

Analisis Kinerja Boiler Pada PLTU Ombilin Berdasarkan Variabel Operasional Menggunakan Model *Machine Learning Random Forest*

Devky Meinaldi Fernanda¹, Arwizet², Yolli Fernanda³, Andre Kurniawan⁴

¹⁻⁴ Universitas Negeri Padang, Indonesia

Alamat: Jl.Prof.Dr.Hamka,Air Tawar Bar.,Kec. Padang Utara,Kota Padang,Sumatera Barat
25171,Indonesia

Corresponding Author: devkymeinaldifernanda@gmail.com

Published: August , 2025

ABSTRAK

Penelitian ini mengembangkan model prediksi beban PLTU menggunakan algoritma Random Forest dengan proses tuning hyperparameter melalui grid search untuk memperoleh konfigurasi optimal, yaitu `n_estimators` 200, `max_depth` 30, `min_samples_split` 10, dan `min_samples_leaf` 2. Model dilatih dan diuji menggunakan data historis dan data eksternal sebanyak 4.149 sampel, menghasilkan performa prediksi yang sangat baik dengan nilai MAE sebesar 0,6517, RMSE 1,4305, dan R-squared 0,9650. Hasil visualisasi scatter plot menunjukkan prediksi yang akurat dengan sebagian besar titik dekat garis ideal. Model ini membuktikan kemampuan generalisasi dan reliabilitas yang tinggi dalam memprediksi beban PLTU, sehingga dapat digunakan efektif dalam monitoring dan pengelolaan operasional pembangkit listrik.

Kata Kunci: Random Forest, Prediksi Beban, PLTU, Machine Learning, Hyperparameter Tuning, MAE, RMSE, R-squared, Generalisasi Model, Monitoring Pembangkit

© The Author(s). 2021 Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third-party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0>

PENDAHULUAN

Pembangkit Listrik Tenaga Uap (PLTU) merupakan salah satu sumber utama pembangkit listrik di Indonesia yang memberikan kontribusi signifikan terhadap pasokan energi nasional. PLTU Ombilin, dengan kapasitas 2×100 MW, berperan penting dalam menjaga keandalan sistem kelistrikan di wilayah Sumatera Barat. Namun, efisiensi operasional PLTU, khususnya pada komponen boiler, masih menjadi tantangan utama yang perlu diatasi untuk mendukung keberlanjutan dan penghematan energi. Kinerja boiler yang optimal sangat menentukan efisiensi siklus termal pembangkit, dimana kehilangan energi yang terjadi dapat menyebabkan pemborosan bahan bakar dan peningkatan emisi gas rumah kaca. Oleh karena itu, analisis dan optimalisasi kinerja boiler menjadi fokus penelitian yang dapat memberikan dampak besar pada efisiensi dan lingkungan.

Dalam era digitalisasi dan industri 4.0, penerapan teknologi machine learning telah berkembang pesat dalam bidang pengelolaan energi dan otomasi industri. Algoritma machine learning mampu mengenali pola pada data operasional yang kompleks dan memberikan prediksi atau rekomendasi yang akurat untuk meningkatkan efisiensi sistem. Khususnya, algoritma Random Forest menjadi pilihan populer karena kemampuannya dalam menangani data non-linear, mengurangi risiko overfitting, dan memberikan interpretasi yang dapat digunakan untuk analisis sensitivitas variabel.

Penelitian ini bertujuan untuk menganalisis kinerja boiler pada PLTU Ombilin berdasarkan variabel operasional menggunakan model machine learning Random Forest. Dengan mengolah data historis operasional yang telah melalui praproses dan tuning hyperparameter, diharapkan model ini dapat

memberikan prediksi yang akurat, mengidentifikasi faktor-faktor penting yang mempengaruhi efisiensi boiler, serta mendukung pengambilan keputusan yang lebih efisien dan berkelanjutan.

Nomenklatur

- a. m_{bakar} : Laju aliran massa bahan bakar (kg/s)
- b. m_{uap} : Laju aliran massa uap (kg/s)
- c. LHV: Lower Heating Value bahan bakar (Kj/kg)
- d. h_{uap} : Entalpi uap keluar (Kj/kg)
- e. h_{air} : Entalpi air keluar (Kj/kg)
- f. Q_{loss} : Kehilangan panas pada boiler (kW)
- g. MAE: Mean Absolute Error
- h. RMSE: Root Mean Square Error
- i. R^2 : Koefisien Determinasi
- j. $n_{\text{estimators}}$: Jumlah pohon dalam model Random Forest
- k. max_depth : Kedalaman maksimum pohon keputusan
- l. min_samples_split : Jumlah minimum sampel untuk membagi simpul
- m. min_samples_leaf : Jumlah minimum sampel pada daun pohon

Penelitian ini mengkaji kinerja boiler pada PLTU Ombilin melalui analisis variabel operasional menggunakan model machine learning berbasis algoritma Random Forest. Studi dilakukan dengan mengumpulkan data historis operasional yang mencakup parameter penting seperti tekanan uap, temperatur uap, laju aliran massa uap, tekanan udara dan air, serta konsumsi bahan bakar. Data ini telah dibersihkan dan dinormalisasi menggunakan metode Min-Max Scaler untuk meningkatkan performa model. Model Random Forest dikembangkan dan dioptimasi melalui tuning hyperparameter dengan pendekatan grid search, menghasilkan konfigurasi terbaik pada parameter $n_{\text{estimators}}$ sebesar 200, max_depth 30, min_samples_split 10, dan min_samples_leaf 2. Evaluasi performa model pada data pelatihan mencapai nilai Mean Absolute Error (MAE) 0,659, Root Mean Square Error (RMSE) 0,942, dan koefisien determinasi (R^2) sebesar 0,9744, yang menunjukkan tingkat akurasi prediksi yang sangat tinggi.

Pengujian pada data eksternal sebanyak 4.149 sampel mengonfirmasi kemampuan generalisasi model dengan nilai MAE 0,6517, RMSE 1,4305, dan R^2 0,9650, menunjukkan kestabilan dan keandalan model dalam menghadapi variasi data baru. Visualisasi hasil melalui scatter plot memperlihatkan bahwa sebagian besar titik prediksi berada sangat dekat dengan garis ideal, menandakan model mampu melakukan prediksi yang akurat dan dapat diandalkan dalam monitoring kinerja operasional boiler.

Analisis sensitivitas model mengidentifikasi variabel-variabel utama yang mempengaruhi efisiensi boiler, seperti tekanan dan temperatur uap, laju aliran uap dan air, serta konsumsi bahan bakar. Penerapan model ini dapat membantu operator PLTU dalam mengambil keputusan operasional yang lebih cepat, akurat, dan mendukung peningkatan efisiensi energi serta pengurangan emisi. Studi ini juga menghasilkan framework tata kelola data dan standar validasi model sesuai dengan best practice industri sebagai landasan implementasi machine learning di lingkungan pembangkit listrik berbasis boiler. Penelitian ini membuktikan bahwa integrasi teknologi machine learning dengan data operasional dapat memberikan solusi efektif untuk meningkatkan kinerja boiler PLTU, mendukung transisi energi yang lebih berkelanjutan dan pengelolaan pembangkit listrik yang lebih adaptif dan efisien.

TUJUAN PENELITIAN

1. Menganalisis kinerja boiler PLTU Ombilin berdasarkan data historis variabel operasional.
2. Mengidentifikasi variabel operasional yang paling mempengaruhi Beban yang dihasilkan.
3. Menerapkan algoritma *Random Forest* untuk memprediksi Beban yang dihasilkan
4. Mengidentifikasi faktor utama yang mempengaruhi perfoma boiler berdasarkan analisis *Random Forest*.

METODE PENELITIAN

Metode penelitian eksperimen adalah metode penelitian yang dilakukan dengan percobaan, yang merupakan metode kuantitatif, digunakan untuk mengetahui pengaruh variable independent (treatment/perlakuan) terhadap variable dependen (hasil) dalam kondisi yang terkendalikan. Kondisi dikendalikan agar tidak ada variable lain (selain variable treatment) yang mempengaruhi variable dependen. Agar kondisi dapat dikendalikan, maka dalam penelitian eksperimen menggunakan kelompok control. Penelitian ini menggunakan pendekatan machine learning berbasis algoritma Random Forest untuk menganalisis kinerja boiler pada PLTU Ombilin berdasarkan variabel operasional. Data historis operasional yang relevan dikumpulkan, kemudian melalui proses pembersihan termasuk penghapusan data duplikat, imputasi nilai hilang, dan identifikasi outlier. Data yang telah bersih selanjutnya dinormalisasi menggunakan metode Min-Max Scaler untuk menyamakan skala variabel dalam rentang 0 hingga 1, yang bertujuan meningkatkan stabilitas dan performa model. Hyperparameter Random Forest dituning menggunakan grid search dengan cross-validation untuk menemukan kombinasi optimal yang meminimalkan kesalahan prediksi dan mencegah overfitting.

Studi ini terdiri dari beberapa tahap utama, yakni pengumpulan dan persiapan data, pembangunan dan optimasi model Random Forest, evaluasi model menggunakan metrik kinerja seperti Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan koefisien determinasi (R-squared), serta analisis sensitivitas terhadap variabel operasional boiler. Visualisasi hasil prediksi dilakukan untuk memvalidasi kemampuan model dalam merepresentasikan data aktual. Selanjutnya, model diuji pada dataset eksternal untuk menguji kemampuan generalisasi dan kehandalan prediksi dalam kondisi nyata.

Analisis statistik deskriptif diterapkan pada variabel operasional untuk memahami karakteristik seperti rentang nilai minimum, maksimum, dan nilai rata-rata. Indeks Gini digunakan dalam proses pemisahan atribut untuk membangun pohon keputusan dalam Random Forest, dimana impurity dan perbedaan nilai Gini dihitung sebagai dasar pemilihan split terbaik. Korelasi antar variabel dianalisis untuk mengidentifikasi hubungan yang signifikan antara parameter operasional dan efisiensi boiler. Metrik evaluasi prediksi seperti MAE, RMSE, dan R-squared digunakan untuk mengukur akurasi dan kemampuan model dalam menjelaskan variasi data beban. Pendekatan ini memastikan model bukan hanya akurat pada data pelatihan, tetapi juga dapat diaplikasikan secara efektif pada data baru di lapangan.

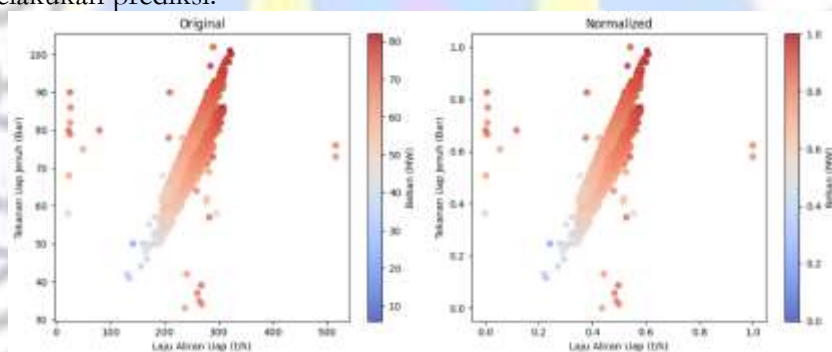
HASIL DAN PEMBAHASAN

Data untuk penelitian ini diambil dari ruang kontrol boiler di Pembangkit Listrik Tenaga Uap (PLTU) Ombilin, di mana operator secara rutin mengumpulkan data operasional harian sebagai laporan. Variabel yang dicatat meliputi arus motor fan (Ampere), tekanan uap jenuh (Bar), suhu uap utama ($^{\circ}\text{C}$), tekanan uap utama (Bar), laju aliran uap (t/h), jumlah air tambahan (t), aliran air (t/h), tekanan air umpan (Bar), total batu bara (t/h), jumlah udara pembakaran (Nm^3/h), suhu udara masuk ($^{\circ}\text{C}$), suhu gas buang ($^{\circ}\text{C}$), dan beban (MW). Pemilihan variabel ini berdasarkan literatur yang menyatakan bahwa parameter-parameter tersebut berperan penting dalam kinerja dan efisiensi boiler. Tekanan dan suhu uap utama sangat mempengaruhi efisiensi termal, sementara laju aliran uap dan aliran air terkait langsung dengan kebutuhan energi dan output daya. Arus motor fan dan jumlah udara pembakaran menunjukkan efisiensi pembakaran dalam ruang bakar, dan suhu gas buang digunakan untuk menilai kehilangan energi melalui cerobong. Beban dalam megawatt dipilih sebagai variabel target karena mencerminkan performa keseluruhan sistem pembangkit. Dengan total 13 variabel, penelitian ini bertujuan menganalisis secara komprehensif dinamika operasional boiler di PLTU Ombilin.

Variabel	Type	Min	Max	Mean
Arus motor Fan (Amp)	Input	28	46	42
Tekanan Uap Jenuh (Bar)	Input	33	102	76,5
Suhu Uap Utama (Deg C)	Input	310	747	508,19
Tekanan Uap Utama (Bar)	Input	35	92	68,29
Laju Aliran Uap (t/h)	Input	22	516	259,47
Jumlah Air Tambahan (t)	Input	0,02	94	6,43
Aliran Air (t/h)	Input	25	666	251,2
Tekanan Air Umpan (Bar)	Input	20	102	77,5
Total Batu Bara (t/h)	Input	11,43	56,83	46,67
Jumlah Udara Pembakaran (Nm3/h)	Input	152	422	334,2
Suhu Udara Masuk(deg C)	Input	202	359	324,95
Suhu gas buang (deg C)	Input	53	252	152,01
Beban (MW)	Output	6	82	65,03

Tabel 1. Variabel Boiler PLTU Ombilin

Data boiler dinormalisasi menggunakan metode Min-Max Scaler untuk mengubah nilai-nilai dataset menjadi rentang seragam antara 0 dan 1. Normalisasi ini umum diterapkan dalam prediksi beban (MW) karena dapat meningkatkan kinerja model machine learning dengan mencegah masalah gradien yang besar akibat perbedaan skala variabel. Selain itu, normalisasi memastikan bahwa model tidak hanya memprioritaskan variabel dengan skala besar, tetapi juga memberikan perhatian seimbang pada variabel dengan skala kecil, sehingga secara keseluruhan meningkatkan efektivitas model dalam melakukan prediksi.

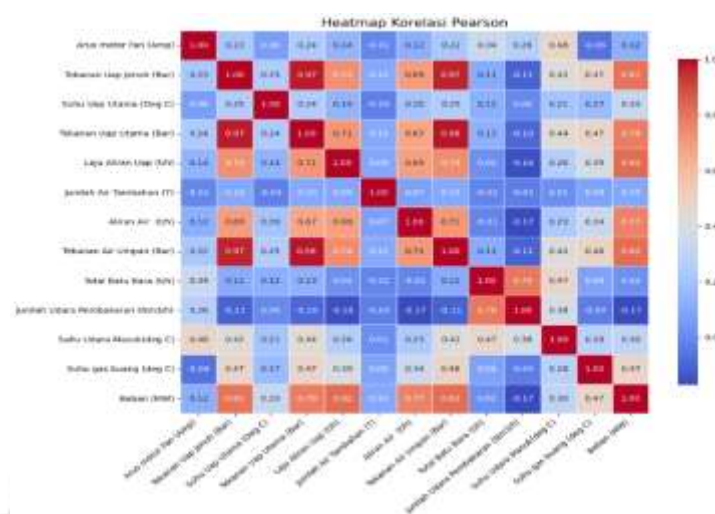


Gambar 1. Ilustrasi normalisasi dengan metode Min-MaxScaler

Gambar scatter plot berikut memperlihatkan perbandingan data sebelum dan sesudah normalisasi. Pada grafik sebelah kiri (Original), variabel laju aliran uap (t/h) diplot terhadap tekanan uap jenuh (Bar) dengan nilai beban (MW) ditampilkan menggunakan warna, di mana sebagian besar data berada pada laju aliran uap 200–350 t/h dan tekanan uap jenuh 60–100 Bar, serta menunjukkan kecenderungan peningkatan beban yang sejalan dengan kenaikan kedua variabel tersebut. Setelah normalisasi (grafik kanan), nilai variabel dikonversi ke rentang 0 hingga 1, namun pola distribusi antar variabel tetap terjaga, menandakan struktur hubungan yang konsisten. Normalisasi ini sangat penting terutama untuk algoritma machine learning yang sensitif terhadap skala data, seperti k-Nearest Neighbor, Support Vector Machine, dan Neural Networks, karena memungkinkan model mengenali pola dengan lebih optimal tanpa dominasi variabel tertentu akibat perbedaan satuan atau skala.

Sebelum melatih model, dilakukan analisis korelasi untuk mengetahui hubungan antara beban yang dihasilkan dan parameter-parameter boiler. Analisis korelasi ini menggunakan metode Pearson, sebuah teknik statistik yang mengukur sejauh mana dua variabel memiliki hubungan linier satu sama lain dengan menghasilkan koefisien korelasi (r) yang berkisar antara -1 hingga 1. Nilai r yang mendekati 1 menunjukkan korelasi positif sempurna, artinya kedua variabel meningkat secara bersamaan. Nilai r yang mendekati -1 menunjukkan korelasi negatif sempurna, di mana kenaikan

pada satu variabel diikuti oleh penurunan pada variabel lainnya. Sedangkan nilai r mendekati 0 menunjukkan tidak adanya hubungan linier antara kedua variabel.



Gambar 2. Heatmap Korelasi Person Beban (MW) dengan parameter Boiler

Heatmap korelasi Pearson menggambarkan hubungan antar variabel operasional pada sistem pembangkit listrik PLTU, di mana beberapa variabel menunjukkan korelasi kuat terhadap beban (MW) seperti tekanan uap jenuh ($r = 0,81$), tekanan uap utama ($r = 0,78$), laju aliran uap ($r = 0,82$), tekanan air umpan ($r = 0,82$), dan aliran air ($r = 0,77$). Korelasi ini mengindikasikan bahwa peningkatan tekanan dan aliran fluida dalam sistem berkontribusi pada peningkatan energi uap dan daya listrik yang dihasilkan, sesuai dengan prinsip termodinamika siklus Rankine. Sebaliknya, jumlah udara pembakaran memiliki korelasi negatif terhadap beban ($-0,17$), yang menjelaskan bahwa peningkatan jumlah udara secara berlebihan dapat menyebabkan heat loss melalui gas buang sehingga menurunkan efisiensi boiler dan output energi efektif. Oleh karena itu, pengaturan jumlah udara pembakaran perlu dioptimalkan agar efisiensi tetap tinggi dan beban stabil.

Model machine learning dalam penelitian ini dibangun menggunakan algoritma Random Forest Regressor, sebuah metode ensemble learning yang menggabungkan banyak pohon keputusan untuk mengolah data non-linear dan mengurangi risiko overfitting. Model ini dilatih menggunakan data operasional PLTU yang mencakup tekanan, suhu, aliran fluida, dan parameter pembakaran untuk memprediksi beban pembangkitan (MW). Proses pengembangan model dilakukan secara sistematis, mulai dari pemilihan fitur, normalisasi data, pembagian data menjadi set latihan dan uji, hingga tuning hyperparameter. Arsitektur Random Forest terdiri dari beberapa pohon keputusan yang dibangun secara acak dari subset data dan fitur, di mana prediksi akhir merupakan rata-rata dari seluruh hasil pohon tersebut (bagging).

Metode	Nama Parameter	Pengaturan Parameter
DT	Maximal Depth	1,2,3,4,5,6,7,8,9,10,15,20,25,30
RF	Number of tree	20,50,80,110,140,150
	Maximal Depth	5,10,15,25,30
ANN	Training Cycle	10.000,20.000,30.000, 40.000
	Learning Rate	0,001, sd 0,1
	Momentum	0,1 sd 1
	Neuron Hidden Layer 1	5, 10, 15, 20, 25, 30, 40, 50
	Neuron Hidden Layer 2	5, 10, 15, 20, 25, 30, 40, 50

Tabel 2. Nilai Hyperparameter

Dalam pembangunan model, akan dilakukan proses tuning hyperparameter untuk meningkatkan akurasi dan efisiensi model, di mana hasil dari kombinasi tuning ini akan dianalisis guna menentukan konfigurasi terbaik meskipun hasil tersebut belum disajikan pada tahap ini. Hyperparameter yang merupakan parameter eksternal yang ditetapkan sebelum pelatihan memegang peran penting dalam menentukan struktur dan perilaku model, sehingga penelitian ini akan mengeksplorasi berbagai kombinasi nilai hyperparameter dengan tujuan mencegah overfitting melalui evaluasi menggunakan metrik R-squared, MAE, dan RMSE.

No	N_ estimators	Max_ depth	Min_ samples_split	Min_ samples_ leaf	MAE	RMSE	R- squared
1	100	None	2	1	0,657	1,012	0,970
2	200	None	2	1	0,657	0,995	0,971
3	300	None	2	1	0,657	1,004	0,970
4	400	None	2	1	0,658	1,009	0,970

Tabel 3. Perbandingan Matriks model dengan variasi *n_estimators*

No	N_ Estimators	Max_ Depth	Min_ samples_ split	Min_ samples_ leaf	MAE	RMSE	R- Squared
1	200	None	2	1	0,657	0,995	0,971
2	200	10	2	1	0,692	1,023	0,969
3	200	20	2	1	0,657	1,001	0,9712
4	200	30	2	1	0,658	0,998	0,9714

Tabel 4. Perbandingan Matriks model dengan variasi *max_depth*

No	N_ estimators	Max_ depth	Min_ samples_ split	Min_ samples_ leaf	MAE	RMSE	R- squared
1	200	30	2	1	0,658	0,998	0,971
2	200	30	5	1	0,656	0,978	0,9724
3	200	30	10	1	0,667	0,964	0,9732
4	200	30	17	1	0,687	0,970	0,9728

Tabel 5. Perbandingan Matriks model dengan variasi *Min_samples_split*

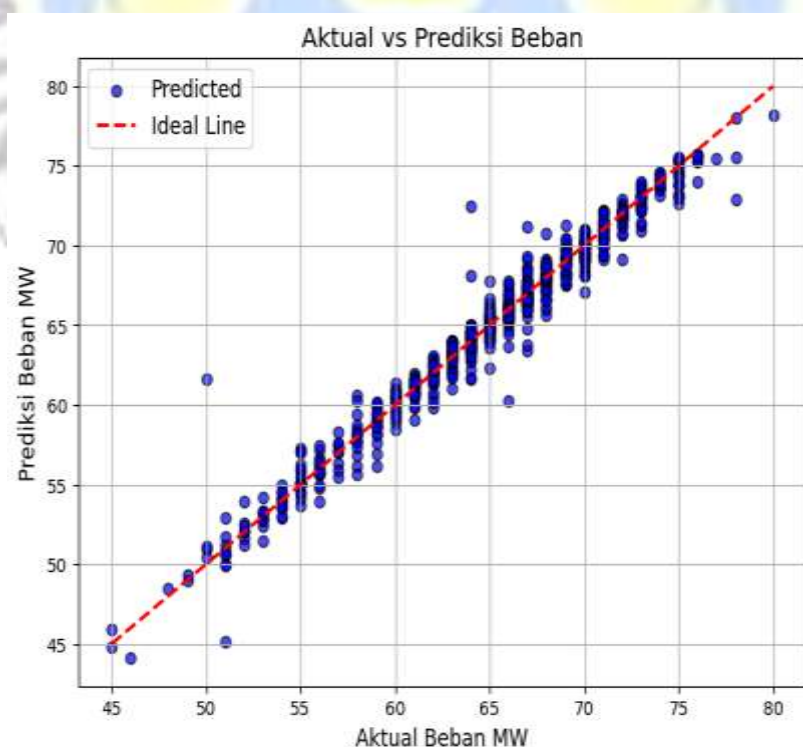
No	N_ estimators	Max_ depth	Min_ samples_ split	Min_ samples_ leaf	MAE	RMSE	R- squared
1	200	30	10	1	0,657	0,964	0,9732
2	200	30	10	2	0,659	0,942	0,9744
3	200	30	10	4	0,665	0,946	0,9742
4	200	30	10	6	0,683	0,973	0,972

Tabel 6. Perbandingan Matriks model dengan variasi *min_samples_leaf*

Hasil tuning menunjukkan bahwa nilai MAE tetap stabil di rentang 0,657–0,658 yang menandakan kesalahan absolut rata-rata model relatif konsisten di seluruh variasi jumlah pohon, dengan RMSE antara 0,995 hingga 1,012 dan R-squared berkisar pada 0,970 hingga 0,971. Konfigurasi dengan

$n_estimators = 200$ memberikan performa terbaik dengan MAE 0,657, RMSE terendah 0,995, dan R^2 tertinggi 0,971. Setelah menetapkan $n_estimators$ 200 sebagai nilai optimal, tuning dilanjutkan pada parameter max_depth yang mengatur kedalaman maksimum pohon keputusan, penting untuk mengendalikan kompleksitas model dan mencegah overfitting sambil menangkap pola data. Pengujian variasi max_depth dengan $min_samples_split = 2$ dan $min_samples_leaf = 1$ menunjukkan bahwa max_depth 30 memberikan hasil terbaik dengan R^2 0,9714 dan RMSE 0,998, meskipun perbedaan performa antar variasi kecil, sehingga max_depth 30 dianggap sebagai konfigurasi optimal berikutnya. Selanjutnya, tuning $min_samples_split$ diuji pada nilai 2, 5, 10, dan 17 dengan $min_samples_leaf$ tetap 1, dimana nilai 10 memberikan performa terbaik (R^2 0,9732, RMSE 0,964, MAE 0,667), meskipun MAE sedikit lebih tinggi dibanding 5, namun peningkatan R^2 menunjukkan kemampuan model menjelaskan variasi data dengan lebih baik sehingga nilai 10 dipilih sebagai kompromi terbaik antara kesalahan prediksi dan generalisasi. Tuning terakhir pada $min_samples_leaf$ dengan pengujian nilai 1, 2, 4, dan 6 menunjukkan $min_samples_leaf = 2$ sebagai konfigurasi terbaik dengan R^2 0,9744, RMSE 0,942, dan MAE 0,659, memberikan keseimbangan optimal antara akurasi dan kemampuan generalisasi model serta mengurangi risiko overfitting akibat data dengan sampel kecil.

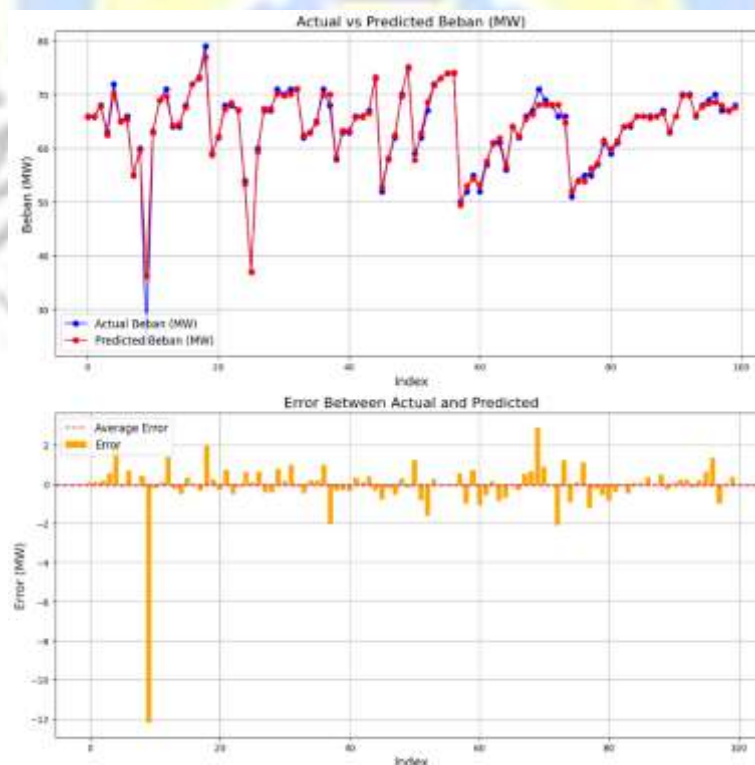
Setelah melakukan tuning hyperparameter dengan metode grid search, ditemukan kombinasi parameter terbaik untuk model Random Forest Regressor, yaitu $n_estimators$ 200, max_depth 30, $min_samples_split$ 10, dan $min_samples_leaf$ 2, dengan $random_state$ 42 untuk memastikan konsistensi hasil. Model kemudian dilatih menggunakan data historis beban PLTU dan dievaluasi performanya pada data uji, menghasilkan nilai Mean Absolute Error (MAE) 0,659, RMSE 0,942, serta koefisien determinasi (R-squared) 0,9744 yang menunjukkan kemampuan model menjelaskan 97,44% variasi data beban aktual dengan akurasi tinggi. Untuk memperkuat analisis, dilakukan visualisasi scatter plot antara nilai aktual dan prediksi beban, yang memperlihatkan mayoritas titik prediksi sangat dekat dengan garis ideal ($y = x$), sehingga dapat disimpulkan bahwa model Random Forest yang dioptimasi ini mampu memberikan prediksi beban yang akurat dan handal untuk keperluan monitoring maupun peramalan performa operasional PLTU.



Gambar 3. Visualisasi grafik latihan model aktual vs beban

Gambar tersebut memperlihatkan grafik scatter plot yang memetakan nilai beban aktual pada sumbu X dan beban hasil prediksi model Random Forest pada sumbu Y, di mana titik-titik biru mewakili hasil prediksi model terhadap data uji, dan garis merah putus-putus menunjukkan garis ideal ($y = x$) yang mencerminkan kondisi prediksi sempurna. Sebagian besar titik prediksi berada sangat dekat dengan garis ideal, menandakan performa model yang sangat baik dalam memprediksi beban, meskipun terdapat beberapa outlier terutama pada beban di atas 70 MW dan di bawah 50 MW. Penyebaran titik yang simetris terhadap garis ideal juga menunjukkan tidak adanya bias signifikan baik overestimasi maupun underestimasi secara konsisten. Evaluasi model sebelumnya mendukung temuan ini dengan nilai MAE 0,659, RMSE 0,942, dan R-squared 0,9744, yang mengindikasikan model mampu menjelaskan 97,44% variasi beban aktual. Dapat disimpulkan bahwa model Random Forest yang telah dioptimasi ini memberikan prediksi beban PLTU yang akurat dan layak digunakan sebagai dasar dalam pengambilan keputusan operasional pembangkit listrik.

Setelah model Random Forest dilatih dengan kombinasi hyperparameter optimal yaitu `n_estimators` 200, `max_depth` 30, `min_samples_split` 10, `min_samples_leaf` 2, dan `random_state` 42, model diuji menggunakan data eksternal sebanyak 4.149 sampel untuk mengevaluasi kemampuan generalisasinya. Evaluasi menghasilkan nilai Mean Absolute Error (MAE) 0,6517, Root Mean Squared Error (RMSE) 1,4305, dan koefisien determinasi (R-squared) 0,9650, yang menunjukkan rata-rata kesalahan absolut dan kesalahan kuadrat relatif kecil serta kestabilan prediksi terhadap outlier, sementara nilai R^2 menunjukkan bahwa model mampu menjelaskan sekitar 96,5% variasi beban aktual, menandakan kemampuan prediktif yang sangat baik dan dapat diandalkan. Hasil ini konsisten dengan penelitian terdahulu, seperti studi Blackburn et al. (2022) yang menggunakan optimasi machine learning GRU dan mencapai peningkatan efisiensi termal boiler dengan kesalahan prediksi rendah, serta penelitian Qiao et al. (2020) yang menunjukkan performa tinggi model Random Forest dengan R^2 di atas 0,90 dalam memodelkan sistem pembakaran boiler. Oleh karena itu, temuan penelitian ini menguatkan bukti bahwa algoritma Random Forest efektif dalam memodelkan hubungan antara parameter operasional boiler dan output beban listrik pada PLTU.



Gambar 4. Visualisasi pengujian model aktual vs beban

Grafik yang ditampilkan memperlihatkan perbandingan antara beban aktual dan beban hasil prediksi model Random Forest beserta distribusi errornya. Pada grafik atas, garis prediksi merah mengikuti

pola beban aktual yang berwarna biru, menunjukkan bahwa model mampu memetakan perubahan beban dengan baik, ditandai oleh kedekatan titik-titik prediksi dengan nilai aktual yang mencerminkan akurasi tinggi. Di bagian bawah, distribusi error dalam satuan MW sebagian besar berkisar di sekitar nol, menandakan prediksi yang cukup akurat meskipun terdapat beberapa outlier dengan selisih error besar terutama pada indeks ke-12. Secara keseluruhan, error tersebar merata dan kecil, mendukung hasil evaluasi metrik yang menunjukkan performa model yang baik.

KESIMPULAN

Penelitian ini berhasil menganalisis kinerja boiler dengan memanfaatkan data historis dari 13 variabel operasional utama, termasuk tekanan dan suhu uap, laju aliran air dan uap, serta jumlah bahan bakar dan udara pembakaran. Hasil analisis menunjukkan bahwa variasi pada parameter-parameter tersebut secara langsung memengaruhi output beban listrik yang dihasilkan oleh sistem pembangkitan. Analisis korelasi Pearson mengidentifikasi variabel yang paling berpengaruh terhadap beban listrik, yaitu laju aliran uap, tekanan air umpan, tekanan uap jenuh, dan aliran air, dengan nilai korelasi masing-masing berkisar antara 0,77 hingga 0,82. Sebaliknya, jumlah udara pembakaran menunjukkan korelasi negatif terhadap beban, yang menandakan bahwa kelebihan udara dapat menurunkan efisiensi pembakaran dan performa boiler secara keseluruhan.

Algoritma Random Forest Regressor diterapkan untuk memprediksi beban listrik berdasarkan parameter operasional boiler. Model yang telah dioptimasi dengan kombinasi hyperparameter terbaik, yakni `n_estimators` 200, `max_depth` 30, `min_samples_split` 10, dan `min_samples_leaf` 2, menunjukkan performa prediksi yang sangat baik dengan nilai evaluasi pada data uji mencapai MAE sebesar 0,659, RMSE sebesar 0,942, dan koefisien determinasi (R^2) sebesar 0,9744. Selain itu, pengujian pada data eksternal sebanyak 4.149 sampel mengonfirmasi kemampuan generalisasi model ini dengan menghasilkan MAE 0,6517, RMSE 1,4305, dan R^2 0,9650, yang mengindikasikan stabilitas dan kehandalan prediksi meskipun menghadapi data baru. Dengan demikian, model Random Forest yang dikembangkan dalam penelitian ini terbukti efektif dalam mengidentifikasi pola-pola kompleks antara variabel operasional boiler dan performa output, sehingga dapat dijadikan alat bantu yang andal untuk pemantauan dan prediksi beban dalam pengambilan keputusan operasional di PLTU Ombilin.

SARAN

Sebagai rekomendasi bagi pengelola PLTU Ombilin, sangat disarankan untuk mempertimbangkan penerapan algoritma machine learning, khususnya Random Forest, dalam sistem monitoring prediktif. Beban dapat diperkirakan secara real-time berdasarkan parameter operasional, yang akan membantu meningkatkan efisiensi operasional serta pengambilan keputusan. Selain itu, perlu diberikan perhatian khusus terhadap pengaturan jumlah udara pembakaran yang menunjukkan korelasi negatif terhadap beban, sehingga pengaturan ini hendaknya dilakukan secara optimal dengan mempertimbangkan rasio stoikiometri yang tepat, guna menghindari penurunan efisiensi akibat kelebihan udara. Untuk pengembangan penelitian berikutnya, dianjurkan agar faktor-faktor eksternal seperti kualitas batubara, kelembaban udara, dan temperatur lingkungan juga diperhitungkan karena memiliki pengaruh terhadap performa boiler. Penelitian selanjutnya juga dapat mengeksplorasi perbandingan algoritma machine learning lainnya, seperti Gradient Boosting, XGBoost, atau metode deep learning. Terakhir, pengembangan sistem prediktif yang terintegrasi dengan model machine learning ke dalam dashboard monitoring real-time akan memberikan manfaat besar bagi para operator dan teknisi di PLTU. Sistem ini tidak hanya dapat memprediksi beban, tetapi juga memberikan peringatan dini terhadap kondisi operasi boiler yang menyimpang dari pola normal, sehingga memungkinkan tindakan preventif dilakukan dengan lebih cepat dan efektif.

REFERENSI

- Ashfania, G. A. M., Prahasto, T., Widodo, A., Warsokusumo, T., Soedarto, J., Tembalang, K., Semarang, K., Tengah, J., Gatot Subroto No Kav, J., Tim, K., Setiabudi, K., Jakarta Selatan, K., & Khusus Ibukota Jakarta, D. (2021). Penggunaan Algoritma Random Forest untuk Klasifikasi berbasis Kinerja Efisiensi Energi pada Sistem Pembangkit Daya. *Vol. 24*(Issue 3).
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Blackburn, L. D., Tuttle, J. F., Andersson, K., Hedengren, J. D., & Powell, K. M. (2022). Dynamic Machine Learning-based Optimization Algorithm to Improve Boiler Efficiency.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.
- Eastop, T. D., & McConkey, A. (1993). *Applied Thermodynamics and Engineering* (5th ed.).
- Ding, Y., & Shi, Y. (2019). Real-Time Boiler Control Optimization with Machine Learning. *arXiv*. <http://arxiv.org/abs/1903.04958>
- Eastop, T. D., & McConkey, A. (2003). *Applied Thermodynamics for Engineering Technologists*. Pearson Education.
- Efri Yendri. (2024). BOILER.
- El Wakil, M. M. (1984). *Power Plant Technology*.
- Fitriyaningsih, I., Basani, Y., & Ginting, L. M. (2018). Machine Learning: Prosperity of Rainfall, Water Discharge, and Flood with Web Application in Deli Serdang. *Jurnal Penelitian Komunikasi dan Opini Publik*, 22(2). <https://doi.org/10.33299/jpkop.22.2.1752>
- Ganapathy, V. (2003). *Industrial Boilers and Heat Recovery Steam Generators*. Marcel Dekker.
- Ganapathy, V. (2011). Boiler Efficiency Improvement. *Journal of Energy Engineering*, 137(2), 56–63.
- Grochowalski, J., Jachymek, P., Andrzejczyk, M., Klajny, M., Widuch, A., Morkisz, P., Hernik, B., Zdeb, J., & Adamczyk, W. (2021). Towards application of machine learning algorithms for prediction temperature distribution within CFB boiler based on specified operating conditions. *Energy*, 237, Article 121538. <https://doi.org/10.1016/j.energy.2021.121538>
- Ho, T. K. (1998). The Random Subspace Method for Constructing Decision Forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 832–844.
- Lianzhai, D., Roestam, R., Sen, T. W., Fahmi, H., ChungKiat, O., & Hariyanto, D. T. (2024). Machine Learning Algorithms for Prediction of Boiler Steam Production. *International Journal of Advances in Data and Information Systems*, 5(2). <https://doi.org/10.59395/ijadis.v5i2.1339>
- Liparas, D., Moumtzidou, A., Vrochidis, S., & Kompatsiaris, I. (2014). Concept-oriented labelling of patent images based on Random Forests and proximity-driven generation of synthetic data.
- Pandia, M., Sihombing, P., Simamora, P., & Kaban, R. (2024). Kajian Literatur Multimedia Retrieval: Machine Learning untuk Pengenalan Wajah. *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, 7(1), 161–166. <https://doi.org/10.55338/jikoms.v7i1.2758>
- PLN. (2024). *Laporan Ketenagalistrikan Semester 1 Tahun 2024*.
- Putri, A., Syaficha Hardiana, C., Novfujia, E., Try Puspa Siregar, F., Fatma, Y., & Wahyuni, R. (2023). Comparison of K-NN, Naive Bayes and SVM Algorithms for Final-Year Student Graduation Prediction. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(1), 20–26.
- Ramadhon, R. N., Ogi, A., Agung, A. P., Putra, R., Febrihartina, S. S., & Firdaus, U. (2024). Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak Aktif pada Data Bank. *Vol. 3*.
- Soebroto, A. A. (2019). Buku Ajar AI, Machine Learning & Deep Learning. <https://www.researchgate.net/publication/348003841>
- Sugiharto, A. (2020). Perhitungan Efisiensi Boiler dengan Metode Secara Langsung pada Boiler Pipa Api.

- Syahputra, R. (2020). Teknologi Pembangkit Tenaga Listrik.
- Tirumala Srinivas, G. (2017). Efficiency of a Coal Fired Boiler in a Typical Thermal Power Plant. *American Journal of Mechanical and Industrial Engineering*, 2(1), 32. <https://doi.org/10.11648/j.ajmie.20170201.15>
- Xu, X., & Wang, R. (2013). Neurodynamics of up and down transitions in network model. *Abstract and Applied Analysis*, 2013, Article ID 486178. <https://doi.org/10.1155/2013/486178>
- Yuniardi, D., & Puspita, R. (2023). Meningkatkan Efisiensi Pengendalian Suhu Boiler pada PLTSa Burangkeng melalui Analisis Komprehensif Berbasis Machine Learning. *JTS*, 2(3).
- Zhou, H., et al. (2020). Machine Learning Applications in Power Plant Optimization, 456–469.

