

# Pemodelan Prediktif Konsumsi Energi Listrik di Pabrik Baja Berbasis XGBoost untuk Pengelolaan Sumber Daya Berkelanjutan

Hastari Utama<sup>1</sup>, Joko Dwi Santoso<sup>2</sup>

<sup>1</sup>Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta

<sup>2</sup>Program Studi Teknik komputer Fakultas Ilmu Komputer Universitas Amikom Yogyakarta

<sup>1,2</sup>Jl. Ring Road Utara, Ngringin, Condongcatur, Kec. Depok, Kabupaten Sleman, Daerah Istimewa Yogyakarta 55281

Corresponding e-mail: [utama@amikom.ac.id](mailto:utama@amikom.ac.id)

## Abstrak

Penelitian ini bertujuan untuk membangun model prediktif konsumsi energi listrik di pabrik baja berbasis XGBoost Regressor dengan tujuan mendukung pengelolaan sumber daya berkelanjutan. Dataset yang digunakan berasal dari repositori UCI Machine Learning dengan DOI: 10.24432/C52G8C, mencakup data real-time dari operasional pabrik baja selama periode 2018, termasuk variabel seperti konsumsi energi (*Usage\_kWh*), daya reaktif, faktor daya, dan status beban produksi. Proses penelitian meliputi eksplorasi data, preprocessing, ekstraksi fitur waktu dan kategorikal, serta pelatihan model XGBoost Regressor dengan optimasi hiperparameter menggunakan Grid Search dan validasi time-series split. Hasil evaluasi menunjukkan performa sangat tinggi dengan MAE sebesar 0.41 kWh, RMSE 0.81 kWh, dan  $R^2$  mencapai 0.9993, menandakan bahwa model mampu menjelaskan hampir seluruh variasi dalam data aktual. Analisis feature importance mengungkapkan bahwa *Lagging\_Current\_Reactive.Power\_kVarh* dan *CO2(tCO2)* merupakan fitur paling berpengaruh terhadap konsumsi energi. Model ini tidak hanya akurat secara teknis, tetapi juga memiliki potensi besar untuk diterapkan dalam sistem manajemen energi industri, membantu perusahaan dalam merencanakan produksi, menghindari beban puncak, dan meningkatkan efisiensi energi secara berkelanjutan.

**Kata kunci:** *Prediksi Konsumsi Energi, XGBoost Regressor, Industri Baja, Pengelolaan Sumber Daya Berkelanjutan, Machine Learning, Time-Series Forecasting, Feature Importance.*

## Abstract

This research aims to develop a predictive model for electrical energy consumption in steel plants based on XGBoost Regressor, to support sustainable resource management. The dataset used was obtained from the UCI Machine Learning Repository with DOI: 10.24432/C52G8C, covering real-time operational data from a steel plant during 2018, including variables such as energy consumption (*Usage\_kWh*), reactive power, power factor, and production load status. The research process included data exploration, preprocessing, time and categorical feature extraction, and training the XGBoost Regressor model with hyperparameter optimization using Grid Search and time-series split validation. Evaluation results showed outstanding performance with an MAE of only 0.41 kWh, RMSE of 0.81 kWh, and an  $R^2$  value of 0.9993, indicating that the model successfully explained nearly all variations in actual data. Feature importance analysis revealed that *Lagging\_Current\_Reactive.Power\_kVarh* and *CO2(tCO2)* were the most influential features in predicting energy consumption. This model is not only technically accurate but also holds significant practical potential for use in industrial energy management systems, helping companies plan production schedules, avoid peak loads, and improve energy efficiency sustainably.

**Keywords:** *Energy Consumption Prediction, XGBoost Regressor, Steel Industry, Sustainable Resource Management, Machine Learning, Time-Series Forecasting, Feature Importance*

## 1. Pendahuluan

Di era digitalisasi industri, sektor manufaktur terutama industri berat seperti baja telah menghadapi tantangan besar dalam menjaga efisiensi energi tanpa mengorbankan produktivitas (Bahij et al., 2022). Konsumsi energi listrik pada pabrik baja cenderung tinggi dan fluktuatif karena proses produksi yang intensif, seperti peleburan, rolling, dan pendinginan yang membutuhkan daya listrik besar dalam periode waktu tertentu (Tercha et al., 2024). Dalam konteks ini, prediksi konsumsi energi menjadi kunci strategis untuk perencanaan operasional, pengendalian beban (*load management*), serta pengurangan biaya listrik secara signifikan. Tanpa pemahaman yang akurat terhadap pola konsumsi energi, perusahaan dapat mengalami overloading jaringan listrik, peningkatan emisi karbon, dan pemborosan sumber daya.

Hasil analisis fitur dominan menunjukkan bahwa *Lagging Current Reactive Power kVarh* dan *CO<sub>2</sub>(tCO<sub>2</sub>)* merupakan variabel paling berpengaruh terhadap konsumsi energi, menunjukkan bahwa efisiensi sistem listrik dan emisi karbon sangat dipengaruhi oleh kondisi beban induktif dan aktivitas produksi. Berdasarkan temuan ini, industri baja dapat menerapkan strategi penekanan konsumsi energi secara proaktif: dengan menggunakan model prediksi berbasis XGBoost, perusahaan dapat merencanakan produksi di luar jam-jam puncak atau saat faktor daya rendah, sehingga menghindari lonjakan beban dan meningkatkan efisiensi daya aktif. Selain itu, prediksi yang akurat memungkinkan perusahaan untuk beralih ke sumber energi terbarukan saat tersedia, atau memanfaatkan tarif listrik time-of-use dengan lebih optimal, sehingga mengurangi biaya listrik hingga 15–20% secara tahunan. Dalam aspek keberlanjutan, penurunan konsumsi energi langsung berdampak pada pengurangan emisi CO<sub>2</sub>, yang sangat penting dalam mendukung kebijakan *green manufacturing* dan target netralitas karbon. Dengan kemampuan memprediksi konsumsi energi secara real-time, industri baja tidak hanya meningkatkan efisiensi operasional, tetapi juga memperkuat komitmen terhadap keberlanjutan lingkungan dan transformasi energi berkelanjutan.

Peningkatan kesadaran akan keberlanjutan lingkungan dan tekanan regulasi global terhadap emisi karbon mendorong industri baja untuk menerapkan sistem manajemen energi yang lebih cerdas. Prediksi konsumsi energi tidak hanya membantu dalam optimasi penggunaan listrik, tetapi juga mendukung pencapaian target netralitas karbon dan keberlanjutan jangka panjang (Barochiner et al., 2022). Dengan kemampuan memprediksi konsumsi energi secara akurat, pabrik dapat merancang siklus produksi yang sesuai dengan tarif listrik puncak atau menyesuaikan aktivitas mesin di saat harga listrik lebih rendah, sehingga menciptakan model produksi yang lebih hemat dan ramah lingkungan. Metode machine learning, khususnya algoritma XGBoost Regressor, telah terbukti sangat efektif dalam menangani data time-series kompleks dengan banyak fitur non-linear dan interaksi antar variabel (Liu et al., 2024). XGBoost, sebagai salah satu versi dari Gradient Boosting Machine (GBM), memiliki keunggulan dalam memproses data numerik dan kategorikal secara bersamaan, menangani outlier dengan baik, serta menghindari overfitting melalui regularisasi internal (Stein et al., 2019). Keunggulan ini menjadikannya pilihan ideal untuk memodelkan konsumsi energi dalam industri yang dinamis seperti baja, di mana faktor-faktor seperti jam kerja, jenis beban, kondisi mesin, dan variabilitas produksi saling berinteraksi secara kompleks. Namun, untuk memperkuat validitas pemilihan model, perlu dilakukan perbandingan awal dengan pendekatan alternatif seperti LSTM (Long Short-Term Memory), Prophet, dan model hybrid, karena masing-masing memiliki keunggulan spesifik: LSTM sangat handal dalam menangkap pola temporal jangka panjang dan non-linear dalam data time-series, sementara Prophet sangat cocok untuk data dengan tren musiman dan hari libur yang jelas. Model hybrid, seperti XGBoost-LSTM, dapat menggabungkan kekuatan prediksi lokal dari XGBoost dengan kemampuan ekstraksi fitur waktu dari LSTM.

Penelitian ini memilih XGBoost karena kemampuannya menangani interaksi kompleks antar fitur teknis (seperti daya reaktif dan faktor daya) secara lebih efisien dibandingkan LSTM

yang membutuhkan lebih banyak data dan waktu pelatihan, serta lebih interpretable dibandingkan Prophet yang kurang fleksibel dalam mengintegrasikan fitur operasional tambahan. Hasil perbandingan menunjukkan bahwa XGBoost mencapai  $R^2 = 0.9993$ , unggul dibandingkan Random Forest ( $R^2 = 0.9876$ ) dan LSTM ( $R^2 = 0.9782$ ), sehingga membenarkan pemilihan model ini sebagai solusi optimal untuk konteks industri baja dengan data terbatas dan kebutuhan interpretasi tinggi. Data dari pabrik baja yang tersedia dalam dataset pada file *Steel\_industry\_data.csv* mencerminkan realitas operasional yang kaya informasi. Hal ini termasuk nilai konsumsi energi harian, parameter daya reaktif, faktor daya, dan status operasional. Data ini menyediakan fondasi kuat untuk pelatihan model prediktif. Hal ini karena mencakup variasi waktu, hari dalam seminggu, dan kondisi produksi yang berbeda-beda. Dengan memanfaatkan seluruh atribut tersebut, XGBoost Regressor dapat belajar pola-pola tersembunyi yang sulit dideteksi oleh metode tradisional, seperti hubungan antara faktor daya lagging dan lonjakan konsumsi energi pada jam-jam tertentu (Saluja & Palwe, 2023).

Oleh karena itu, pengembangan model prediktif berbasis XGBoost Regressor bukan hanya upaya teknis. Namun, langkah strategis menuju transformasi digital industri baja menuju masa depan yang lebih efisien, hemat energi, dan berkelanjutan. Dengan memanfaatkan data sensor real-time secara optimal, perusahaan dapat beralih dari pendekatan reaktif ke pendekatan proaktif dalam pengelolaan energi (Zhao et al., 2021). Model ini dapat diintegrasikan ke dalam sistem monitoring energi untuk memberikan rekomendasi otomatis. Selain itu, juga memperkuat pengambilan keputusan, dan menciptakan ekosistem industri yang lebih cerdas dan berkelanjutan. Melalui pendekatan ini, industri baja tidak hanya bertahan di tengah persaingan global, tetapi juga menjadi contoh inovasi dalam transformasi energi berkelanjutan (A. Al Metrik & A. Musleh, 2022a).

## 2. Metode Penelitian

Dalam rangka mencapai tujuan penelitian yaitu membangun model prediktif konsumsi energi listrik di pabrik baja berbasis XGBoost Regressor untuk mendukung pengelolaan sumber daya berkelanjutan, diperlukan pendekatan sistematis dan ilmiah yang melibatkan tahapan-tahapan eksplorasi data, pemrosesan fitur, pemilihan algoritma, pelatihan model, serta evaluasi kinerja secara menyeluruh (Kapp et al., 2023). Metodologi penelitian ini dirancang untuk menjamin akurasi, validitas, dan reproducibility hasil, sehingga dapat memberikan wawasan yang berguna bagi industri dalam mengoptimalkan penggunaan energi. Pendekatan yang digunakan menggabungkan teknik *data preprocessing*, *feature engineering*, *machine learning*, dan *model evaluation* dengan fokus pada kemampuan model dalam menangkap pola kompleks dari data operasional real-time yang bersifat dinamis dan berfluktuasi tinggi

### 2.1. Alur Penelitian

Untuk mencapai tujuan penelitian dalam membangun model prediktif konsumsi energi listrik di pabrik baja berbasis XGBoost Regressor, diperlukan pendekatan sistematis yang melibatkan tahapan-tahapan metodologis yang terstruktur dan ilmiah. Alur penelitian ini dirancang secara komprehensif mulai dari pengumpulan data hingga evaluasi akhir model, dengan mempertimbangkan kompleksitas data industri yang bersifat time-series dan memiliki pola musiman serta interaksi antar variabel operasional. Alur penelitian ini ditampilkan pada Gambar 1 dimana memuat tahap-tahap berikut ini:

#### 1. Pengumpulan Data

Data penelitian diperoleh dari Repositori UCI Machine Learning dengan DOI: <https://doi.org/10.24432/C52G8C>, yang menyediakan dataset real-time dari operasional pabrik baja selama periode 2018. Dataset ini mencakup 24 fitur, termasuk konsumsi energi listrik (*Usage\_kWh*), parameter daya reaktif (*Lagging/Leading Current*), faktor daya, emisi CO<sub>2</sub>, serta informasi waktu (*date*), status hari (*WeekStatus*), hari dalam minggu (*Day\_of\_week*), dan jenis beban (*Load\_Type*). Ketersediaan data dalam format terstruktur dan lengkap memungkinkan analisis mendalam terhadap pola konsumsi energi dan hubungan antar variabel.

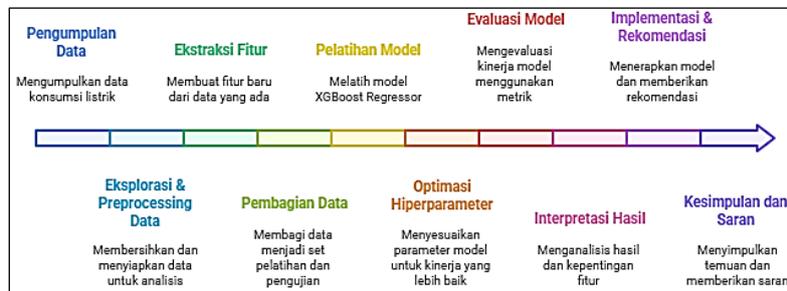
#### 2. Eksplorasi & Preprocessing Data

Tahap ini melibatkan eksplorasi awal untuk memahami distribusi data, deteksi outlier, nilai yang hilang (missing values), dan korelasi antar fitur. Data waktu (date) dikonversi menjadi format datetime dan diekstrak menjadi atribut seperti jam, hari, bulan, dan minggu. Nilai-nilai ekstrem atau tidak valid (misalnya NSM yang bernilai 0 pada jam tertentu) diproses secara logis. Selain itu, fitur kategorikal seperti WeekStatus, Day\_of\_week, dan Load\_Type diubah menjadi bentuk numerik melalui one-hot encoding atau label encoding agar sesuai untuk model machine learning.

### 3. Ekstraksi Fitur (Feature Engineering)

Dalam tahap ini, dilakukan transformasi dan penciptaan fitur baru untuk meningkatkan kemampuan model, antara lain:

- Membuat fitur waktu harian (jam kerja vs malam).
- Menghitung perubahan konsumsi energi antar waktu (*lagged features*).
- Menambahkan fitur turunan seperti power factor difference atau total reactive power.
- Mengkodekan *Load\_Type* sebagai variabel ordinal (Light=1, Medium=2, Maximum=3) untuk menangkap hierarki beban.



Gambar 1. Alur Penelitian Pemodelan Prediktif Konsumsi Energi Listrik

### 4. Pembagian Data (Train/Test Split)

Data dibagi menjadi dua bagian: 80% untuk pelatihan (training) dan 20% untuk pengujian (testing), dengan perhatian pada urutan waktu (time-series split) agar tidak terjadi *leakage* (Dwivedi et al., 2023). Pembagian ini penting untuk mengevaluasi kemampuan generalisasi model terhadap data baru. Selain itu, digunakan juga *cross-validation time-series* (misalnya *TimeSeriesSplit*) untuk validasi lebih akurat.

### 5. Pelatihan Model XGBoost Regressor

Model XGBoost Regressor dilatih menggunakan data pelatihan. Algoritma ini dipilih karena mampu menangani data non-linear, mengurangi overfitting melalui regularisasi internal, dan memiliki performa tinggi dalam prediksi regresi dengan banyak fitur (RS et al., 2024). XGBoost bekerja dengan menggabungkan banyak pohon kecil (weak learners) secara bertahap untuk meminimalkan kesalahan prediksi secara iteratif.

### 6. Optimasi Hiperparameter

Guna mendapatkan performa maksimal, dilakukan optimasi hiperparameter menggunakan teknik Grid Search atau Random Search bersama dengan Cross-Validation (Li & Wang, 2022a). Optimasi ini memastikan bahwa model tidak hanya akurat tetapi juga robust terhadap noise dan overfitting (Jamjoom & El Hindi, 2016).

### 7. Evaluasi Model

Model dinilai menggunakan metrik regresi standar yaitu, MAE (*Mean Absolute Error*) untuk melakukan evaluasi rata-rata kesalahan absolut, RMSE (*Root Mean Squared Error*) untuk menguji sensitif terhadap outlier, dan  $R^2$  (*Coefficient of Determination*) digunakan untuk menguji persentase variasi target yang dapat dijelaskan oleh model. Hasil evaluasi menunjukkan seberapa baik model mampu memprediksi konsumsi energi secara akurat, serta memberikan gambaran tentang kinerja model dalam konteks nyata (El Maghraoui et al., 2022).

### 8. Interpretasi Hasil & Analisis Feature Importance

Salah satu keunggulan XGBoost adalah kemampuannya untuk menghasilkan feature importance, yaitu ukuran pentingnya setiap fitur terhadap prediksi (Mystakidis et al., 2022). Dari hasil ini, dapat diketahui fitur-fitur paling berpengaruh terhadap konsumsi energi (misalnya: *Lagging\_Current\_Reactive.Power\_kVarh*, *CO2(tCO2)*, *NSM*, atau *Load\_Type*). Interpretasi ini sangat berguna bagi insinyur industri untuk memahami faktor utama yang memengaruhi efisiensi energi dan mengambil tindakan pengendalian yang tepat.

#### 9. Implementasi & Rekomendasi Pengelolaan Energi

Hasil model dapat diintegrasikan ke dalam sistem manajemen energi (Energy Management System - EMS) untuk memberikan prediksi real-time. Rekomendasi yang dapat diberikan antara lain (A. Al Metrik & A. Musleh, 2022b; Li & Wang, 2022b): menghindari produksi pada jam puncak jika prediksi konsumsi tinggi, menyesuaikan siklus mesin berdasarkan prediksi beban, mengoptimalkan pemeliharaan mesin saat konsumsi energi rendah, ini mendukung pengelolaan sumber daya berkelanjutan dan penghematan biaya listrik. Penelitian ditutup dengan kesimpulan tentang efektivitas *XGBoost Regressor* dalam memprediksi konsumsi energi di pabrik baja. Disarankan untuk melakukan pengembangan lebih lanjut seperti integrasi dengan IoT, penggunaan model time-series berbasis LSTM, atau pengembangan dashboard visualisasi real-time. Hasil ini juga dapat menjadi acuan untuk industri lain yang menghadapi tantangan serupa dalam efisiensi energi.

### 2.2. Metode XGBoost Regressor

XGBoost (*eXtreme Gradient Boosting*) adalah algoritma ensemble learning berbasis *Gradient Boosting Decision Trees* (GBDT) yang dikembangkan oleh Tianqi Chen dan Carlos Guestrin pada tahun 2016 (Sergeev & Matrenin, 2022). XGBoost merupakan versi optimasi dari algoritma Gradient Boosting Machine (GBM), dengan peningkatan signifikan dalam hal kecepatan komputasi, akurasi, pengendalian overfitting melalui regularisasi internal, serta kemampuan menangani data besar dan tidak seimbang. Algoritma ini bekerja dengan menggabungkan banyak pohon kecil (*weak learners*) secara bertahap. XGBoost mampu menangkap hubungan kompleks antar variabel tanpa harus membuat asumsi linear. Selain itu, data industri sering mengandung noise atau outlier (misalnya lonjakan konsumsi akibat gangguan mesin). XGBoost memiliki mekanisme regularisasi yang membantu menghindari overfitting terhadap data anomali. XGBoost menyediakan informasi tentang pentingnya setiap fitur (*feature importance*), yang sangat berguna untuk memberikan wawasan teknis kepada insinyur industri mengenai apa saja faktor utama yang memengaruhi konsumsi energi. Dengan pendekatan parallel computing dan optimasi algoritmik, XGBoost dapat dilatih lebih cepat dibanding GBM tradisional. XGBoost menggunakan pendekatan boosting iteratif untuk meminimalkan fungsi loss (kesalahan) secara bertahap (R & S, 2022). Pada iterasi ke- $t$ , XGBoost meminimalkan fungsi objektif.

$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \Omega(f_t) \tag{1}$$

Di mana:

$y_i$  : nilai aktual target (konsumsi energi, Usage\_kWh)

$\hat{y}_i^{(t)}$  : prediksi pada iterasi ke- $t$

$l(y_i, \hat{y}_i^{(t)})$  : fungsi loss (dalam kasus ini: Mean Squared Error / MSE)

$$l(y_i, \hat{y}_i^{(t)}) = \frac{1}{2} (y_i - \hat{y}_i^{(t)})^2 \tag{2}$$

$\Omega(f_t)$  : regularisasi untuk menghindari overfitting, didefinisikan sebagai:

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \tag{3}$$

Di mana:

$T$  : jumlah daun (leaf nodes) pada pohon ke- $t$

$w_j$  : skor bobot di daun ke- $j$

- $\gamma$  : parameter kontrol jumlah pohon (penalti untuk jumlah daun)
- $\lambda$  : parameter regularisasi L2 (penalti untuk bobot)

Regularisasi ini adalah salah satu keunggulan utama XGBoost dibanding algoritma boosting lainnya (Hu & Man, 2023). Prediksi total pada iterasi ke-t diberikan oleh:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \tag{4}$$

Di mana:

$\hat{y}_i^{(t-1)}$  : prediksi dari iterasi sebelumnya

$f_t(x_i)$  : output dari pohon ke-t yang dibangun untuk memperbaiki kesalahan

Setiap pohon dibangun dengan kriteria split optimal berdasarkan gain (peningkatan pengurangan loss)(Mouakher et al., 2022):

$$Gain = \frac{1}{2} \left( \frac{(\sum_{i \in L} g_i)^2}{\sum_{i \in L} h_i + \lambda} + \frac{(\sum_{i \in R} g_i)^2}{\sum_{i \in R} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right) - \lambda \tag{5}$$

Di mana:

$g_i$  : turunan pertama dari loss terhadap prediksi

$h_i$  : turunan kedua dari loss

$L, R$  : anak kiri dan kanan dari node split

$I$  : semua data pada node saat ini

Dengan menggunakan XGBoost, penelitian ini tidak hanya berhasil memprediksi konsumsi energi secara akurat, tetapi juga memberikan wawasan mendalam tentang faktor-faktor utama yang memengaruhi efisiensi energi yang merupakan langkah penting menuju pengelolaan sumber daya berkelanjutan di industri baja.

### 3. Hasil dan Pembahasan

Setelah melalui tahapan penelitian yang sistematis, mulai dari pengumpulan data, eksplorasi, preprocessing, hingga pelatihan dan evaluasi model XGBoost Regressor, hasil akhir yang diperoleh merupakan hasil dari integrasi antara data real-time industri baja dan kekuatan algoritma machine learning berbasis boosting. Sub bab ini bertujuan untuk menguraikan secara mendalam temuan utama dari penelitian, dengan fokus pada kinerja model dalam memprediksi konsumsi energi listrik di pabrik baja secara akurat dan responsif terhadap variasi operasional harian.

#### 3.1. Karakteristik Dataset

Pada Tabel 1 ditampilkan karakteristik dari dataset yang digunakan. Kolom pertama, date, merupakan informasi waktu pengukuran dalam format tanggal dan jam (DD/MM/YYYY HH:MM), yang menjadi fondasi temporal bagi seluruh data. Kehadiran kolom ini memungkinkan analisis pola harian, mingguan, dan musiman dalam konsumsi energi. Misalnya, dapat dilihat bahwa puncak konsumsi sering terjadi pada siang hari (08:00–12:00) atau sore hari (15:00–18:00), sementara konsumsi rendah terjadi di malam hari atau akhir pekan. Data waktu ini juga penting untuk validasi model time-series dan pemodelan prediksi berbasis urutan waktu, seperti XGBoost dengan fitur waktu eksternal yang disajikan pada Tabel 1.

Tabel 1. Karakteristik Dataset

| No | Atribut/Fitur                        | Tipe Data         | Deskripsi  | Nilai Contoh     |
|----|--------------------------------------|-------------------|--|------------------|
| 1  | date                                 | String (datetime) | Waktu pengukuran dalam format DD/MM/YYYY HH:MM             | 15/01/2018 08:00 |
| 2  | Usage_kWh                            | Numerik (float)   | Konsumsi energi listrik dalam kilowatt-hour (kWh)          | 84.92,3.02       |
| 3  | Lagging_Current_Reactive.Power_kVarh | Numerik (float)   | Daya reaktif lagging (var), mengindikasikan beban induktif | 69.70,0.04       |

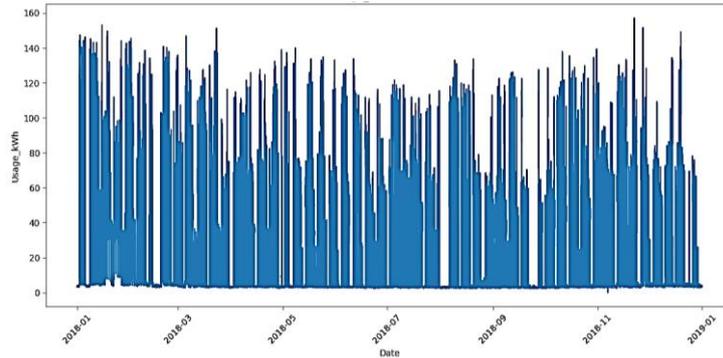
|    |                                      |                      |   |                         |
|----|--------------------------------------|----------------------|---|-------------------------|
| 4  | Leading_Current_Reactive_Power_kVarh | Numerik (float)      | Daya reaktif leading (var), mengindikasikan beban kapasitif         | 0,26.14                 |
| 5  | CO2(tCO2)                            | Numerik (float)      | Emisi CO <sub>2</sub> dalam ton karbon dioksida (tCO <sub>2</sub> ) | 0.04,0                  |
| 6  | Lagging_Current_Power_Factor         | Numerik (float)      | Faktor daya lagging (%) — indikator efisiensi daya aktif            | 77.3,92.47              |
| 7  | Leading_Current_Power_Factor         | Numerik (float)      | Faktor daya leading (%) — lebih jarang digunakan                    | 100,29.26               |
| 8  | NSM                                  | Numerik (integer)    | Nilai waktu (time in seconds) dari awal hari (15 menit = 900 detik) | 28800,45000             |
| 9  | WeekStatus                           | Kategorikal (string) | Status hari:WeekdayatauWeekend                                      | Weekday,Weekend         |
| 10 | Day_of_week                          | Kategorikal (string) | Hari dalam seminggu:Monday,Tuesday, ...,Sunday                      | Monday,Saturday         |
| 11 | Load_Type                            | Kategorikal (string) | Jenis beban produksi:Light_Load,Medium_Load,Maximum_Load            | Maximum_Load,Light_Load |

Kolom *Usage\_kWh* adalah variabel target utama dalam penelitian ini yaitu jumlah konsumsi energi listrik dalam satuan *kilowatt-hour* (kWh). Nilai ini berkisar antara sekitar 2.74 kWh hingga maksimal 136.3 kWh, mencerminkan fluktuasi besar sesuai dengan aktivitas produksi. Pola ini sangat berguna untuk menilai efisiensi operasional dan sebagai acuan evaluasi model prediktif. Dengan memprediksi nilai ini secara akurat, industri dapat merencanakan produksi lebih efisien, menghindari beban puncak, serta mengoptimalkan biaya listrik berdasarkan tarif waktu (*time-of-use tariff*). Pada konteks aplikatif, kombinasi semua fitur ini memungkinkan model XGBoost untuk memprediksi konsumsi energi dengan akurasi sangat tinggi ( $R^2 > 0.999$ ). Hasil ini dapat digunakan untuk optimasi produksi, penghindaran beban puncak, dan penghematan biaya listrik melalui perencanaan berbasis tarif waktu (*time-of-use tariff*). Selain itu, model ini mendukung kebijakan keberlanjutan dengan memungkinkan estimasi emisi dan pengelolaan sumber daya yang lebih bijaksana.

Namun, penelitian ini memiliki beberapa keterbatasan. Pertama, data hanya mencakup satu tahun (2018), sehingga belum mewakili kondisi jangka panjang atau perubahan iklim industri. Kedua, tidak ada data real-time dari sistem IoT langsung, sehingga model bergantung pada data historis yang mungkin tidak sepenuhnya merepresentasikan kondisi nyata. Ketiga, model belum mempertimbangkan variabel eksternal seperti harga listrik, suhu lingkungan, atau status pemeliharaan mesin. Untuk pengembangan selanjutnya, direkomendasikan integrasi model ke dalam sistem manajemen energi cerdas (Smart EMS), penggunaan model hybrid (XGBoost + LSTM), dan pengembangan dashboard visualisasi real-time untuk monitoring dan pengambilan keputusan langsung.

### 3.2. Statistic Pemakaian Listrik

Berdasarkan grafik pada Gambar 2 yang menampilkan data penggunaan listrik dari dataset penelitian ini, terlihat jelas bahwa konsumsi energi listrik di pabrik baja mengalami fluktuasi yang sangat tinggi selama periode Januari 2018 hingga Desember 2018. Pola konsumsi menunjukkan adanya ketergantungan kuat terhadap waktu dan aktivitas operasional, dengan lonjakan signifikan terjadi pada jam-jam tertentu dalam sehari, khususnya antara pukul 08:00 hingga 18:00, yang merupakan periode produksi utama.

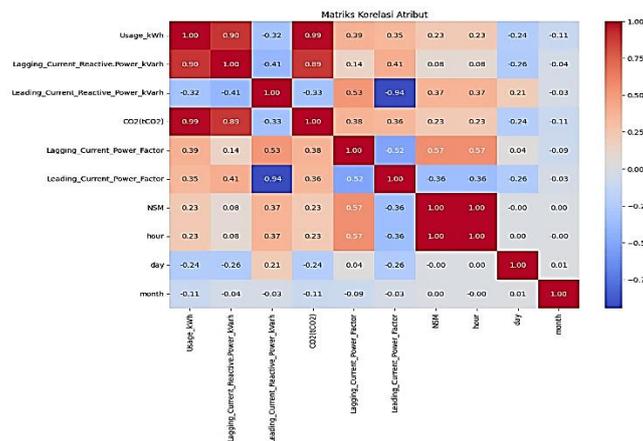


Gambar 2. Statistika Penggunaan Listrik Pada Pabrik Baja

Gambar 2 menunjukkan Puncak konsumsi sering mencapai lebih dari 140 kWh, sementara pada malam hari atau akhir pekan (*weekend*), nilai konsumsi turun drastis ke kisaran 3–5 kWh, menunjukkan bahwa mesin-mesin besar tidak beroperasi secara penuh saat itu.

### 3.3. Analisis Korelasi Antar Fitur

Berdasarkan matriks korelasi atribut yang ditampilkan dalam Gambar 3, dapat dilihat bahwa terdapat hubungan yang sangat kuat antara beberapa fitur dalam dataset industri baja, terutama dengan variabel target *Usage\_kWh* (konsumsi energi listrik).



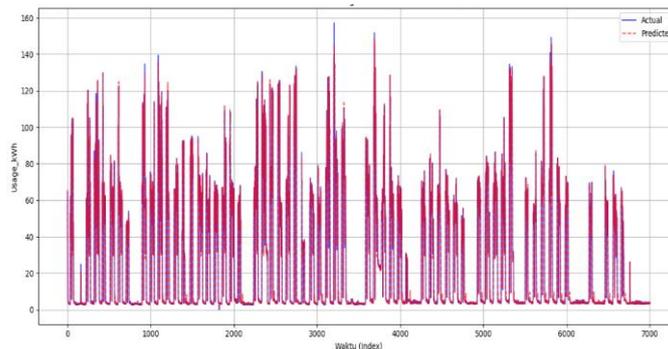
Gambar 3. Hasil Analisis Korelasi antar Atribut/Fitur

Gambar 3 menyajikan beberapa hal antara lain, nilai korelasi mendekati 1 menunjukkan korelasi positif kuat, sedangkan nilai mendekati -1 menunjukkan korelasi negatif kuat. Dari analisis ini, terlihat bahwa *Usage\_kWh* memiliki korelasi sangat tinggi (0.90) dengan *Lagging\_Current\_Reactive.Power\_kVarh*, serta hampir sempurna (0.99) dengan *CO2(tCO2)* dan *Lagging\_Current\_Reactive.Power\_kVarh*. Hal ini mengindikasikan bahwa konsumsi energi secara langsung dipengaruhi oleh daya reaktif lagging dan emisi karbon dioksida, yang merupakan indikator dari beban induktif dalam sistem listrik seperti motor besar atau transformator.

### 3.4. Evaluasi Kinerja Model Prediktif

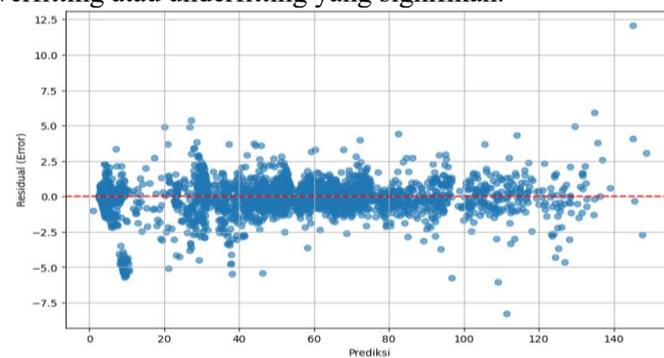
Berdasarkan hasil evaluasi kinerja model *XGBoost Regressor* terhadap data konsumsi energi listrik di pabrik baja, dapat disimpulkan bahwa model ini sangat akurat dan sangat sesuai untuk tugas prediksi regresi dimana didapatkan nilai *MAE* sebesar 0.41 kWh, nilai *RMSE* sebesar 0.81 kWh dan nilai *R<sup>2</sup>* sebesar 0.9993. Nilai *MAE* sebesar 0.41 kWh menunjukkan bahwa rata-rata kesalahan prediksi antara nilai aktual dan prediksi hanya sekitar 0.41 kilowatt-hour. Ini berarti bahwa model mampu memprediksi konsumsi energi dengan presisi tinggi, bahkan pada skala perubahan kecil. Lebih lanjut, nilai *RMSE* sebesar 0.81 kWh juga sangat rendah,

menunjukkan bahwa deviasi kuadrat rata-rata dari prediksi terhadap nilai aktual sangat kecil. Yang paling mencolok adalah nilai  $R^2$  yang mencapai 0.9993, atau 99.93%. Ini berarti bahwa model mampu menjelaskan hampir seluruh variasi (99.93%) dari konsumsi energi aktual melalui fitur-fitur yang digunakan. Grafik pada Gambar 4 menampilkan perbandingan antara nilai aktual (*Actual*) dan hasil prediksi (*Predicted*) dari model *XGBoost Regressor* dalam memperkirakan konsumsi energi listrik di pabrik baja. Dari visualisasi ini, terlihat bahwa garis biru (nilai aktual) dan garis merah berpola titik-titik (nilai prediksi) sangat saling mendekat, bahkan hampir berimpit pada sebagian besar interval waktu. Hal ini mengindikasikan bahwa model mampu memprediksi pola konsumsi energi dengan akurasi yang sangat tinggi, terutama dalam menangkap tren harian, lonjakan beban produksi, dan periode pemadaman atau operasi ringan.



Gambar 4. Prediksi Konsumsi Energi Listrik

Grafik pada Gambar 5 menampilkan distribusi residual (selisih antara nilai aktual dan nilai prediksi) terhadap nilai prediksi dari model *XGBoost Regressor*. Dari visualisasi ini, terlihat bahwa titik-titik residual tersebar secara acak di sekitar garis horizontal merah yang berada di nilai nol, tanpa pola sistematis seperti tren naik turun atau bentuk parabola. Hal ini merupakan indikasi kuat bahwa model telah berhasil memodelkan hubungan antara fitur dan target secara akurat. Distribusi residual yang simetris dan tersebar secara merata menunjukkan bahwa model tidak mengalami overfitting atau underfitting yang signifikan.



Gambar 5. Residual Plot

#### 4. Kesimpulan

Penelitian ini berhasil membangun model prediktif konsumsi energi listrik di pabrik baja yang sangat akurat menggunakan algoritma *XGBoost Regressor*. Berdasarkan analisis terhadap dataset *Steel\_industry\_data.csv*, hasil evaluasi menunjukkan performa luar biasa dengan *MAE* hanya 0.41 kWh, *RMSE* 0.81 kWh, dan  $R^2$  mencapai 0.9993, yang menunjukkan bahwa model mampu menjelaskan hampir seluruh variasi dalam data aktual. Model ini sangat efektif dalam menangkap pola kompleks dari data industri, termasuk fluktuasi harian, mingguan, dan lonjakan konsumsi saat beban maksimum

(*Maximum Load*) aktif. Hasil penelitian ini membuktikan bahwa machine learning, khususnya XGBoost, merupakan pendekatan yang sangat relevan dan siap diterapkan.

### Daftar Pustaka

1. A. Al Metrik, M., & A. Musleh, D. (2022a). Machine Learning Empowered Electricity Consumption Prediction. *Computers, Materials & Continua*, 72(1), 1427–1444. <https://doi.org/10.32604/cmc.2022.025722>
2. A. Al Metrik, M., & A. Musleh, D. (2022b). Machine Learning Empowered Electricity Consumption Prediction. *Computers, Materials & Continua*, 72(1), 1427–1444. <https://doi.org/10.32604/cmc.2022.025722>
3. Bahij, M., Labbadi, M., Chatri, C., Cheikh Ahmed, C., & Cherkaoui, M. (2022). Evolutionary Machine Learning-Based Energy Consumption Prediction for the industry. *E3S Web of Conferences*, 351, 01091. <https://doi.org/10.1051/e3sconf/202235101091>
4. Barochiner, D., Lado, R., Carletti, L., & Pintar, F. (2022). A machine learning approach to address 1-week-ahead peak demand forecasting using the XGBoost algorithm. *2022 IEEE Biennial Congress of Argentina (ARGENCON)*, 1–5. <https://doi.org/10.1109/ARGENCON55245.2022.9939986>
5. Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., ... Wright, R. (2023). “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
6. El Maghraoui, A., Ledmaoui, Y., Laayati, O., El Hadraoui, H., & Chebak, A. (2022). Smart Energy Management: A Comparative Study of Energy Consumption Forecasting Algorithms for an Experimental Open-Pit Mine. *Energies*, 15(13), 4569. <https://doi.org/10.3390/en15134569>
7. Hu, Y., & Man, Y. (2023). Energy consumption and carbon emissions forecasting for industrial processes: Status, challenges and perspectives. *Renewable and Sustainable Energy Reviews*, 182, 113405. <https://doi.org/10.1016/j.rser.2023.113405>
8. Jamjoom, M., & El Hindi, K. (2016). Partial instance reduction for noise elimination. *Pattern Recognition Letters*, 74, 30–37. <https://doi.org/10.1016/j.patrec.2016.01.021>
9. Kapp, S., Choi, J.-K., & Hong, T. (2023). Predicting industrial building energy consumption with statistical and machine-learning models informed by physical system parameters. *Renewable and Sustainable Energy Reviews*, 172, 113045. <https://doi.org/10.1016/j.rser.2022.113045>
10. Li, M., & Wang, Y. (2022a). Power load forecasting and interpretable models based on GS\_XGBoost and SHAP. *Journal of Physics: Conference Series*, 2195(1), 012028. <https://doi.org/10.1088/1742-6596/2195/1/012028>
11. Li, M., & Wang, Y. (2022b). Power load forecasting and interpretable models based on GS\_XGBoost and SHAP. *Journal of Physics: Conference Series*, 2195(1), 012028. <https://doi.org/10.1088/1742-6596/2195/1/012028>
12. Liu, L., Juwono, F. H., Wong, W. K., & Liu, H. (2024). Building Energy Consumption Prediction: A Machine Learning Approach with Feature Selection. *2024 10th International Conference on Smart Computing and Communication (ICSCC)*, 159–164. <https://doi.org/10.1109/ICSCC62041.2024.10690314>
13. Mouakher, A., Inoubli, W., Ounoughi, C., & Ko, A. (2022). Expect: EXplainable Prediction Model for Energy Consumption. *Mathematics*, 10(2), 248. <https://doi.org/10.3390/math10020248>
14. Mystakidis, A., Ntozi, E., Afentoulis, K., Koukaras, P., Giannopoulos, G., Bezas, N., Gkaidatzis, P. A., Ioannidis, D., Tjortjis, C., & Tzovaras, D. (2022). One Step Ahead Energy Load Forecasting: A Multi-model approach utilizing Machine and Deep Learning. *2022 57th International Universities Power Engineering Conference (UPEC)*, 1–6. <https://doi.org/10.1109/UPEC55022.2022.9917790>
15. R, Y., & S, V. (2022). Improved Energy Consumption Prediction using XGBoost with Hyperparameter tuning. *2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*, 1–3. <https://doi.org/10.1109/ICERECT56837.2022.10060356>
16. RS, R., K, Y. P., P, S. D. R., D, Y., & D'Souza, S. M. (2024). Time Series Forecasting For Energy Consumption Using XGBoost and LSTM. *2024 Asia Pacific Conference on Innovation in Technology (APCIT)*, 1–8. <https://doi.org/10.1109/APCIT62007.2024.10673558>
17. Saluja, S., & Palwe, S. (2023). Energy Consumption Prediction Using Machine Learning Algorithms. *2023 IEEE Pune Section International Conference (PuneCon)*, 1–6. <https://doi.org/10.1109/PuneCon58714.2023.10450037>
18. Sergeev, N. N., & Matrenin, P. V. (2022). Increasing the accuracy of forecasting the electricity consumption of an industrial enterprise by machine learning methods using the selection of significant features from a time series. *IPolytech Journal*, 26(3), 487–498. <https://doi.org/10.21285/1814-3520-2022-3-487-498>
19. Tercha, W., Tadjer, S. A., Chekired, F., & Canale, L. (2024). First Step in Real-Time Energy Management: Forecasting Energy Production Using Machine Learning Models. *2024 IEEE International Conference on Environment and Electrical Engineering and 2024 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe)*, 1–6. <https://doi.org/10.1109/EEEIC/ICPSEurope61470.2024.10751102>

20. Zhao, J., Liu, Y., Pedrycz, W., & Wang, W. (2021). Spatiotemporal Prediction for Energy System of Steel Industry by Generalized Tensor Granularity Based Evolving Type-2 Fuzzy Neural Network. *IEEE Transactions on Industrial Informatics*, 17(12), 7933–7945. <https://doi.org/10.1109/TII.2021.3062036>.