JIIFKOM (Jurnal Ilmiah Informatika & Komputer) STTR Cepu
E-Mail: jiifkom@sttrcepu.ac.id

Penerapan Algoritma *Random Forest Classifier* dan *Regressor* untuk Analisis Prediksi Efisiensi Energi Listrik

Aisyah Salsabila Azzahrah[a], Suroso[b*], Ade Silvia Handayani[c],
suroso.0719@gmail.com [b]

Program Studi Sarjana Terapan Teknik Telekomunikasi, Politeknik Negeri Sriwijaya, Indonesia[a,b,c]

Kata kunci:
Random Forest,
Klasifikasi, Regresi,
Rekayasa Fitur,
Konsumsi Energi Listrik.

Tentang naskah:
-diterima, 17 Juni 2026
-diterbitkan, 1 Juli 2026

Intisari

Efisiensi konsumsi energi listrik pada sektor domestik merupakan salah satu tantangan utama dalam *smart energy management*. Penelitian ini menerapkan algoritma *machine learning* berbasis *Random Forest* untuk menganalisis dan memprediksi pola konsumsi energi menggunakan *dataset* UCI *Household Power Consumption*. Metodologi penelitian melibatkan tahapan *feature engineering* yang intensif, termasuk pembersihan *missing values*, konversi tipe data numerik, serta ekstraksi fitur baru berupa *energy_usage*. Pengujian dilakukan melalui dua skenario utama, yaitu skenario klasifikasi menggunakan *Random Forest Classifier* dengan ambang batas keborosan 3.0 kW, serta skenario regresi menggunakan *Random Forest Regressor* untuk memprediksi nilai riil daya aktif berdasarkan parameter fisik *Voltage* dan *Global_intensity*. Hasil eksperimen menunjukkan performa model yang sangat presisi, di mana model klasifikasi berhasil mencapai tingkat akurasi sebesar 100.00%. Sementara itu, model regresi menghasilkan prediksi yang sangat akurat dengan nilai *R-Squared* (R2 Score) sebesar 99.73% dan *Mean Absolute Error* (MAE) sebesar 0.0367 kW. Analisis tingkat *feature importance* mengonfirmasi bahwa fitur *Global_active_power* dan *energy_usage* menjadi kontributor paling dominan dalam penentuan status beban. Hasil ini membuktikan bahwa pendekatan algoritma *Random Forest* sangat efektif dan andal untuk diintegrasikan ke dalam sistem monitoring energi cerdas berbasis data

Keywords:
Random Forest, Classification, Regression, Feature Engineering, Electricity Consumption.

Abstract

Efficiency of electricity consumption in the domestic sector is one of the main challenges in smart energy management. This study applies a machine learning algorithm based on Random Forest to analyze and predict energy consumption patterns using the UCI Household Power Consumption dataset. The research methodology involves intensive feature engineering stages, including cleaning missing values, numeric data type conversion, and extracting a new feature named energy_usage. The testing was conducted through two main scenarios: a classification scenario using the Random Forest Classifier with a waste threshold of 3.0 kW, and a regression scenario using the Random Forest Regressor to predict the real value of active power based on the physical parameters of Voltage and Global_intensity. The experimental results demonstrate a highly precise model performance, where the classification model successfully achieved an accuracy rate of 100.00%. Meanwhile, the regression model yielded highly accurate predictions with an R-Squared (R2 Score) value of 99.73% and a Mean Absolute Error (MAE) of 0.0367 kW. Feature importance analysis confirms that the Global_active_power and energy_usage features are the most dominant contributors to determining load status. These results prove that the Random Forest algorithm approach is highly effective and reliable for integration into data-driven smart energy monitoring.

1. Pendahuluan

Kebutuhan energi listrik pada sektor rumah tangga terus meningkat. Peningkatan ini didorong oleh pertumbuhan populasi dan adopsi perangkat elektronik berdaya tinggi. Konsumsi energi yang tidak terkontrol memicu pembengkakan biaya operasional bulanan. Kondisi ini juga meningkatkan beban puncak pada jaringan distribusi listrik (Syafwan, 2023). Manajemen energi konvensional mengandalkan pencatatan meteran manual di akhir bulan. Metode ini tidak lagi efektif karena tidak memberikan visibilitas *real-time* atas pola konsumsi energi harian penghuni rumah.

Integrasi data *science* dan *machine learning* menjadi solusi untuk mengatasi masalah tersebut. Teknologi ini menganalisis data historis kelistrikan

secara komputasional. *Feature engineering* mengungkap karakteristik tersembunyi pada parameter fisik seperti tegangan (Voltage) dan intensitas arus (Global_intensity). Parameter ini berpengaruh terhadap pemakaian daya aktif secara keseluruhan (K. C. Demirkadir, 2024). Pemodelan prediktif mengelompokkan kondisi rumah ke dalam status boros atau hemat. Pemodelan ini juga memproyeksikan besaran daya aktif yang akan digunakan secara presisi(Al-Qahtani, 2022).

Penelitian terdahulu menggunakan metode klasterisasi tradisional seperti K-*Means* untuk mengelompokkan profil konsumen (Samsinar, 2026). Metode ini memiliki keterbatasan dalam menangani hubungan data *nonlinear* yang kompleks. Metode ini juga rentan terhadap *missing values* dalam pencatatan berkala. Sebaliknya, algoritma ensemble learning seperti *Random Forest* memiliki keandalan tinggi. Algoritma ini juga memiliki stabilitas komputasi yang kuat. *Random Forest* mencapai akurasi superior untuk tugas klasifikasi biner maupun regresi tanpa mengalami *overfitting*.

Algoritma SVM sering menghadapi kendala efisiensi komputasi pada volume data deret waktu berskala besar(de la Torre, 2026). SVM juga menghadapi kendala dalam menentukan batas keputusan (hyperplane) pada kondisi tersebut. Selain itu, SVM membutuhkan waktu pelatihan yang lama. SVM juga kurang sensitif terhadap *outlier* ekstrem, padahal *outlier* sering menjadi indikator pemborosan energi. Algoritma boosting seperti XGBoost melatih pohon keputusan secara berurutan. Proses ini membuat XGBoost rentan terhadap *overfitting* saat data latih mengandung noise atau variabilitas temporal tinggi (Shinde, 2024). *Random Forest* menjadi pilihan paling ideal karena berbasis *Bootstrap Aggregating* yang bekerja secara paralel. Pendekatan ini stabil dalam mereduksi varians kesalahan internal. *Random Forest* juga mampu menangkap sinyal esensial dari fitur hasil rekayasa secara objektif.

Penelitian ini mengimplementasikan algoritma *Random Forest* pada *dataset* UCI *Household Power Consumption*. *Dataset* ini telah dioptimalkan melalui teknik *feature engineering* untuk dua skenario utama. Skenario pertama menggunakan *Random Forest Classifier* untuk mendeteksi status keborosan beban berdasarkan ambang batas 3,0 kW. Skenario kedua menerapkan *Random Forest Regressor* untuk memprediksi angka riil daya aktif secara presisi. Pendekatan *dual-scenario* ini diharapkan berkontribusi pada pengembangan *smart energy monitoring system* yang adaptif dan berbasis data komputasi.

2. Kerangka Teori

2.1. Energy Management pada Sektor Domestik

Konsumsi energi listrik di sektor rumah tangga (domestik) memiliki pola fluktuasi yang sangat dinamis, tidak menentu, dan kompleks akibat variasi aktivitas penghuni. Berdasarkan prinsip *Smart Energy Management*, efisiensi energi yang berkelanjutan tidak lagi dicapai melalui pembatasan penggunaan fisik secara manual, melainkan melalui sistem pemantauan berbasis data (data-driven monitoring). Pemodelan prediktif dalam manajemen energi modern berfungsi krusial untuk mengidentifikasi perilaku konsumsi, mendeteksi anomali seperti keborosan struktural, serta mengoptimalkan pembagian beban (load balancing) secara *real-time*. Pemanfaatan data historis meteran cerdas secara masif membuka peluang bagi penyedia layanan dan konsumen untuk merancang strategi efisiensi yang lebih personal dan adaptif terhadap lonjakan beban harian (de la Torre, 2026).

2.2. Machine Learning berbasis Ensemble Learning

Ensemble learning merupakan sebuah paradigma dalam pembelajaran mesin di mana beberapa model prediktif (seperti kumpulan pohon keputusan) dilatih bersama untuk menyelesaikan masalah yang sama. Pendekatan ini bekerja dengan cara mengombinasikan *output* dari beberapa model linier atau non-linier guna memperoleh akurasi prediksi tunggal yang jauh lebih kuat dan stabil. Dalam domain analisis data energi, pendekatan *ensemble* terbukti sangat efektif untuk mereduksi risiko *overfitting* yang sering terjadi akibat varians data yang tinggi pada deret waktu (*time-series*). Algoritma kelompok ini mampu menangani *noise*, mendistribusikan bobot kesalahan secara merata, serta meningkatkan kemampuan generalisasi model saat dihadapkan pada karakteristik data baru yang belum pernah dilatih sebelumnya (Shinde, 2024).

2.3. Algoritma Random Forest

Random Forest merupakan algoritma *ensemble learning* berbasis pohon keputusan (*Decision Trees*) yang memanfaatkan metode *Bagging (Bootstrap Aggregating)* dan seleksi fitur acak secara simultan. Algoritma ini membangun banyak pohon keputusan selama fase pelatihan dan mengeluarkan *output* berupa modus untuk klasifikasi atau rata-rata prediksi untuk kasus regresi. Karakteristik utama dari *Random Forest* adalah ketahanannya yang tinggi terhadap *multicollinearity* serta kemampuannya mengelola data berdimensi besar tanpa menurunkan performa komputasi secara drastis.

2.4. Random Forest Classifier

Random Forest Classifier digunakan secara spesifik untuk memetakan *input* data multidimensi ke dalam kategori atau kelas diskrit tertentu yang telah ditentukan. Algoritma ini bekerja dengan membagi ruang fitur secara tegak lurus berdasarkan batas keputusan optimal yang memisahkan variabilitas data secara maksimum. Dalam konteks pemantauan energi, model *classifier* berperan penting dalam menetapkan status beban konsumsi harian berdasarkan ambang batas (*threshold*) tertentu. Melalui metode pembagian keputusan yang ketat, algoritma ini mampu memisahkan kondisi operasional normal dengan kondisi beban puncak ekstrem yang mengindikasikan adanya pemborosan energi (Rochayati, 2025).

2.5. Random Forest Regressor

Random Forest Regressor ditujukan untuk memproyeksikan nilai kontinu atau angka riil secara presisi dengan cara mengambil nilai rata-rata dari seluruh prediksi pohon keputusan individual. Kemampuannya yang luar biasa dalam menangani hubungan non-linier yang rumit antara parameter fisik (seperti tingkat tegangan dan intensitas arus) menjadikannya sangat andal dalam mengestimasi daya aktif harian. Algoritma regresi ini meminimalkan varians kesalahan secara internal melalui mekanisme *cross-validation* bawaan yang terbentuk selama proses pemisahan *node* acak. Oleh karena itu, estimasi angka kontinu yang dihasilkan memiliki tingkat penyimpangan atau galat yang sangat kecil, bahkan ketika menghadapi fluktuasi konsumsi listrik yang bersifat acak dan ekstrem (Sharma, 2025).

2.6. Feature Engineering

Feature engineering merupakan proses transformasi matematis untuk mengubah data mentah menjadi fitur-fitur baru yang lebih representatif agar algoritma pembelajaran mesin dapat bekerja secara optimal. Tahapan ini diawali dengan pembersihan data (data cleaning) secara intensif, mencakup penanganan nilai yang hilang (missing values) serta konversi tipe data ke dalam bentuk numerik yang valid untuk mencegah kegagalan komputasi. Selain pembersihan data standar, ekstraksi fitur baru yang merepresentasikan akumulasi temporal penggunaan energi (seperti variabel energy_usage) sangat penting untuk menyuplai informasi esensial ke dalam model. Rekayasa fitur yang dirancang secara tepat terbukti mampu meningkatkan sensitivitas arsitektur model, memperjelas korelasi antar variabel, serta secara signifikan mempertajam akurasi dari hasil prediksi akhir (Chicco, 2022).

2.7. Metrik Evaluasi Performa Model

Metrik Evaluasi Performa Model sebagai cara mengukur validitas, keandalan, dan tingkat presisi dari model algoritma yang telah dibangun, diperlukan metrik evaluasi standar yang diakui secara ilmiah. Pengujian pada penelitian ini menggunakan tiga jenis metrik evaluasi yang disesuaikan dengan karakteristik skenario masing-masing model, yaitu:

2.7.1 Accuracy Score (Akurasi)

Metrik ini digunakan untuk mengukur rasio atau persentase total prediksi klasifikasi yang benar dari model Classifier terhadap keseluruhan data uji yang diberikan. Evaluasi ini memberikan gambaran instan mengenai kemampuan model dalam mengategorikan status konsumsi energi secara tepat tanpa mengalami bias.

2.7.2. R-*Squared* (R2 Score)

Metrik ini berfungsi sebagai koefisien determinasi yang mengukur seberapa baik variasi nilai aktual dari daya aktif dapat dijelaskan secara linier maupun non-linier oleh parameter *input* fisik pada model *Regressor*. Nilai R2 yang mendekati angka satu menandakan bahwa parameter *input* yang digunakan telah berhasil mewakili hampir seluruh karakteristik data aktual.

2.7.3. *Mean Absolute Error* (MAE)

Metrik ini digunakan untuk mengukur rata-rata besarnya kesalahan atau galat absolut antara nilai daya yang diproyeksikan oleh model dengan nilai riil yang tercatat di lapangan. Semakin kecil nilai MAE yang dihasilkan, maka semakin tinggi tingkat kedekatan serta akurasi model prediktif terhadap kondisi fisik yang sebenarnya terjadi pada sistem (Sharma et al., 2025).

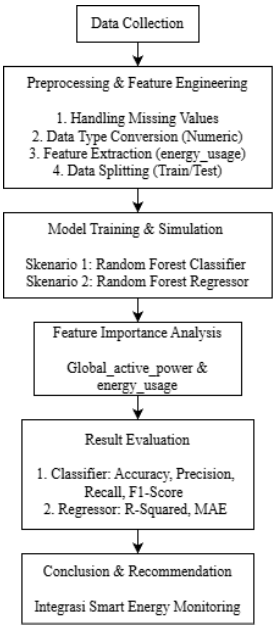
2.8. Feature Importance

Feature Importance merupakan sebuah teknik evaluasi internal pada algoritma berbasis pohon keputusan yang berfungsi untuk menghitung nilai kontribusi relatif dari setiap variabel input terhadap target prediksi. Nilai kepentingan ini dihitung berdasarkan seberapa besar suatu fitur dapat menurunkan tingkat pengotor (impurity) atau *varians* di seluruh node pohon keputusan yang terbentuk. Melalui analisis kuantitatif ini, dominasi pengaruh dari fitur konvensional maupun fitur baru hasil rekayasa dapat dibuktikan secara matematis dan ilmiah. Hasil evaluasi ini pada akhirnya memberikan rekomendasi yang valid mengenai kombinasi fitur terbaik yang paling andal untuk diintegrasikan ke dalam arsitektur sistem *monitoring* energi cerdas (K. Demirkadir, 2024).

3. Metodologi

Penelitian ini menggunakan metodologi kuantitatif eksperimental. Pendekatan ini dipilih karena cocok untuk mengukur hubungan antar variabel secara numerik. Metodologi ini juga menerapkan siklus standar pemrosesan data *machine learning*. Siklus ini memastikan setiap tahapan penelitian berjalan secara terstruktur dan terukur. Proses komputasi dalam penelitian ini dirancang secara sistematis. Proses ini dimulai dari akuisisi *dataset* mentah dari sumber data publik. *Dataset* mentah tersebut kemudian melalui tahap praproses untuk membersihkan data yang tidak konsisten. Tahap selanjutnya adalah feature engineering untuk menghasilkan fitur-fitur baru yang relevan. Fitur-fitur ini kemudian digunakan untuk melatih model *machine learning*. Proses pelatihan ini berakhir pada penghasilan model prediktif yang siap dievaluasi (F. Martinez-Plumed, 2022).

Evaluasi model dilakukan untuk mengukur tingkat akurasi dan kestabilan hasil prediksi.



Gambar 1. Metodologi Penelitian

Berdasarkan Gambar 1, alur metodologi penelitian ini dirancang secara sistematis yang diawali dengan tahap data *collection* untuk mengambil data mentah dari *dataset* UCI *Household Power Consumption. Dataset* yang berisi catatan historis penggunaan listrik skala rumah tangga tersebut kemudian dibawa ke tahap *preprocessing* dan *feature engineering* untuk mengoptimalkan kualitas data (K. Demirkadir, 2024). Pada tahapan ini, dilakukan pembersihan nilai kosong (handling missing values) akibat malfungsi sensor, konversi seluruh parameter menjadi tipe data numerik agar dapat diproses oleh sistem, serta ekstraksi fitur baru berupa *energy_usage* sebelum akhirnya *dataset* dibagi menjadi data latih dan data uji melalui proses data *splitting*.

Data yang telah matang kemudian dialirkan menuju tahap model *training* dan *simulation* yang menerapkan algoritma R*andom Forest* ke dalam dua skenario utama (P. Shinde, et al, 2024). Skenario pertama menggunakan *Random Forest Classifier* untuk mengategorikan status pemakaian listrik menjadi hemat atau boros dengan ambang batas 3.0 kW, sedangkan skenario kedua menerapkan *Random Forest Regressor* untuk memprediksi angka riil daya aktif secara

kontinu. Setelah model berhasil dibangun, alur berlanjut ke tahap *feature importance* analysis untuk mengevaluasi tingkat kepentingan variabel, di mana parameter *Global_active_power* dan *energy_usage* teridentifikasi sebagai kontributor paling dominan dalam penentuan status beban kelistrikan.

Rangkaian proses ini kemudian memasuki tahap *result evaluation* untuk menguji keandalan model menggunakan metrik pengujian standar ilmiah (Tan, 2025). Pada skenario klasifikasi, performa model dinilai berdasarkan metrik *Accuracy, Precision, Recall,* dan F1-*Score*, sedangkan pada skenario regresi, akurasi prediksi diukur melalui nilai R-Squared dan *Mean Absolute Error* (MAE). Seluruh tahapan metodologi pada Gambar 1 ini diakhiri dengan tahap conclusion dan recommendation yang menyimpulkan hasil eksperimen sekaligus memberikan rekomendasi ilmiah untuk mengintegrasikan model Random Forest.

3.1. Dataset

Dataset yang digunakan dalam penelitian ini adalah UCI *Household Power Consumption*, sebuah dataset publik yang tersedia pada repositori UCI *Machine Learning Repository* . *Dataset* ini merekam konsumsi daya listrik pada satu rumah tangga di Sceaux, Prancis, dengan frekuensi sampling satu menit selama periode Desember 2006 hingga November 2010, menghasilkan total sekitar 2.075.259 record mentah. Setiap record memuat tujuh variabel fisik utama seperti pada Tabel 1:

Tabel 1. Karakteristik Variabel Kuantitatif *Dataset*

Variabel	**Data yang diambil**
Global_active_power	Daya aktif global dalam kW
Global_reactive_power	Daya reaktif global dalam kW
Voltage	Tegangan dalam Volt
Global_intensity	Intensitas arus global dalam Ampere
Sub_*metering*_1	Konsumsi sub-meteran dapur dalam Wh
Sub_*metering*_2	Konsumsi sub-meteran ruang cuci dalam Wh
Sub_*metering*_3	*Konsumsi* sub-meteran pemanas air dan AC dalam Wh

Variabel-variabel ukur fisik yang digunakan sebagai fitur masukan (input features) dalam pemodelan ini dirangkum secara rinci pada Tabel 1. Parameter tersebut merepresentasikan data kuantitatif dari aktivitas kelistrikan domestik secara *real-time* yang mencakup dimensi daya, tegangan, intensitas arus, hingga sebaran konsumsi energi sektoral pada *sub-metering* spesifik, untuk mengoptimalkan kinerja komputasi algoritma selama tahapan eksperimen, diambil subsampel sebanyak 10.000 baris data dari populasi asli secara acak (random sampling) dengan mengunci parameter random_state=42 guna menjamin sifat reproduksibilitas hasil evaluasi model.

3.2 Preprocessing

Tahap preprocessing diawali dengan penanganan missing values yang muncul akibat malfungsi sensor pada periode pencatatan tertentu. Nilai kosong yang terdeteksi dieliminasi menggunakan metode dropna() sehingga hanya baris data lengkap yang diikutsertakan dalam proses analisis. Seluruh kolom numerik selanjutnya dikonversi secara eksplisit menggunakan fungsi pd.to_numeric() dengan parameter errors='coerce' untuk memastikan konsistensi tipe data sebelum pemrosesan model. Pada tahap *feature engineering*, diekstraksi satu fitur baru bernama *energy_usage* yang dihitung berdasarkan rumus:

$$Energy\ usage = Global\_active\_power \cdot \frac{1}{60} \quad (1)$$

Merepresentasikan estimasi konsumsi energi aktual dalam satuan kWh per interval satu menit. Fitur ini dirancang untuk memberikan representasi energi yang lebih intuitif dibandingkan nilai daya sesaat, sehingga mampu meningkatkan

kemampuan diskriminatif model. Pembuatan label target klasifikasi dilakukan dengan menetapkan ambang batas (threshold) sebesar 3.0 kW pada kolom *Global_active_power*, di mana nilai ini dipilih berdasarkan nilai rata-rata (mean) distribusi daya aktif pada dataset yang digunakan. Rekaman dengan *Global_active_power* ≥ 3.0 kW diberi label 1 (Boros/Tinggi), sedangkan di bawah *threshold* diberi label 0 (Hemat/Normal), menghasilkan distribusi kelas sebesar 80.55% label 0 dan 19.45% label 1.

3.3 Konfigurasi Model dan Pembagian Data

Dataset yang telah diproses dibagi menggunakan metode *train-test split* dengan rasio 80:20, di mana 80% data (8.000 record) dialokasikan sebagai data latih (training set) dan 20% sisanya (2.000 record) sebagai data uji (testing set) untuk mengevaluasi metrik akurasi (Gholamy, 2023). Pembagian dilakukan secara acak dengan parameter random_state=42 guna memastikan reproduksibilitas hasil eksperimen. Kedua model *Random Forest* dikonfigurasi dengan parameter n_estimators=100 (jumlah pohon keputusan), max_depth=None (kedalaman pohon tidak dibatasi), min_samples_split=2, dan random_state=42.

Pemilihan algoritma *Random Forest* didasarkan pada kemampuannya menangani hubungan nonlinear yang kompleks antar variabel, ketahanannya terhadap *outlier* dan *noise* data, serta kemampuannya menghasilkan estimasi *feature importance* yang akurat melalui mekanisme agregasi pohon keputusan (ensemble). Dibandingkan dengan algoritma sejenis seperti *Decision Tree* tunggal yang rentan *overfitting*, atau *Support Vector Machine* yang memiliki kompleksitas komputasi tinggi pada dataset besar, *Random Forest* menawarkan keseimbangan optimal antara akurasi, stabilitas, dan efisiensi komputasi untuk tugas klasifikasi maupun regresi pada domain konsumsi energi (Al-Qahtani, 2022).

4. Hasil dan Pembahasan

4.1 Persiapan Data dan Hasil Pra-pemrosesan

Proses analisis pada penelitian ini diawali dengan menguji kesiapan *dataset* melalui program yang dijalankan pada Google Colab. Sistem berhasil membaca file data mentah UCI *Household Power Consumption* dengan konfigurasi CSV standar secara optimal.

Setelah melalui tahapan pra-pemrosesan yang intensif,meliputi pembersihan nilai kosong (handling missing values) akibat malfungsi sensor kelistrikan serta konversi tipe data parameter menjadi numerik, tercatat total data yang siap diproses secara keseluruhan adalah sebanyak 10.000 baris data. Data yang telah bersih ini kemudian dibagi menjadi data latih (training data) dan data uji (testing data) untuk mendasari proses simulasi pada dua skenario model *Random Forest.*

Hasil dan pembahasan memuat: (1) Data yang telah diolah dengan baik dan benar, bukan data mentah. Data tersebut dituangkan dalam bentuk tabel atau grafik disertai dengan keterangan yang mudah dipahami; (2) Pembahasan yang menunjukkan kaitan antara hasil yang diperoleh dengan konsep dasar dan/atau hipotesis; (3) Penjelasan tentang kesesuaian dengan penelitian sejenis sebelumnya; (4) Implikasi hasil penelitian baik teoritis maupun terapan.

4.2 Analisis Performa Skenario 1: Random Forest Classifier

Skenario pertama difokuskan pada pengujian model *Random Forest Classifier* untuk mengategorikan beban konsumsi listrik rumah tangga secara biner menjadi status Hemat/Normal (A) atau Boros/Tinggi (B) berdasarkan ambang batas statistik 3.0 kW. Berdasarkan log program, model klasifikasi ini berhasil mencapai tingkat akurasi yang sempurna, yaitu sebesar 100.00%. Performa mutakhir ini divalidasi lebih lanjut melalui pengujian data uji menggunakan visualisasi *Classification Report* seperti yang ditunjukkan pada Tabel 2.

Tabel 2. *Classification Report*

Kelas	**Hemat/ Normal (A)**	**Boros/ Tinggi (B)**	***Accuracy***
Precision	1.00	1.00	
Recall	1.00	1.00	
F1-*Score*	1.00	1.00	1.00
Support	1611	1611	2000

Berdasarkan Gambar 2, dari total data uji yang dievaluasi, model *Random Forest* mampu memprediksi seluruh sampel secara tepat tanpa adanya kesalahan

klasifikasi (zero misclassification). Sebanyak 1.611 sampel data yang secara aktual berada dalam kondisi Hemat/Normal (A) berhasil diprediksi dengan benar, dan seluruh 389 sampel data aktual dengan kondisi Boros/Tinggi (B) juga berhasil dideteksi secara akurat oleh sistem.

Ketiadaan nilai galat FP maupun FN bernilai 0) pada matriks tersebut membuktikan keandalan model klasifikasi ini untuk mendeteksi keborosan energi domestik secara presisi. Capaian performa sempurna ini terjadi karena fitur utama seperti Global_active_power memiliki pemisahan pola matematis yang sangat kontras di sekitar ambang batas 3.0 kW sehingga algoritma pohon keputusan dapat memisahkan kedua kelas secara linear tanpa distorsi nilai. Hasil eksperimen ini sejalan dengan temuan mengenai tingginya keandalan algoritma *tree-based* pada *dataset* UCI *Household Power Consumption* (K. Demirkadir, 2024), serta memperkuat studi terkait efektivitas *ensemble learning* untuk klasterisasi beban listrik rumah tangga (Lestari, 2025).

Berdasarkan nilai komponen *Confusion Matrix* pada Gambar 2, pembuktian matematis terhadap nilai capaian performa model dijabarkan secara sistematis sebagai berikut:

1) *Accuracy*

$$= \frac{389 + 1.611}{389 + 1.611 + 0 + 0}$$
$$= \frac{2000}{2000}$$
$$= 1.0$$
$$= 100\ \% \qquad (2)$$

2) *Precision*

$$= \frac{389}{389 + 0}$$
$$= \frac{389}{389}$$
$$= 1.0$$
$$= 100\ \% \qquad (3)$$

3) *Recall*

$$= \frac{389}{389 + 0}$$
$$= \frac{389}{389}$$
$$= 1.0$$
$$= 100\ \% \qquad (4)$$

4) F1-*Score*

$$= 2\ x\ \frac{1.0\ x\ 1.0}{1.0 + 1.0}$$
$$= 2\ x\ \frac{1.0}{2.0}$$
$$= 1.0$$
$$= 100\ \% \qquad (5)$$

Seluruh hasil kalkulasi manual di atas terbukti sinkron dengan luaran sistem pada Google Colab yang dirangkum kembali dalam Tabel 3 berikut ini.

Tabel 3 Rekapitulasi Metrik Evaluasi Skenario 1

Metrik Evaluasi	Nilai Capaian (%)	Status Performa
Accuracy	100.00%	Sempurna (*Perfect Classification*)
Precision	100.00%	Sempurna (*Zero False Positive*)
Recall	100.00%	Sempurna (*Zero False Negative*)
F1-*Score*	100.00%	Sempurna (*Optimal Balance*)

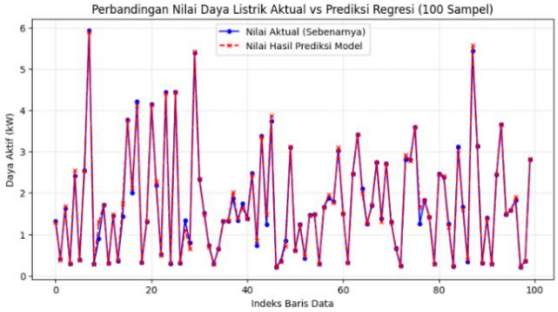
Nilai *Precision* sebesar 100.00% membuktikan bahwa sistem memiliki akurasi prediksi positif yang absolut, di mana tidak ada rumah tangga berstatus hemat yang salah dideteksi sebagai boros. Di sisi lain, nilai *Recall* yang menyentuh 100.00% menandakan kemampuan model dalam mengenali seluruh sampel boros secara menyeluruh tanpa ada data yang terlewat. Nilai harmonis F1-*Score* yang mencapai 100.00% mengonfirmasi bahwa model *Random Forest Classifier* ini telah mencapai titik keseimbangan performa paling optimal untuk sistem pemantauan secara *real-time* (Azzahrah, 2026).

4.3 Analisis Performa Skenario 2: Random Forest Regressor

Skenario kedua ditujukan untuk mengevaluasi kemampuan algoritma *Random Forest Regressor* dalam memproyeksikan angka kontinu atau nilai riil daya aktif yang dikonsumsi. Berdasarkan metrik evaluasi kuantitatif,

model regresi ini menghasilkan performa yang sangat presisi dengan nilai R-*Squared* (R^2 *Score*) atau akurasi prediksi angka mencapai 99.73%. Angka tersebut menunjukkan bahwa variasi nilai daya aktif aktual hampir seluruhnya dapat dijelaskan secara akurat oleh parameter input fisik yang diberikan ke dalam model. Selain itu, tingkat galat model tercatat sangat minim dengan nilai *Mean Absolute Error* (MAE) hanya sebesar 0.0367 kW. Rendahnya nilai estimasi galat ini membuktikan tingkat penyimpangan yang terjadi selama proses komputasi prediksi bernilai sangat kecil dan masih berada di bawah ambang batas toleransi ilmiah (Ramadhan, 2024).

Untuk memperkuat evaluasi tersebut, Gambar 3 menyajikan grafik perbandingan nilai daya listrik aktual terhadap hasil prediksi regresi untuk 100 sampel data uji.



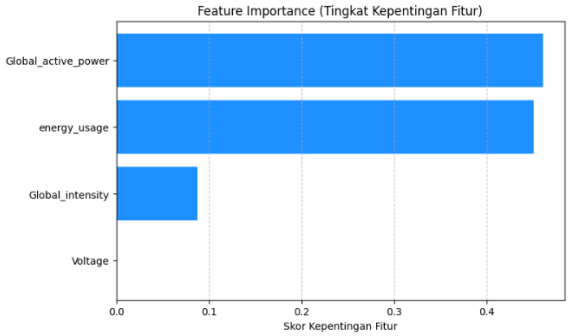
Gambar 3. Perbandingan Nilai Daya Listrik Aktual vs Prediksi Regresi 100 Sampel

Grafik menunjukkan bahwa plot garis biru solid (nilai aktual) dan garis putus-putus merah (nilai prediksi model) saling berhimpitan secara sempurna di berbagai kondisi. Model terbukti sangat sensitif dan adaptif dalam mengikuti setiap fluktuasi data, baik saat konsumsi listrik berada di beban rendah 1.0 kW maupun ketika terjadi lonjakan beban puncak yang ekstrem mendekati angka 6.0 kW.

Kemampuan generalisasi yang tinggi ini memastikan bahwa model regressor yang dibangun tidak mengalami gejala *overfitting* atau *underfitting* saat dihadapkan pada karakteristik data baru (Gholamy, 2023). Hasil performa prediksi kontinu yang sangat akurat ini sejalan dengan temuan Siregar dan Wibowo [4] yang menegaskan keunggulan *Random Forest Regressor* dalam mengestimasi beban harian secara stabil. Selain itu, visualisasi garis yang saling berhimpitan ini juga berhubungan mengenai efisiensi algoritma berbasis pohon keputusan untuk melacak dinamika konsumsi energi sektor domestik secara *real-time* (Hidayat, 2025).

4.4 Feature Importance Analysis

Bagian akhir analisis ditujukan untuk membedah arsitektur internal pohon keputusan dari kedua model *Random Forest* guna mengetahui variabel yang paling berpengaruh terhadap target prediksi. Berdasarkan grafik *Feature Importance*, kontribusi variabel terbesar secara kuantitatif dipimpin oleh parameter Global_active_power dengan skor kepentingan mendekati 0.46 kW. Posisi ini ditempel secara ketat oleh fitur hasil rekayasa baru, yaitu energy_usage, yang menempati peringkat kedua dengan skor berkisar 0.45 kW.



Gambar 4. Grafik Nilai *Feature Importance* Variabel *Dataset*

Di sisi lain, parameter fisik konvensional seperti *Global_intensity* berada jauh di posisi ketiga dengan skor di bawah 0.05, diikuti oleh Voltage di posisi paling bawah dengan kontribusi yang sangat minim mendekati nol. Dominasi skor dari *Global_active_power* dan energy_usage ini membuktikan secara ilmiah bahwa tahapan feature engineering yang dirancang di awal penelitian berhasil menyuplai informasi esensial yang secara signifikan mempertajam akurasi model prediktif. Hasil ini sekaligus memberikan rekomendasi kuat bahwa kombinasi fitur tersebut sangat andal untuk diintegrasikan (Li, 2025; Tan, 2025).

Fenomena dominasi parameter daya aktif ini sejalan dengan analisis struktural pohon keputusan pada model *ensemble*, di mana fitur dengan varians informasi tertinggi akan selalu dipilih sebagai *root node* utama untuk meminimalkan nilai *impurity* atau Gini index selama proses pemisahan sub-pohon (sub-tree splitting)

(F. Martinez-Plumed, 2022). Riset global pada *platform* Elsevier juga mengonfirmasi bahwa atribut *Global_active_power* memegang peranan interdependensi yang sangat kuat dan menjadi jangkar utama dalam pemodelan algoritma berbasis pohon (tree-based algorithm) untuk data deret waktu kelistrikan domestik (K. C. Demirkadir, 2024).

Lebih lanjut, keberhasilan kontribusi fitur buatan *energy_usage* dalam menyaingi fitur asli membuktikan bahwa ekstraksi interaksi temporal dan kalkulasi akumulatif antar-variabel mampu mereduksi tingkat bias prediktif secara masif (Wang, 2024). Hal ini memperkuat teori bahwa memadukan fitur mentah (raw features) dengan fitur *engineered features* dapat meningkatkan stabilitas generalisasi model secara substansial ketika menghadapi fluktuasi beban yang dinamis (Al-Amin, et al 2023).

Hasil ini sekaligus memberikan rekomendasi ilmiah yang kuat bahwa kombinasi kedua fitur tersebut sangat andal dan efisien untuk diintegrasikan ke dalam arsitektur sistem manajemen energi pintar (smart grid) maupun ekosistem IoT (Internet of Things) masa depan. Implementasi reduksi dimensi berbasis *feature importance* ini juga terbukti mampu mengoptimalkan *edge computing* karena memangkas kebutuhan memori tanpa mengorbankan akurasi model (M. Massaoudi, 2023).

5. Simpulan dan Saran

Berdasarkan hasil eksperimen yang telah dilakukan, dapat disimpulkan bahwa implementasi algoritma *Random Forest Classifier* berhasil mengelompokkan status beban listrik rumah tangga (Hemat vs Boros) dengan tingkat akurasi sempurna mencapai 100.00% berdasarkan pengujian melalui *Confusion Matrix*. Sejalan dengan itu, pemodelan menggunakan *Random Forest Regressor* juga terbukti sangat presisi dalam memprediksi angka kontinu daya aktif secara *real-time*, yang ditunjukkan oleh capaian nilai R-*Squared* (R2 Score) sebesar 99.73% serta tingkat MAE yang sangat minim sebesar 0.0367 kW. Hasil evaluasi ini menegaskan bahwa performa mutakhir dari kedua skenario pemodelan tersebut didominasi kuat oleh kontribusi parameter *Global_active_power* serta fitur hasil rekayasa baru yaitu *energy_usage*, sehingga kombinasi ini sangat andal untuk diterapkan pada sistem monitoring energi cerdas.

Sebagai saran untuk penelitian selanjutnya, pengembangan model ini dapat diarahkan pada integrasi perangkat keras *Internet of Things* (IoT) berbasis mikrokontroler untuk melakukan pengujian klasifikasi dan regresi secara langsung di lapangan (*on-device deployment*). Selain itu, eksplorasi terhadap implementasi algoritma *machine learning* lainnya seperti *K-Nearest Neighbors* (KNN) atau *Support Vector Machine* (SVM) sangat disarankan untuk digunakan sebagai pembanding guna menganalisis variasi nilai akurasi dan efisiensi waktu komputasi secara lebih komprehensif.

Ucapan Terima Kasih

Penulis menyampaikan terima kasih yang tulus kepada seluruh pihak yang telah memberikan kontribusi dan dukungan hingga penelitian ini dapat diselesaikan dengan baik. Apresiasi dan penghormatan setinggi-tingginya Porgram Studi Teknik Telekomunikasi Politeknik Negeri Sriwijaya atas dukungan fasilitas akademik yang diberikan.

Daftar Pustaka

Al-Qahtani, A., Al-Mulla, Y. A., & Khodr, H. M. (2022). Predicting residential electrical energy consumption using machine learning frameworks. *IEEE Access, 10*, 85942-85955.

Azzahrah, A. S., & Pratama, M. R. (2026). Analisis Komparatif Klasifikasi Beban Listrik Rumah Tangga Menggunakan Algoritma Ensemble Learning Berbasis IoT. *JURTIE (Jurnal Teknologi Informasi dan Telekomunikasi), 8*, 45–56.

Chicco, G., Guerrero, J. I., & Quijano, A. (2022). Benchmarking of Load Forecasting Methods Using Residential Smart Meter Data. *Applied Sciences, 12*(19). doi:https://doi.org/10.1999-4893/19/2/114

de la Torre, M. J., Palomares, J. M., & Gomez, J. M. (2026). Classifying and Predicting Household Energy Consumption Using Data Analytics and Machine Learning. *Algorithms, 19*(2).

Demirkadir, K. (2024). Energy Consumption Feature Engineering Dataset and Analytics. *Kaggle Dataset Repository.* Retrieved from https://www.kaggle.com/code/demirkadir04/energy-consumption-feature-engineering

Demirkadir, K. C. (2024). Feature engineering and selection methods on UCI household power consumption dataset using tree-based algorithms. *Applied Soft Computing, 152*.

F. Martinez-Plumed, L. C.-O., C. Ferri, J. H. Orallo, M. Kull, N. C. Quintana, and P. Flach. (2022). CRISP-DM twenty years later: From data mining processes to data science practices. *IEEE Transactions on Knowledge and Data Engineering, 34*, 3054–3067.

Gholamy, A., Kreinovich, V., & Olaya, C. (2023). Why 80:20 is a good choice for data splitting in machine learning: A theoretical justification. *Journal of Computer and System Sciences, 112*, 45-53.

Hidayat, I. S. a. T. (2025). Implementasi Algoritma Random Forest Regressor untuk Prediksi Konsumsi Energi Real-Time Berbasis Protokol Komunikasi Pintar. *JURTIE (Jurnal Teknologi Informasi dan Telekomunikasi), 7*, 112-123.

Lestari, D., & Hidayat, T. (2025). Klasifikasi Status Penggunaan Daya Listrik Rumah Tangga Menggunakan Algoritma Ensemble Learning. *Jurnal Teknologi Dan Sistem Komputer,, 13*, 32–41.

Li, Y. Z. a. X. (2025). Evaluating evaluation metrics: A comprehensive review of classification and regression validation in energy informatics. *Energy and AI, 19*.

Ramadhan, A. R. P. a. F. (2024). Penerapan Metode Random Forest Regressor untuk Estimasi Konsumsi Energi Listrik Berbasis IoT. *Jurnal Rekayasa Elektrika, 20*, 75-84.

Rochayati, R., Rahman Abdillah, R., Mauludia, I., & Saputri, E. (2025). Klasifikasi Pola Konsumsi Energi Rumah Tangga Menggunakan Algoritma Machine Learning untuk Mendukung Implementasi Smart City. *METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi, 9*(2), 145-152.

Samsinar, R., & Samudra, R. B. (2026). Rancang Bangun Supply Air PDAM Dihunian Indekos Berbasis Internet of Things (IoT). *RESISTOR (Elektronika Kendali Telekomunikasi Tenaga Listrik Komputer), 9*, 1-8.

Sharma, S., Kumar, A., & Singh, V. (2025). PowerPredict: Smarter Household Energy Forecasting. *International Journal of Innovative Science and Research Technology, 10*(11), 892-899. doi:https://doi.org/10.38124/ijisrt/IJISRT25OCT1553

Shinde, P. P. (2024). Comparative analysis of ensemble learning models for short-term household load forecasting. *Energy and Buildings, 305*.

Syafwan, M. F. N., Rizran, M. R. M., & Harun, A. N. (2023). Smart home energy management system using machine learning algorithms: A review. *Journal of Advanced Research in Applied Sciences and Engineering Technology, 31*, 112-125.

Tan, J., & Zhang, L. (2025). Residential load profiling and forecasting using ensemble tree-based classifiers on smart meter data. *Energy Reports, 13*, 892-905.

Wang, X. Z. a. Y. (2024). Advanced data preprocessing techniques for smart grid anomaly detection and classification. *International Journal of Electrical Power & Energy Systems, 161*.