan prediksi mendekati aktual, misalnya 42.729 ton aktual dan 42.734 ton prediksi. Hasil ini membuktikan bahwa *XGBoost* dengan *Optuna* efektif dan dapat diandalkan untuk mendukung efisiensi operasional pembangkit listrik

Kata kunci— *Prediksi, Konsumsi Batubara, XGBoost, Optuna, GridSearch, Cross Validation, Efisiensi Operasional*

1. PENDAHULUAN

Kebutuhan energi listrik di Indonesia terus mengalami peningkatan seiring dengan pertumbuhan penduduk, ekspansi ekonomi, dan percepatan digitalisasi nasional. Berdasarkan data Kementerian ESDM [6], konsumsi listrik nasional tahun 2024 mencapai 1.408 kWh per kapita, naik dari 1.285 kWh pada tahun sebelumnya. Peningkatan kebutuhan listrik ini menuntut perencanaan pasokan energi yang semakin efisien dan akurat, terutama dalam pengelolaan bahan bakar batu bara yang masih menjadi sumber utama pembangkit listrik, khususnya Pembangkit Listrik Tenaga Uap (PLTU).

Sektor pembangkit listrik juga berkontribusi sebesar 43,83% terhadap total emisi karbon nasional [3], sehingga prediksi kebutuhan bahan bakar batu bara bukan hanya persoalan efisiensi biaya operasional, tetapi juga berkaitan dengan aspek keberlanjutan dan lingkungan hidup. Wilayah Jawa–Bali menjadi fokus utama penelitian ini karena merupakan pusat konsumsi energi nasional dengan pertumbuhan beban listrik yang tinggi [4]. Dengan demikian, diperlukan model prediksi yang mampu memperkirakan kebutuhan batu bara secara akurat untuk mendukung pengambilan keputusan dan perencanaan pasokan oleh PLN [5].

Pendekatan machine learning menjadi solusi potensial dalam menangani kompleksitas dan hubungan non-linear pada data operasional pembangkit. Algoritma *eXtreme Gradient Boosting* (XGBoost) dikenal unggul dalam performa prediksi, stabil terhadap noise, serta memiliki kemampuan generalisasi yang kuat [1]. Pada penelitian ini, XGBoost diterapkan untuk memprediksi kebutuhan bahan bakar batu bara PLTU di

¹ Corresponding author: azis-zulkifli15@gmail.com

wilayah Jawa–Bali dengan dua teknik optimasi hyperparameter, yaitu Optuna dan GridSearchCV [2]. Evaluasi dilakukan menggunakan K-Fold Cross Validation dengan metrik MAE, RMSE, dan R².

Tujuan penelitian ini adalah menganalisis performa model XGBoost dalam memprediksi kebutuhan penggunaan batu bara berdasarkan data operasional pembangkit listrik serta membandingkan efektivitas dua metode tuning hyperparameter.

Cakupan penelitian mencakup pengolahan data operasional pembangkit, penerapan transformasi data dan feature engineering, pembangunan model XGBoost, optimasi hyperparameter Optuna dan GridSearchCV, serta evaluasi performa model. Penelitian dibatasi pada data pembangkit wilayah Jawa–Bali dan tidak mencakup model prediksi untuk sumber energi lain selain batu bara.

METODE

Penelitian ini terdiri dari empat tahapan utama, yaitu pengumpulan data, pra-pemrosesan data, pembangunan model XGBoost, dan evaluasi melalui desain eksperimen berbasis K-Fold Cross Validation.

Pengumpulan Data

Data yang digunakan merupakan data operasional pembangkit listrik wilayah Jawa–Bali, mencakup variabel beban listrik, efisiensi pembangkit, nilai kalor batu bara, faktor operasi, serta total penggunaan bahan bakar. Sumber data berasal dari RUPTL PLN, Handbook of Energy and Economic Statistics of Indonesia (HEESI), dan publikasi resmi BPS. Total terdapat 1.000 entri dengan 10 variabel utama yang relevan dengan konsumsi batu bara.

Pra-Pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk memastikan kualitas data dan mempersiapkannya untuk pemodelan. Proses yang diterapkan meliputi:

1) Pengecekan dan penanganan missing values

Data dibersihkan dan disesuaikan untuk mencegah bias pada model.

$$x_{baru} = \frac{\sum x}{N} \quad (1)$$

Dengan :

x_{baru}	: Nilai untuk menggantikan data yang hilang
$\sum x$	: Total nilai dari seluruh data yang tersedia
N	: Jumlah data yang tersedia

2) Transformasi logaritmik dan Feature interaction

Transformasi log digunakan untuk menstabilkan variansi dan mengurangi efek distribusi yang terlalu lebar pada target

$$X' = \log(1 + x) \quad (2)$$

Dan Interaksi antar-variabel tertentu dibentuk untuk meningkatkan kemampuan model menangkap hubungan non-linear

$$x_3 = x_1 \times x_2 \quad (3)$$

Dengan :

X'	: Nilai setelah transformasi logaritmik
X	: Nilai asli sebelum transformasi
$\log$	: Fungsi logaritma yang digunakan untuk mereduksi rentang nilai yang terlalu besar dan mengatasi distribusi data yang skewed

3) Pemilihan fitur

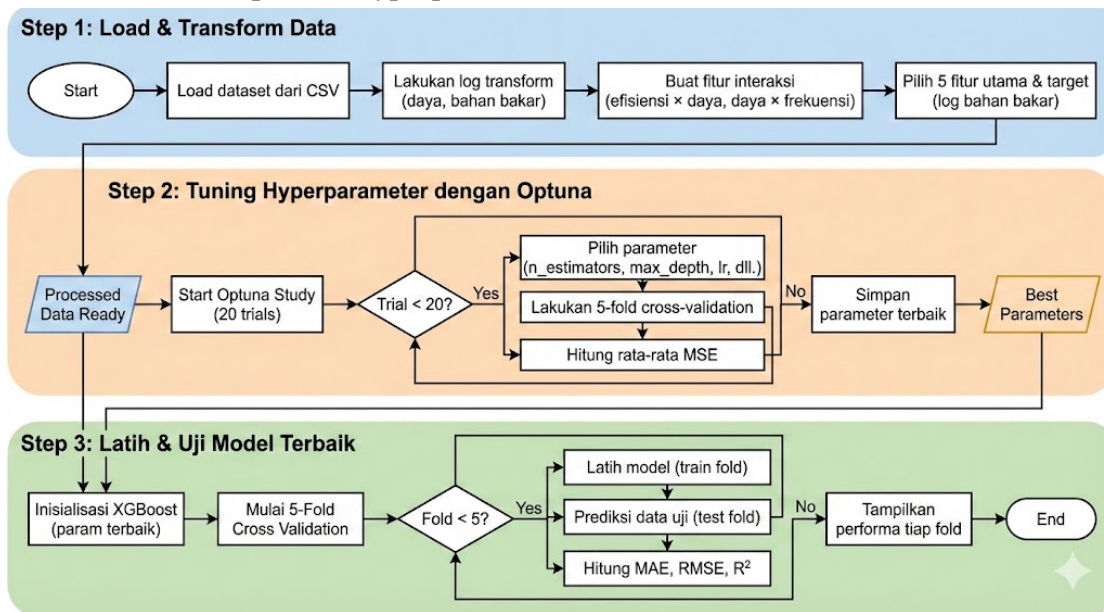
Fitur dengan korelasi rendah atau redundan dieliminasi untuk mempertahankan fitur yang paling berkontribusi pada prediksi

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (4)$$

Dengan :

d_i^2 : selisih antara peringkat dua variabel untuk observasi ke- i
 n : jumlah observasi

c. Model XGBoost dan Optimasi Hyperparameter



Gambar 1 Model XGBoost

Algoritma eXtreme Gradient Boosting (XGBoost) digunakan karena kemampuannya menangani data non-linear, robust terhadap outlier, dan efisien dalam proses training. Model dibangun menggunakan parameter dasar meliputi:

- learning_rate*
- max_depth*
- subsample*
- colsample_bytree*
- n_estimators*

Target yang diprediksi adalah total kebutuhan batu bara (dalam satuan ton) setelah transformasi logaritmik. Fungsi objektif XGBoost terdiri dari dua komponen utama, yaitu loss function dan regularization term:

$$L = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{m=1}^M \Omega(h_m) \quad (5)$$

Dengan :

$l(y_i, \hat{y}_i)$: Model pada iterasi ke- m
 $\Omega(h_m)$: Model pada iterasi sebelumnya

Optimasi hyperparameter dilakukan menggunakan framework Optuna yang berbasis algoritma *Tree-structured Parzen Estimator* (TPE). TPE memisahkan ruang parameter menjadi dua distribusi probabilitas dengan tujuan optimasi adalah memaksimalkan rasio probabilitas :

$$\arg \max_x \frac{g(x)}{l(x)} \quad (6)$$

Dengan :

$g(x)$: Distribusi ini mewakili hyperparameter dengan nilai loss rendah, yaitu kandidat berkinerja baik.

$l(x)$: Distribusi ini menggambarkan hyperparameter dengan nilai loss tinggi atau kinerja yang kurang optimal.

arg max: Operator argmax memilih kombinasi hyperparameter x yang memaksimalkan rasio probabilitas antara distribusi hyperparameter baik $g(x)$ dan buruk $l(x)$.

d. Desain Eksperimen dan Evaluasi

Eksperimen dilakukan menggunakan K-Fold Cross Validation (K=5) untuk memastikan generalisasi model. Metrik evaluasi yang digunakan meliputi:

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y'_i| \quad (7)$$

Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2} \quad (8)$$

Koefisien Determinasi (R^2)

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (9)$$

Model dengan nilai MAE dan RMSE terendah serta R^2 tertinggi dipilih sebagai model terbaik. Hasil menunjukkan bahwa metode XGBoost dengan tuning hyperparameter Optuna memberikan performa paling optimal.

3. HASIL DAN PEMBAHASAN

Tabel 1 Deskripsi Dataset

Variabel	Tipe data	Outlier
Bulan	Int64	0
Minggu	Int64	0
jumlah-pelanggan	Float64	25
jumlah pembangkit	Int64	0
daya terpasang(MW)	Float64	111
penggunaan daya(kWh)	Float64	0
jenis pembangkit	Obj	0
penggunaan bahan bakar batubara(Ton)	Int64	80
effisiensi pembangkit(%)	Int64	173
frekuensi(kWh)	Float64	9

Eksplorasi awal mengidentifikasi beberapa outlier, khususnya pada variabel daya terpasang (111), efisiensi pembangkit (173), dan penggunaan bahan bakar batubara (80). Kondisi ini mencerminkan dinamika operasional PLTU yang secara natural mengalami fluktuasi teknis antar periode. Statistik deskriptif menunjukkan variasi yang cukup lebar pada fitur-fitur utama, misalnya penggunaan daya (mean: 28.062 kWh; max: 41.851 kWh) dan penggunaan batubara (mean: 162.788 ton; max: 987.488 ton). Variasi ini menjadi dasar perlunya model prediksi yang mampu menangkap pola non-linear.

a. Transformasi logaritmik dan Feature interaction

Tabel 2 Hasil transformasi lima dari 1000 data

<i>penggunaan_daya(kWh)</i>	<i>penggunaan_bahan_bakar_batubara(Ton)</i>	<i>log_penggunaan_daya</i>	<i>log_penggunaan_bahan_bakar</i>
34055.7	42729	10.43578206	10.66265653
29814.7	97400	10.30279038	11.48659176
29128.1	37804	10.27949295	10.54019665
28228.9	60226	10.24813698	11.00587604
27335.7	6798	10.2159854	8.824530822

Transformasi logaritmik pada variabel target berhasil menstabilkan variansi dan menurunkan skewness distribusi, sehingga meningkatkan kemampuan model dalam menangkap pola prediktif.

$$\begin{aligned} \text{effisiensi_x_day} &== \text{effisiensi_pembangkit} \times \text{daya_terpasang} \\ \text{penggunaan_daya_x_frekuensi} &= \text{penggunaan_daya} \times \text{frekuensi} \end{aligned} \quad (10)$$

Feature interaction yang dibentuk pada persamaan 10 memberikan kontribusi signifikan dalam meningkatkan korelasi terhadap target.

b. Seleksi Fitur

Tabel 3 Korelasi Spearman Fitur terhadap Target dan Log Target

Nama Fitur	Korelasi terhadap target	Korelasi terhadap Log Target
penggunaan bahan bakar batubara(Ton)	1,000	1,000
log penggunaan bahan bakar	1,000	1,000
bulan	0.566	0.566
penggunaan daya	0.346	0.346
log penggunaan daya	0.346	0.346
jumlah-pelanggan	0.144	0.144
effisiensi x daya	0.047	0.047
daya terpasang(MW)	0.046	0.046
minggu	0.037	0.037
penggunaan daya x frekuensi	0.010	0.010
jumlah pembangkit	0.002	0.002
effisiensi pembangkit(%)	-0.022	-0.022
frekuensi(kWh)	-0.117	-0.117

Analisis korelasi Spearman menghasilkan lima fitur dengan nilai korelasi terbaik terhadap target maupun target-log:

- log_penggunaan_daya
- effisiensi_x_daya
- penggunaan_daya_x_frekuensi
- daya terpasang (MW)
- jumlah-pelanggan

Fitur bulan dikeluarkan karena hanya merepresentasikan urutan waktu, bukan faktor penyebab konsumsi batubara.

c. Hasil Pemodelan XGBoost

Model XGBoost dilatih menggunakan lima fitur terpilih, dengan target berupa log_penggunaan_bahan_bakar untuk meningkatkan stabilitas distribusi. Hyperparameter ditentukan menggunakan metode Optuna, dengan konfigurasi terbaik sebagai berikut:

- n_estimators = 500
- max_depth = 8
- learning_rate = 0.05
- subsample = 0.9
- colsample_bytree = 0.9
- reg_alpha = 0.01
- reg_lambda = 0.5

d. Hasil Optimasi Hyperparameter



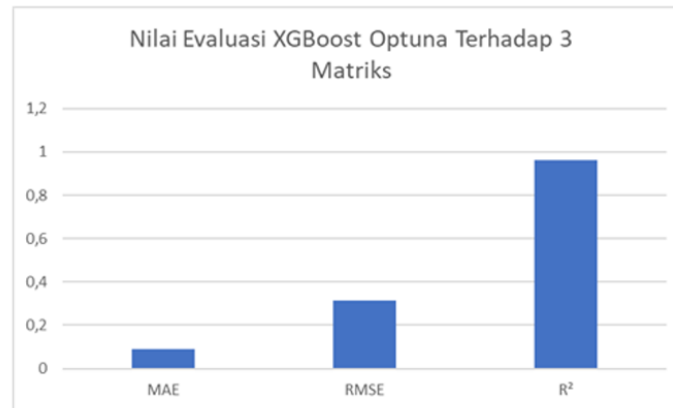
Gambar 2 Perbandingan 12 Hasil Prediksi dengan 12 Data Sebenarnya

Optimasi menggunakan Optuna telah dilakukan dan memperoleh konfigurasi model terbaik. Optuna menghasilkan kombinasi hyperparameter yang memberikan nilai error rendah. Hasil akhir menunjukkan bahwa model dengan tuning Optuna memberikan peningkatan performa signifikan. Hal ini dibuktikan oleh selisih nilai aktual dan prediksi yang sangat kecil, Contohnya pada salah satu sampel data pertama yaitu:

Aktual = 42.729 ton
Prediksi = 42.734 ton
Error = 0.005 ton

Hasil Evaluasi Model

Evaluasi menggunakan 5-Fold Cross Validation menghasilkan performa sebagai berikut:



Gambar 3 Nilai Evaluasi XGBoost Optuna Terhadap 3 Matriks

Nilai R^2 mencapai 0.9641 menunjukkan bahwa model mampu menjelaskan lebih dari 96% variabilitas penggunaan batu bara. Error yang rendah pada MAE dan RMSE memperlihatkan akurasi prediksi yang tinggi. Pada uji sampel, model memberikan selisih prediksi rata-rata 0.005 ton, yang menunjukkan presisi tinggi dalam konteks operasional.

KESIMPULAN

Secara keseluruhan penelitian ini menunjukkan bahwa model XGBoost mampu memprediksi kebutuhan bahan bakar batu bara pada PLTU Jawa–Bali dengan akurasi tinggi. Tahap pra-pemrosesan berupa transformasi logaritmik dan feature interaction meningkatkan stabilitas data, sementara optimasi hyperparameter—khususnya dengan Optuna—berhasil menurunkan error secara signifikan. Hasil evaluasi (MAE 0.0912, RMSE 0.3142, dan R^2 0.9641) serta selisih prediksi rata-rata 0.005 ton membuktikan bahwa XGBoost efektif menangkap hubungan non-linear pada data operasional dan layak digunakan sebagai alat pendukung perencanaan kebutuhan batu bara di wilayah Jawa–Bali.

UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan penelitian ini. Penelitian ini didukung oleh Universitas Trilogi, para dosen pembimbing dan teman – teman rekan mahasiswa, serta PT PLN (Persero) UPP JBB 2 Gandul yang membantu saya menyediakan sampel dan dukungan ilmu selama proses riset. Tidak lupa berterima kasih yang sebesar besarnya kepada kedua orang tua yang sudah mendoakan dan mendukung penelitian ini

- [1] X. Zhou, J. Li, Q. Wang. “Extreme Gradient Boosting Algorithm and Its Applications in Energy Forecasting”, 2020
- [2] J. Chen, X. Li, H. Zhao “Air Pollution Prediction Using XGBoost-Based Ensemble Learning Models, 2022.
- [3] D. Hasnanto and R. Setiawan. Analisis Emisi Karbon pada Sektor Energi di Indonesia, 2024.
- [4] Badan Pusat Statistik. Penduduk, Laju Pertumbuhan Penduduk, Distribusi Persentase Penduduk, Kepadatan Penduduk, Rasio Jenis Kelamin Penduduk Menurut Provinsi, 2024. [Online] from <https://www.bps.go.id/id/statistics->

- <table/3/V1ZSbFRUY3ITbFpEYTNsVWNGcDZjek53YkhsNFFUMDkjMyMwMDAw/jumlah-penduduk--laju-pertumbuhan-penduduk--distribusi-persentase-penduduk--kepadatan-penduduk--rasio-jenis-kelamin-penduduk-menurut-provinsi.html?year=2024>[Accessed on December 2024].
- [5] PLN, Rencana Usaha Penyediaan Tenaga Listrik (RUPTL), 2021. [Online] from <https://web.pln.co.id/stakeholder/ruptl> (2021). [Accessed on December 2024].
- [6] Kementerian Energi dan Sumber Daya Mineral, Handbook of Energy and Economic Statistics of Indonesia, 2024. [Online] from <https://www.esdm.go.id/assets/media/content/content-handbook-of-energy-and-economic-statistics-of-indonesia-2024.pdf> (2024). [Accessed on December 2024].
- [7] A. Sandi, K. Kusriani, and K. Kusnawi, “Analisa Prediksi Kesejahteraan Masyarakat Nelayan Lombok Timur Menggunakan Algoritma Random Forest,” *Infotek: Jurnal Informatika dan Teknologi*, vol. 6, no. 2, pp. 238–248, 2023.
- [8] D. P. Sinambela, H. Naparin, M. Zulfadhilah, and N. Hidayah, “Implementasi Algoritma Decision Tree dan Random Forest dalam Prediksi Perdarahan Pascasalin,” *Jurnal Informasi dan Teknologi*, pp. 58–64, 2023.
- [9] E. Arif, I. Suherman, and A. P. Widodo, *Revolusi Prediksi Saham: Pemanfaatan Machine Learning dan Analisis Sentimen dalam Dunia Keuangan*. Greenbook Publisher, 2025.
- [10] R. Gunawan, M. B. Dimiliu, K. Valerine, and S. P. Tamba, “Analisis Prediksi Penjualan Toko Furnitur Dengan Metode Long Short-Term Memory (LSTM),” *Jurnal Tekinkom (Teknik Informasi dan Komputer)*, vol. 7, no. 2, pp. 716–725, 2024.
- [11] D. Y. Perdana and M. A. I. Parkereng, “Prediksi Tingkat Pengangguran Berdasarkan Data Time Series Menggunakan Regresi Linear (Studi Kasus: Kota Salatiga),” *Jurnal EMT KITA*, vol. 6, no. 2, pp. 361–367, 2022.
- [12] S. Sobari, A. I. Purnamasari, A. Bahtiar, and K. Kaslani, “Meningkatkan Model Prediksi Kelulusan Santri Tahfidz di Pondok Pesantren Al-Kautsar Menggunakan Algoritma Random Forest,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, 2025.