

Prediksi Daya Listrik Pada Pembangkit Listrik Siklus Gabungan Berdasarkan Kondisi Lingkungan Menggunakan Metode *Machine Learning*

Hendra^{1*}

Jurusan Teknik Mesin, Politeknik Negeri Padang

*Corresponding Author Email: hendra@pnp.ac.id

Abstrak - Pemanfaatan metode machine learning dalam simulasi energi memungkinkan untuk mengoptimalkan penggunaan energi dan meningkatkan efisiensi energi. Pada penelitian ini, dilakukan pemodelan prediksi output daya listrik dengan kondisi beban penuh pada pembangkit listrik siklus gabungan (CCPP) berdasarkan kondisi lingkungan sekitar. Data historis pengoperasian CCPP digunakan untuk memodelkan dan memprediksi daya listrik dalam berbagai kondisi lingkungan. Data ini terdiri dari 9568 sampel yang dikumpulkan dari Pembangkit Listrik Siklus Gabungan (CCPP) antara tahun 2006 hingga 2011. Pada penelitian ini, Empat algoritma machine learning seperti Linear Regression (LR), Decision Tree (DT), Random Forest (RF), dan Artificial Neural Network (ANN), dibandingkan dan dievaluasi kinerjanya. Metrik evaluasi yang digunakan untuk mengukur kinerja model adalah Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan R-Squared. Hasil penelitian menunjukkan bahwa model Random Forest (RF) mencapai kinerja terbaik dibandingkan dengan model-model lainnya dengan nilai MAE 2.314, RMSE 3.372 dan R-squared 0.961. Selain itu, model RF juga memiliki kinerja paling baik dibandingkan model lainnya saat dilakukan pengujian eksternal dengan data baru, dimana RF memperoleh nilai MAE 2.579, RMSE 3.315, dan R-squared 0.957. Hasil ini tidak jauh berbeda dengan hasil pengujian sebelumnya, hasil ini mengindikasikan bahwa RF memiliki performa yang stabil dan dapat diandalkan dalam melakukan prediksi pada dataset yang lebih besar dan beragam. Keberlanjutan penelitian ini diharapkan dapat memberikan kontribusi lebih lanjut terhadap pemahaman dan penerapan machine learning dalam industri pembangkit listrik.

Kata kunci: Daya Listrik, Machine Learning, energi, pembangkit listrik

Abstract - The utilization of machine learning methods in energy simulation enables the optimization of energy usage and enhances energy efficiency. In this study, the modeling and prediction of electric power output under full load conditions in a Combined Cycle Power Plant (CCPP) are conducted based on the surrounding environmental conditions. Historical data from the operation of the CCPP are employed to model and predict power output under various environmental conditions. This dataset comprises 9568 samples collected from the Combined Cycle Power Plant (CCPP) between 2006 and 2011. In this research, four machine learning algorithms, namely Linear Regression (LR), Decision Tree (DT), Random Forest (RF), and Artificial Neural Network (ANN), are compared and evaluated for their performance. The evaluation metrics used to measure the model's performance include Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-Squared. The research results indicate that the Random Forest (RF) model achieves the best performance compared to other models, with MAE 2.314, RMSE 3.372, and R-squared 0.961. Furthermore, the RF model also outperforms other models during external testing with new data, where RF obtains MAE 2.579, RMSE 3.315, and R-squared 0.957. These results are consistent with the earlier testing, indicating that RF exhibits stable and reliable performance in predicting larger and more diverse datasets. The sustainability of this research is expected to contribute further to the understanding and application of machine learning in the power generation industry.

Keywords: *Electric Power, Machine Learning, Energy, Power Generation.*

© 2023 Elektron Jurnal Ilmiah

I. PENDAHULUAN

Pembangkit Listrik Tenaga Siklus Gabungan (Combined Cycle Power Plant atau CCPP) merupakan pembangkit listrik yang menggunakan kombinasi turbin gas (GT) dan turbin uap (ST) dalam satu siklus untuk menghasilkan listrik. Konsep ini dirancang untuk meningkatkan efisiensi penggunaan bahan bakar dan mengoptimalkan pemanfaatan panas

yang dihasilkan dalam proses pembangkit listrik [1]. CCPP merupakan teknologi yang terus berkembang dalam upaya meningkatkan efisiensi konversi energi dari bahan bakar menjadi listrik. Efisiensi dari pembangkit ini, sering dipengaruhi oleh faktor lingkungan seperti suhu udara lingkungan, kelembaban, dan tekanan atmosfer. Suhu udara dapat mempengaruhi efisiensi pembangkit listrik, terutama pada turbin gas. Suhu yang tinggi dapat menurunkan

efisiensi turbin gas karena menurunnya perbedaan suhu antara gas pembakaran dan udara di sekitarnya [2]. sedangkan tekanan atmosfer dapat mempengaruhi kerja turbin gas dan turbin uap. Turbin gas dapat mengalami penurunan efisiensi pada tekanan tinggi, sementara turbin uap membutuhkan tekanan yang cukup untuk menghasilkan energy [3]. Pemahaman yang baik tentang faktor lingkungan dapat membantu dalam menyesuaikan operasional untuk memaksimalkan kinerja kedua jenis turbin. Oleh sebab itu, diperlukan sebuah metode yang lebih efektif dan inovatif dalam mengoptimalkan pengoperasian CCPP pada berbagai kondisi lingkungan.

Machine learning merupakan salah satu metode yang sering digunakan dalam pemodelan efisiensi energi, optimasi energi dan prediksi energi [4][5]. Machine learning dapat memperhitungkan hubungan antara suhu udara lingkungan, kelembaban, dan tekanan atmosfer dengan efisiensi turbin gas dan turbin uap [6]. Dengan memahami pola-pola ini, model machine learning dapat memberikan rekomendasi untuk penyesuaian operasional, seperti peningkatan pembakaran pada suhu udara tinggi atau penyesuaian tekanan pada turbin gas. Hal ini bertujuan untuk mengoptimalkan produksi listrik, meningkatkan efisiensi, dan mengurangi dampak lingkungan seiring dengan perubahan kondisi lingkungan. Metode ini memanfaatkan kemampuan algoritma untuk memproses dan memahami pola-pola kompleks dalam data lingkungan serta parameter operasional CCPP. Dengan menganalisis data historis, machine learning dapat membangun model prediktif yang mampu meramalkan kinerja CCPP berdasarkan variasi kondisi lingkungan [7].

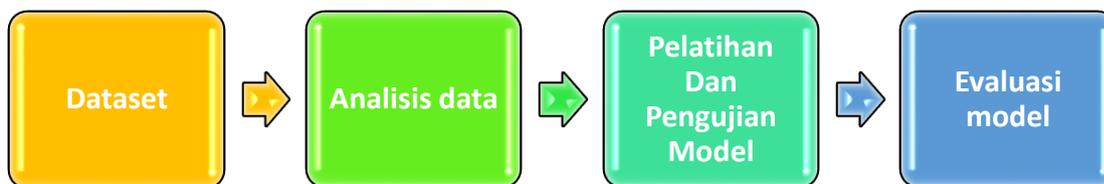
Banyak penelitian terdahulu yang menggunakan machine learning dalam memprediksi daya listrik seperti penelitian [8] tentang prediksi daya output listrik pembangkit siklus gabungan berdasarkan faktor lingkungan. Hasil penelitian ini, menekankan pentingnya pemahaman yang mendalam terhadap faktor lingkungan dalam pengelolaan CCPP. Studi ini menyimpulkan bahwa integrasi machine learning dapat membantu pengelola pembangkit dalam

mengidentifikasi pola-pola kompleks dalam data lingkungan, sehingga memungkinkan pengoptimalan operasional yang lebih cepat. dalam penelitian [9] menyoroti peran penting machine learning dalam prediksi efisiensi energi dan pengoperasian optimal pada pembangkit listrik. Hasil penelitian ini menunjukkan bahwa model machine learning dapat meningkatkan ketepatan prediksi dan memberikan solusi adaptif untuk merespons perubahan lingkungan.

Berdasarkan permasalahan yang sudah diuraikan diatas dan upaya untuk meningkatkan efisiensi energi pada pembangkit listrik siklus gabungan menggunakan metode machine learning. Penelitian ini bertujuan untuk merancang dan membandingkan empat algoritma machine learning yang terdiri dari Linear Regression (LR), Decision Tree (DT), Random Forest (RF), dan Artificial Neural Network (ANN) untuk meprediksi daya listrik pada Pembangkit Listrik Siklus Gabungan berdasarkan faktor lingkungan. Parameter ke akuratan model dalam penelitian ini diukur dengan nilai Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan Koefisien determinasi (R^2), berdasarkan nilai aktual dan nilai prediksi dari data uji.

II. METODE PENELITIAN

Pada penelitian ini, dilakukan perbandingan beberapa model prediktif machine learning untuk meramalkan produksi daya listrik dari pembangkit listrik siklus gabungan berdasarkan kondisi lingkungan. Dataset yang digunakan untuk pelatihan model machine learning terdiri dari data kondisi lingkungan, seperti suhu udara, tekanan atmosfer, kelembaban relatif, vakum pembuangan, dan output daya listrik. Model machine learning dikembangkan menggunakan bahasa pemrograman Python dan diimplementasikan dengan memanfaatkan parameter default dari library scikit-learn versi 1.2.2, serta Keras versi 2.12.0. Tahapan dalam penelitian ini, dapat dilihat dengan lebih jelas pada Gambar 1.



Gambar 1. Tahapan Penelitian

1. Dataset

Dataset yang digunakan dalam penelitian ini diperoleh UCI machine learning repository [6]. Dataset ini terdiri dari 9568 titik data yang dikumpulkan dari Pembangkit Listrik Siklus Gabungan (CCPP) antara tahun 2006 hingga 2011. Selama periode ini, pembangkit listrik bekerja dengan beban penuh. Pada CCPP, turbin gas dan turbin uap menghasilkan listrik bersama-sama dalam siklus yang sama. Pada turbin gas, udara yang terkompresi dicampur dengan bahan bakar yang dipanaskan pada suhu tinggi. Campuran ini melewati bilah turbin dan memutarinya, menghasilkan listrik. Pada saluran buang, campuran ini masih sangat panas dan memiliki potensi untuk pembangkitan daya lebih lanjut. Generator Uap Pemanfaatan Panas (HRSG) menangkap gas buang yang panas dan menggunakannya untuk menghasilkan uap. Uap

ini kemudian disalurkan ke turbin uap untuk menghasilkan listrik tambahan. CCPP tempat dataset saat ini diperoleh terdiri dari dua Turbin Gas ABB 13E2 160 MW, dua HRSG, dan satu Turbin Uap ABB 160 MW. CCPP ini memiliki kapasitas penghasilan nominal sebesar 480 MW (nama perusahaan utilitas yang menyediakan data ini dirahasiakan). Dataset ini mencakup sejumlah variabel ambient yang diukur secara rata-rata per jam, seperti *Temperature* (T) Suhu udara sekitar, *Ambient Pressure* (AP) Tekanan udara sekitar, *Relative Humidity* (RH) Kelembaban relatif udara sekitar, *Exhaust Vacuum* (V) Tekanan vakum di saluran pembuangan, dan *Electrical Energy Output* (EP) produksi energi listrik bersih per jam dari pembangkit listrik tersebut. Statistik dataset CCPP yang digunakan untuk pelatihan model prediksi dapat dilihat pada Tabel 1.

Tabel 1. Statistik dataset pelatihan pembangkit listrik siklus gabungan.

Variabel	Kategori	Mean	Min	Max
Temperature (c)	Input	19.648	1.81	37.11
Vacuum (cmHg)	Input	54.293	25.36	81.56
Pressure (mbar)	Input	1013.265	992.89	1033.3
Humidity (%)	Input	73.292	25.56	100.16
Energy Output (MW)	Output	454.379	420.26	495.76

2. Analisis data

Analisis data dilakukan dengan tujuan untuk memahami pola, hubungan, dan karakteristik dari data operasi pembangkit listrik. Data yang telah melalui proses preprocessing kemudian dianalisis menggunakan metode statistik deskriptif dan menganalisis hubungan antara variabel input dengan variabel output menggunakan korelasi pearson, yang dapat dihitung menggunakan persamaan 2.

$$r_{xy} = \frac{\sum xy}{(n-1)S_x S_y} \quad (1)$$

Dimana r_{xy} adalah koefisien korelasi Pearson, $\sum xy$ adalah jumlah perkalian x dan y , n adalah ukuran sampel, x adalah variabel independen, y adalah variabel dependen, dan S adalah standar deviasi [10]. Nilai koefisien korelasi berkisar dari -1 hingga 1. Nilai -1 menunjukkan korelasi negatif yang kuat antara kedua variabel, nilai 0 menunjukkan tidak ada korelasi, dan nilai 1 menunjukkan korelasi positif yang kuat [11].

3. Pelatihan Dan Pengujian Model

Pada tahap ini, dilakukan perbandingan empat algoritma machine learning yang terdiri dari Linear Regression (LR), Decision Tree (DT), Random Forest (RF), dan Artificial Neural

Network (ANN). Algoritma ini diimplementasikan dengan menggunakan parameter default yang disediakan oleh library scikit-learn versi 1.2.2 dan Keras versi 2.12.0. Data yang digunakan untuk pelatihan dan pengujian model dibagi menjadi dua bagian, yaitu data pelatihan 80% dan data pengujian 20%.

a. Linear Regression (LR)

Linear Regression merupakan metode dalam statistika dan machine learning yang digunakan untuk memodelkan hubungan linear antara satu variabel dependen (output) dengan satu atau lebih variabel independen (input). Tujuannya adalah untuk menemukan garis terbaik yang dapat merepresentasikan hubungan linier antara variabel-variabel tersebut [12]. Pada penelitian ini algoritma LR menggunakan 80% data sebagai data training dan 20% digunakan sebagai data testing kemudian performa model yang dilatih dievaluasi dengan menghitung Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan koefisien determinasi (R^2), dari nilai aktual dan nilai prediksi dari data uji.

b. Random Forest (RF)

Random forest adalah metode pembelajaran ensemble untuk klasifikasi yang digunakan dengan membangun pohon keputusan dari set pelatihan dalam iterasi k. Dalam setiap iterasi, algoritma pelatihan pertamamata secara acak memilih satu set sampel dari set pelatihan. Untuk mereproduksi pohon keputusan dari subset ini, RF secara acak memilih subset fitur sebagai kandidat fitur untuk setiap node [13]. Dengan demikian, setiap pohon keputusan dibangun melalui ansambel menggunakan himpunan bagian independen acak dari fitur dan sampel. Pada penelitian ini algoritma DT menggunakan 80% data sebagai data training dan 20% digunakan sebagai data testing kemudian performa model yang dilatih dievaluasi dengan menghitung Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan koefisien determinasi (R^2), dari nilai aktual dan nilai prediksi dari data uji.

c. Artificial Neural Network (ANN)

Neural network merupakan salah satu metode machine learning classifier atau metode yang sangat powerfull. Neural network hampir selalu cocok dengan berbagai permasalahan machine learning sehingga neural network dapat menjadi hypotheses untuk berbagai persoalan rill sebab Neural network dibentuk berdasarakan cara kerja syaraf otak manusia [14]. Pada penelitian ini algoritma DT menggunakan 80% data sebagai data training dan 20% digunakan sebagai data testing kemudian performa model yang dilatih dievaluasi dengan menghitung Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan koefisien determinasi (R^2), dari nilai aktual dan nilai prediksi dari data uji.

d. Decision Tree (DT)

Decion tree regresion merupakan salah satu predictive model yang digunakan pada machine learning, model ini adalah teknik terawasi yang melakukan teknik klasifikasi dan regresi. Pohon keputusan bekerja dengan baik untuk input kategorikal dan kontinu, variabel output. Ini mengembangkan pohon keputusan terkait dengan membagi dataset menjadi subset yang lebih kecil [15]. Pada penelitian ini algoritma DT menggunakan 80% data sebagai data training dan 20% digunakan sebagai data testing kemudian performa model yang dilatih dievaluasi dengan menghitung Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan

koefisien determinasi (R^2), dari nilai aktual dan nilai prediksi dari data uji.

Evaluasi kinerja model dihitung menggunakan tiga metrik, yakni Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan R-Squared. Metrik evaluasi model dapat dihitung menggunakan persamaan berikut [5] :

a. Mean Absolute Error (MAE)

MAE mengukur rata-rata dari selisih absolut antara prediksi model dan nilai target. Semakin rendah nilai MAE, semakin baik model dalam melakukan prediksi. MAE dapat dihitung menggunakan persamaan 2.

$$MAE = \frac{1}{N} \sum |y_i - z_i| \quad (2)$$

Dimana i adalah indeks dari data pada sample, N adalah jumlah total sample, y_i adalah nilai aktual dari data ke- i , sedangkan z_i adalah nilai prediksi dari model untuk data ke- i .

b. Root Mean Square Error (RMSE)

RMSE adalah akar dari rata-rata kuadrat dari selisih antara prediksi model dan nilai target. RMSE memberikan ukuran yang sama dengan variabel target dan umumnya lebih sensitif terhadap perbedaan yang besar. Semakin rendah nilai RMSE, semakin baik model dalam melakukan prediksi. RMSE dapat dihitung menggunakan persamaan 3.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (f(X_i) - Y_i)^2} \quad (3)$$

Dimana n adalah jumlah data yang digunakan untuk menguji model, $f(X_i)$ adalah nilai yang diprediksi oleh model untuk data ke- i , Y_i adalah nilai sebenarnya untuk data ke- i .

c. R-squared (R^2)

R-squared (R^2) adalah koefisien determinasi yang memberikan informasi tentang seberapa baik model cocok dengan data. R^2 adalah rasio variasi total yang dijelaskan oleh model terhadap variasi total yang ada dalam data. Nilai R^2 berkisar dari 0 hingga 1, dan semakin tinggi nilai R^2 , semakin baik model dalam menjelaskan variasi dalam data.

R-squared dapat dihitung menggunakan persamaan 4.

$$R = \frac{\sum_{i=1}^n (f(X_i) - \bar{f})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (f(X_i) - \bar{f})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (4)$$

Dimana $f(X_i)$ adalah nilai prediksi dari variabel dependen (Y) berdasarkan variabel independen (X) pada observasi ke- i , \bar{f}

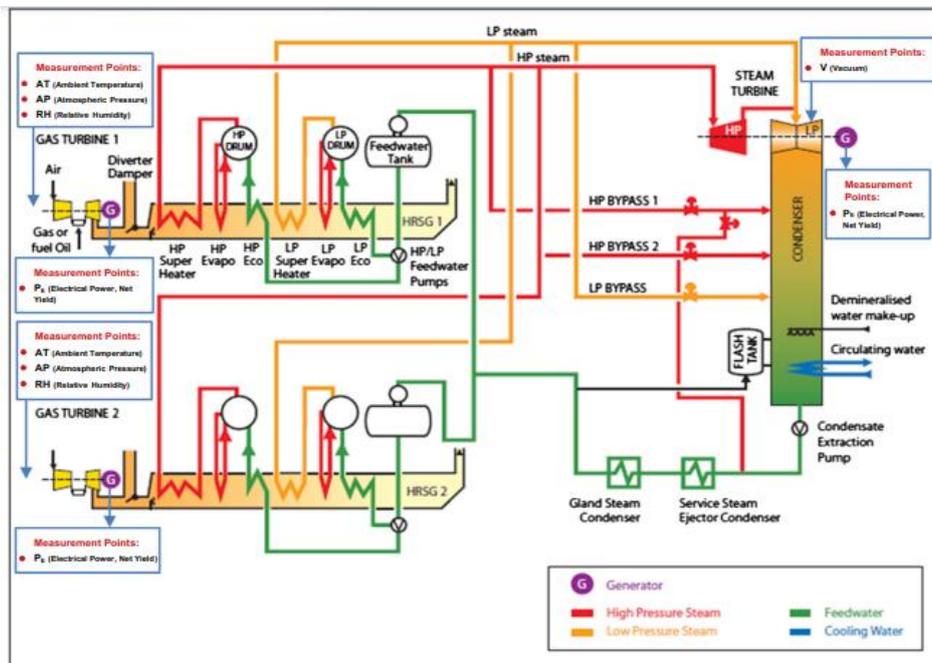
adalah rata-rata dari semua nilai prediksi $f(X_i)$ pada seluruh observasi, Y_i adalah nilai observasi aktual dari variabel dependen pada observasi ke- i , \bar{Y} adalah rata-rata dari semua nilai observasi Y_i pada seluruh observasi, dan n adalah jumlah total observasi.

4. Evaluasi model

Evaluasi model merupakan suatu proses penilaian kinerja suatu model machine learning dengan menggunakan metrik atau indikator tertentu. Tujuan dari evaluasi model adalah untuk memahami sejauh mana model tersebut dapat menggeneralisasi dan membuat prediksi yang akurat pada data yang belum pernah dilihat sebelumnya [16]. Pada penelitian ini, setelah dilakukan pelatihan model, keempat model machine learning dievaluasi melalui pengujian eksternal menggunakan data baru. Hasil dari pengujian eksternal ini dihitung kembali dengan menggunakan metrik evaluasi seperti MAE, RMSE, dan R-squared

III. HASIL DAN PEMBAHASAN

CCPP bekerja dengan dua siklus pembangkit energi secara bersamaan. Pertama, gas dari pembakaran bahan bakar seperti gas digunakan untuk menggerakkan turbin gas, sehingga menghasilkan energi kinetik. Setelah itu, gas buang yang masih panas dialirkan ke suatu sistem penukar panas, di mana panasnya digunakan untuk memanaskan air dan mengubahnya menjadi uap. Kedua, uap yang dihasilkan dari penukar panas digunakan untuk menggerakkan turbin uap, yang juga menghasilkan energi kinetik. Dengan menggunakan dua turbin ini secara bersamaan, Combined Cycle Power Plant dapat mengoptimalkan pemanfaatan panas yang dihasilkan dari pembakaran bahan bakar. Proses ini meningkatkan efisiensi secara keseluruhan, karena panas yang sebelumnya diabaikan dalam siklus tunggal dapat digunakan kembali untuk menghasilkan lebih banyak energi [17]. Hal ini dapat mencapai tingkat efisiensi yang lebih tinggi dibandingkan dengan pembangkit listrik konvensional yang hanya menggunakan satu siklus pembangkitan energi. Ilustrasi dari siklus CCPP dapat dilihat pada Gambar 2.

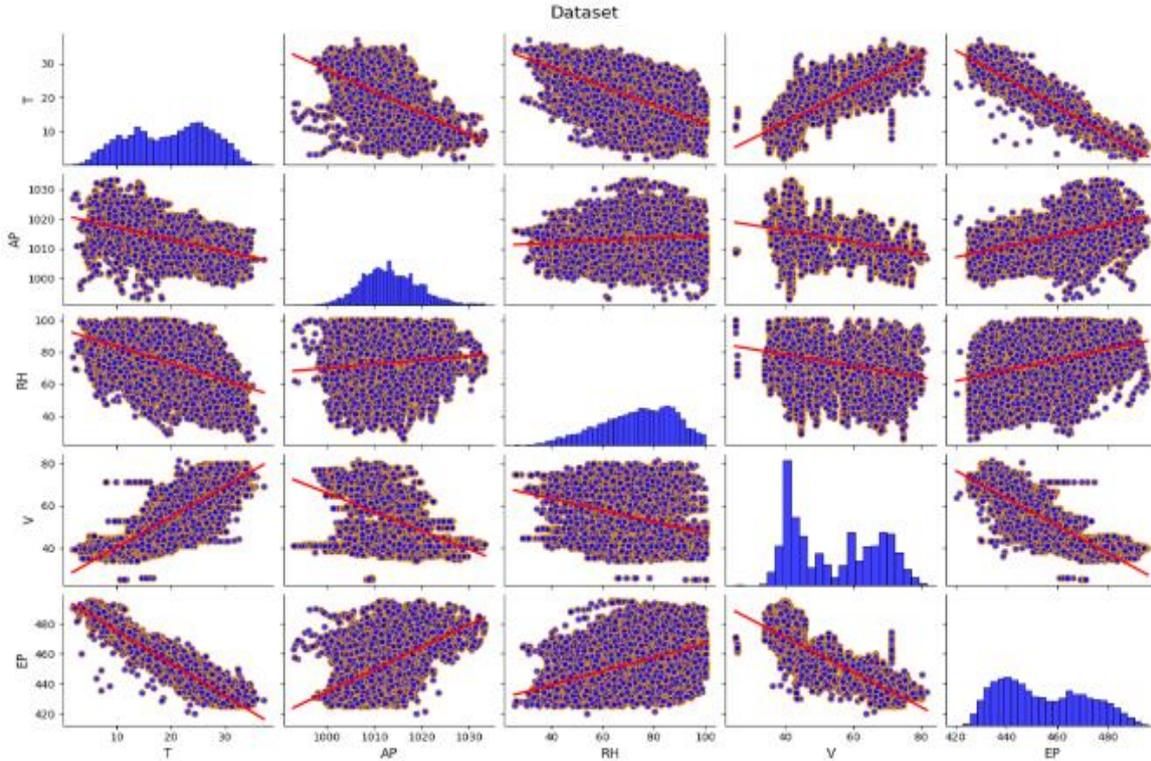


Gambar 2. Tata letak pembangkit listrik siklus gabungan [6].

1. Dataset

Dataset ini terdiri dari total 9568 sampel data, kemudian dibagi menjadi dua bagian yaitu 9468 sampel pada tahap pelatihan model dan dihilangkan sebanyak 100 sampel dari dataset pelatihan untuk

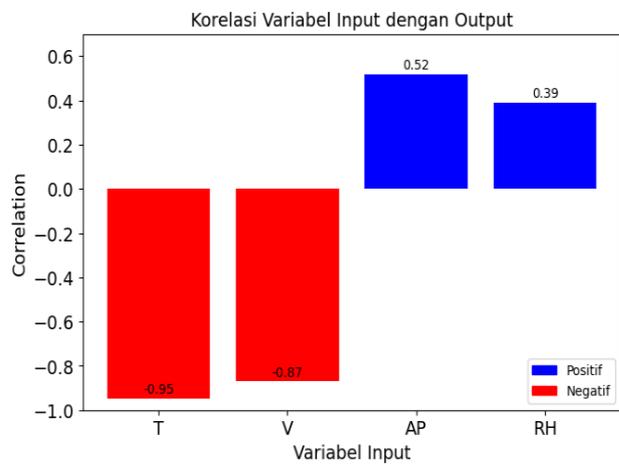
pengujian eksternal nantinya. Dataset ini terdiri dari empat variabel input (Temperature, Vacuum, Pressure, dan Humidity) dan 1 variabel output (EP). Distribusi dataset CCPP pada penelitian ini dapat dilihat pada Gambar 3.



Gambar 3. Distribusi dataset CCPP

2. Analisis Data

Pada tahap ini, dilakukan analisis korelasi untuk melihat hubungan antara variable input seperti temperature, vacuum, pressure, dan humidity dengan variable output yaitu daya listrik. Analisis korelasi yang digunakan adalah korelasi pearson, dimana metode ini digunakan untuk mengukur sejauh mana dua variabel berkaitan satu sama lain dalam hubungan linier. Nilai koefisien korelasi berkisar dari -1 hingga 1. Nilai -1 menunjukkan korelasi negatif yang kuat antara kedua variabel, nilai 0 menunjukkan tidak ada korelasi, dan nilai 1 menunjukkan korelasi positif yang kuat [11]. Hasil analisis korelasi menunjukkan bahwa terdapat hubungan yang signifikan antara variabel-variabel lingkungan (temperatur, vakum, tekanan, dan kelembaban) dengan output daya listrik pada pembangkit listrik siklus gabungan. Hasil koeralasi person antara variable input dan variable output dapat dilihat pada Gambar 4.



Gambar 4. Hasil korelasi person antara variable input dengan variable output.

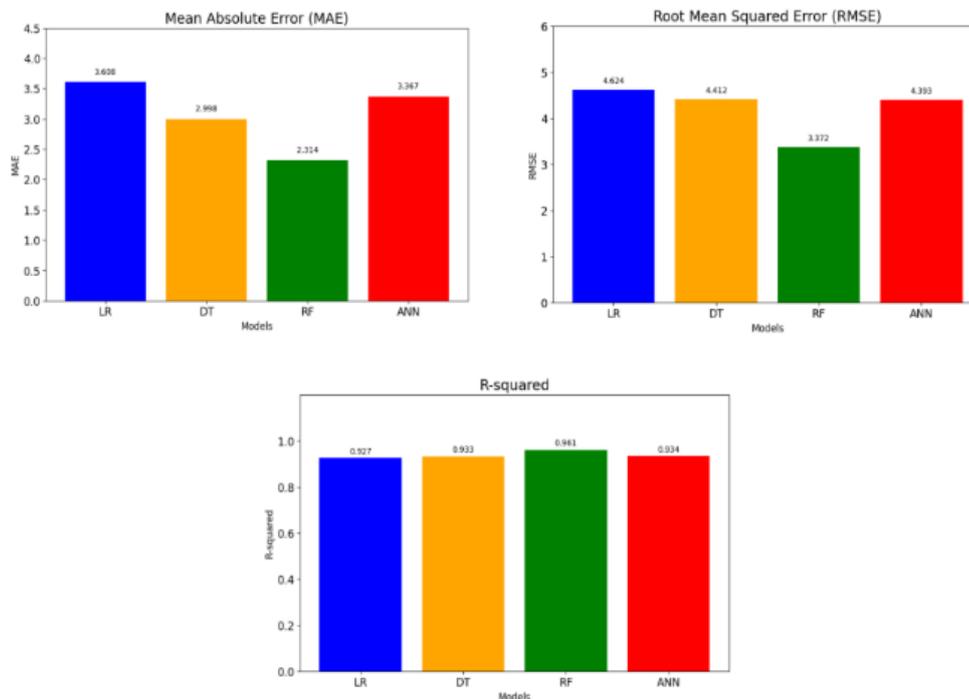
Berdasarkan Gambar 2, dapat dilihat bahwa temperatur dan vakum memiliki korelasi negatif yang sangat kuat dengan daya listrik dengan nilai masing -masing -0.95 dan -0.87. Hal ini, mengindikasikan bahwa peningkatan suhu dan vakum pada CCPP dapat menurunkan daya listrik yang dihasilkan. Peningkatan suhu pada CCPP menyebabkan peningkatan entropi, yang pada gilirannya menyebabkan penurunan efisiensi termal. Sementara itu, peningkatan vakum pada CCPP menyebabkan penurunan tekanan, yang pada gilirannya menyebabkan penurunan efisiensi

termal [18]. Oleh karena itu, peningkatan suhu dan vakum pada CCPP dapat menurunkan daya listrik yang dihasilkan. Sedangkan tekanan dan kelembaban memiliki korelasi positif yang cukup kuat dengan daya listrik yaitu 0.52 dan 0.39. Hal ini, menunjukkan bahwa peningkatan tekanan dan kelembaban dapat meningkatkan daya listrik. Peningkatan tekanan pada CCPP dapat meningkatkan efisiensi termal, karena peningkatan tekanan dapat meningkatkan kerapatan udara, yang pada gilirannya meningkatkan konduktivitas termal [19].

3. Pelatihan dan pengujian model

Pada tahap ini, dilakukan perbandingan empat algoritma machine learning yang terdiri dari Linear Regression (LR), Decision Tree (DT), Random Forest (RF), dan Artificial Neural Network (ANN). Algoritma ini diimplementasikan dengan menggunakan parameter default yang disediakan oleh library scikit-learn versi 1.2.2 dan Keras versi 2.12.0. Setiap model dievaluasi menggunakan metrik evaluasi seperti MAE, RMSE dan R-squared. Jumlah dataset yang digunakan adalah 9468 sampel dan dilakukan pembagian dataset menjadi 80% (7574 sampel) untuk pelatihan dan 20% (1894 sampel) untuk pengujian. Model yang memiliki kinerja terbaik adalah model yang mampu

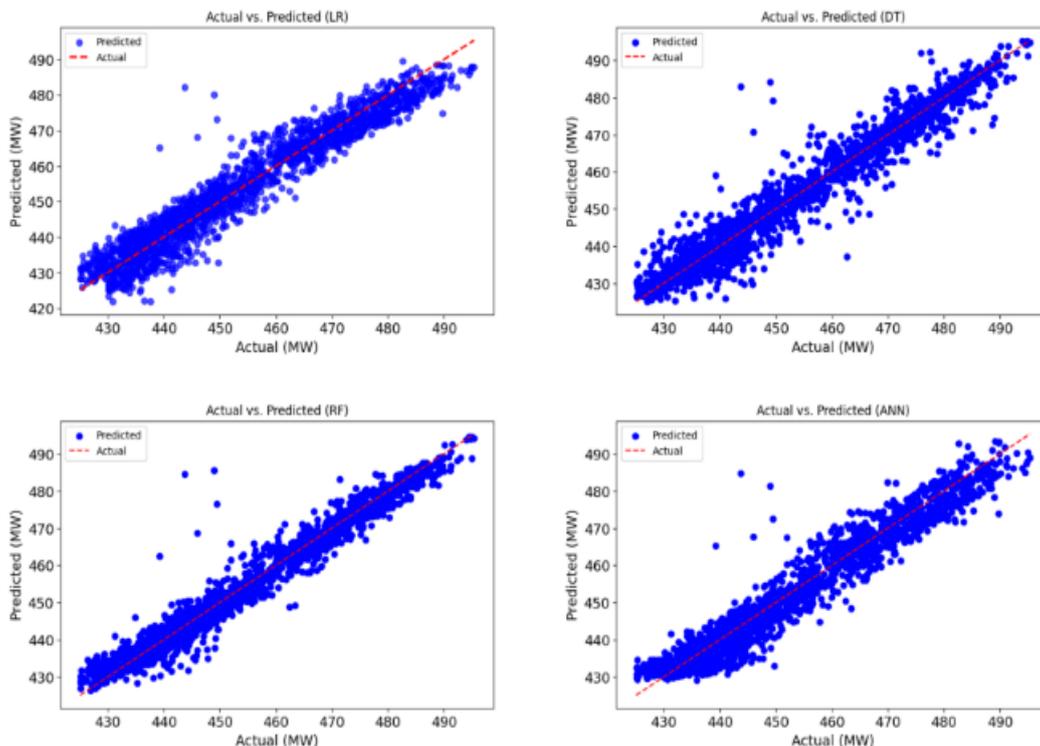
memperoleh nilai MAE dan RMSE terkecil, dan R-squared terbesar. MAE mengukur rata-rata dari selisih absolut antara prediksi model dan nilai target. Semakin rendah nilai MAE, semakin baik model dalam melakukan prediksi. RMSE adalah akar dari rata-rata kuadrat dari selisih antara prediksi model dan nilai target. RMSE memberikan ukuran yang sama dengan variabel target dan umumnya lebih sensitif terhadap perbedaan yang besar. Semakin rendah nilai RMSE, semakin baik model dalam melakukan prediksi. Sedangkan R-squared (R^2) adalah koefisien determinasi yang memberikan informasi tentang seberapa baik model cocok dengan data. R^2 adalah rasio variasi total yang dijelaskan oleh model terhadap variasi total yang ada dalam data. Nilai R^2 berkisar dari 0 hingga 1, dan semakin tinggi nilai R^2 , semakin baik model dalam menjelaskan variasi dalam data [5]. Hasil perbandingan kinerja keempat model dapat dilihat dari metrik evaluasi pada gambar 5.



Gambar 5. Perbandingan metrik evaluasi pengujian model

Berdasarkan hasil pelatihan model menggunakan empat algoritma machine learning, dapat dilihat bahwa Random Forest (RF) memiliki kinerja yang paling baik dibandingkan dengan tiga model lainnya. Gambar 4 menunjukkan bahwa RF mampu memperoleh metrik evaluasi seperti MAE 2.314, RMSE 3.372 dan R-squared 0.961. Sedangkan model LR merupakan model yang paling berkinerja buruk dalam pelatihan ini yaitu dengan nilai MAE 3.608, RMSE 4.624, dan R-squared 0.927. Hasil ini dapat terjadi karena beberapa factor seperti *non-linearity*, *ensemble learning*, dan *parameter tuning*. RF mampu menangani hubungan non-linear antara variabel input dan output. Model LR hanya dapat menangani hubungan linier, sementara RF dapat menangkap pola yang lebih kompleks dan non-linear dalam data, yang terdapat dalam kasus

prediksi daya listrik. RF juga menggunakan teknik ensemble learning, yang menggabungkan prediksi dari beberapa pohon keputusan (trees). Setiap pohon membuat prediksi secara independen, dan hasilnya digabungkan untuk memberikan prediksi akhir. Pendekatan ini dapat mengurangi overfitting dan meningkatkan generalisasi model. Selain itu, RF memiliki lebih banyak parameter yang dapat disesuaikan daripada model linear sederhana. Dengan penyesuaian parameter yang tepat, RF dapat dioptimalkan untuk kinerja yang lebih baik pada data tertentu [20]. Hasil perbandingan keempat model machine learning dapat dilihat pada Gambar 6.



Gambar 6. Perbandingan hasil prediksi model

4. Evaluasi model

Evaluasi model dengan data baru merupakan tahap penting dalam mengukur sejauh mana model machine learning dapat menggeneralisasi dan memberikan prediksi yang akurat pada data yang belum pernah dilihat selama proses pelatihan. Dalam konteks ini, data baru adalah data yang tidak termasuk dalam set pelatihan model, dan pengujian

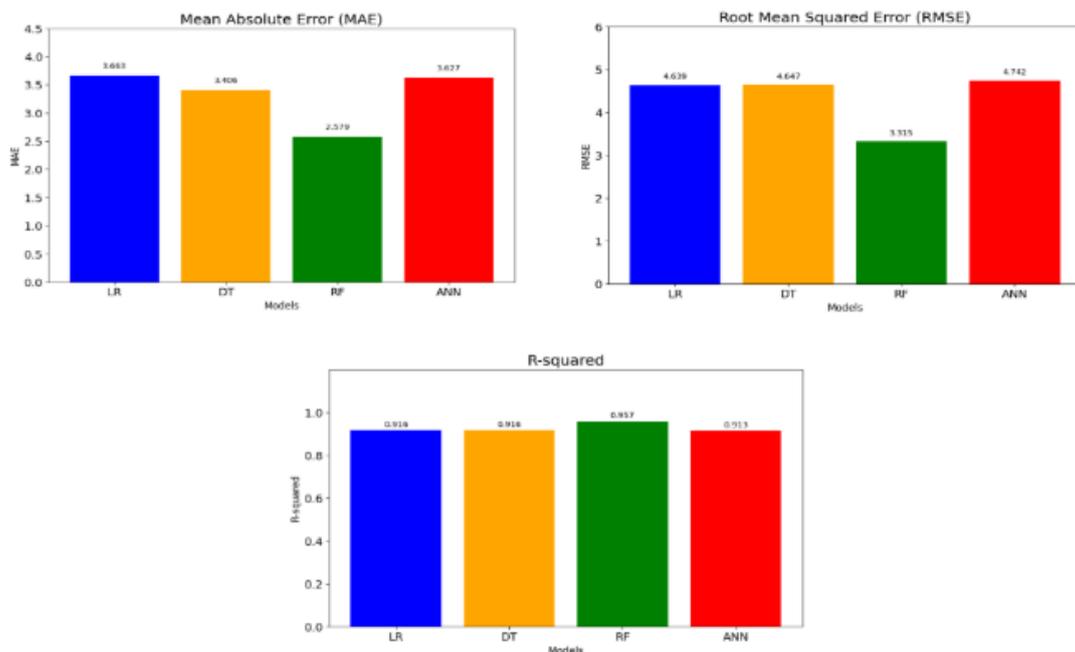
model dilakukan pada data ini untuk mengukur kinerja model di luar sampel yang digunakan selama pelatihan. Evaluasi model dengan data baru membantu mengidentifikasi apakah model mampu melakukan generalisasi dengan baik pada berbagai kondisi atau apakah mungkin terjadi overfitting, di mana model terlalu fokus pada data pelatihan dan kurang fleksibel dalam menghadapi data baru [12].

Evaluasi ini memastikan bahwa model dapat diterapkan dengan efektif dalam situasi dunia nyata. Metrik evaluasi seperti Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), dan R-squared digunakan untuk mengukur seberapa baik model memprediksi nilai target pada data baru. Metrik ini memberikan gambaran tentang seberapa dekat prediksi model dengan nilai sebenarnya.

Hasil evaluasi digunakan untuk mengevaluasi kinerja model secara keseluruhan. Model yang dapat memberikan prediksi yang akurat pada data baru dianggap lebih baik dan lebih dapat diandalkan. Hasil perbandingan metrik evaluasi keempat model dapat dilihat pada Gambar 7, dan statistik data pengujian eksternal dapat dilihat pada Tabel 2.

Tabel 2. Statistik dataset pengujian pembangkit listrik siklus gabungan.

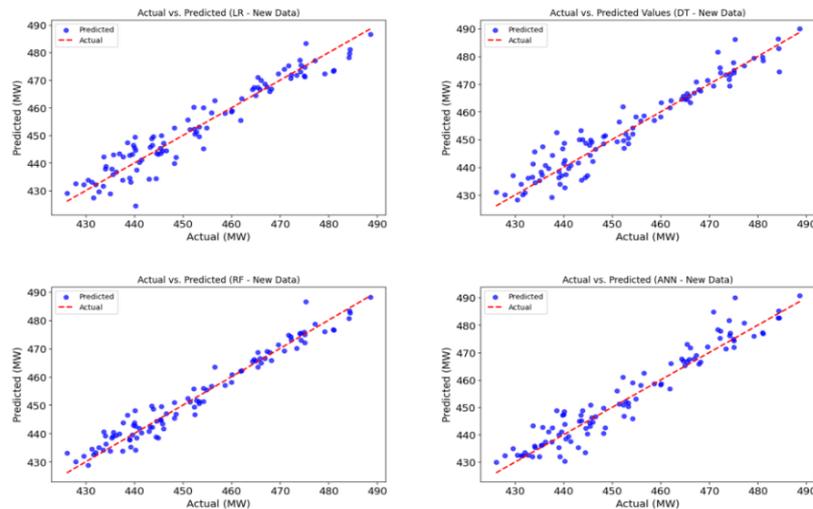
Variabel	Kategori	Mean	Min	Max
Temperature (c)	Input	19.976	5.01	34.03
Vacuum (cmHg)	Input	55.441	25.36	77.95
Pressure (mbar)	Input	1012.639	998.27	1027.12
Humidity (%)	Input	74.894	36.48	100.1
Energy Output (MW)	Output	452.952	36.48	488.65



Gambar 7. Metrik evaluasi pengujian eksternal

Data pengujian eksternal ini, berjumlah 100 sampel dengan karakteristik seperti pada table 1. Pengujian model ini, sama halnya dengan pelatihan model sebelumnya yaitu menggunakan parameter default scikit-learn versi 1.2.2 dan Keras versi 2.12.0. Hasil pengujian eksternal tidak jauh berbeda dengan hasil pelatihan sebelumnya, dimana RF mampu memperoleh kinerja terbaik dengan nilai MAE 2.579, RMSE 3.315, dan R-squared 0.957. Namun,

pada pengujian eksternal ini model yang berkinerja paling buruk adalah ANN dengan nilai MAE 3.627, RMSE 4.742, dan R-squared 0.913. Hasil ini mengindikasikan bahwa RF memiliki performa yang stabil dan dapat diandalkan dalam melakukan prediksi pada dataset yang lebih besar dan beragam. Hasil perbandingan prediksi model menggunakan data baru dapat dilihat pada Gambar 8.



Gambar 8. Perbandingan hasil prediksi model pada pengujian eksternal

IV. KESIMPULAN

Berdasarkan penelitian mengenai Prediksi Daya Listrik pada Pembangkit Listrik Siklus Gabungan (CCPP) dalam Kondisi Lingkungan menggunakan Metode Machine Learning, dapat disimpulkan bahwa data historis pengoperasian sistem pembangkit listrik, terutama pada CCPP, dapat dijadikan alat bantu untuk mengoptimalkan kinerja pembangkit listrik dalam berbagai kondisi lingkungan. Dalam era digitalisasi saat ini, data memegang peran aktif dalam proses pemodelan untuk memprediksi atau mengestimasi daya listrik pada berbagai kondisi lingkungan, menggunakan metode machine learning. Penelitian ini menerapkan empat algoritma machine learning sebagai model prediksi daya listrik pada CCPP berdasarkan kondisi lingkungan. Hasil penelitian menunjukkan bahwa model Random Forest (RF) mencapai kinerja terbaik dibandingkan dengan model-model lainnya dengan nilai MAE 2.314, RMSE 3.372 dan R-squared 0.961. Selain itu, model RF juga memiliki kinerja paling baik dibandingkan model lainnya saat dilakukan pengujian eksternal dengan data baru, dimana RF memperoleh nilai MAE 2.579, RMSE 3.315, dan R-squared 0.957. Hasil ini mengindikasikan bahwa model Random Forest (RF) menunjukkan kinerja yang paling baik dibandingkan dengan tiga model lainnya, yaitu Linear Regression (LR), Decision Tree (DT), dan Artificial Neural Network (ANN). Model RF juga mencapai kinerja optimal pada pengujian eksternal, di mana model diuji menggunakan data baru. Temuan ini mengindikasikan bahwa model machine learning menggunakan Random Forest dapat dijadikan sebagai model prediksi daya listrik yang efektif pada CCPP. Keberlanjutan penelitian ini diharapkan dapat memberikan kontribusi lebih lanjut terhadap pemahaman dan penerapan machine learning dalam industri pembangkit listrik.

Daftar Pustaka

- [1]. Siddiqui, R., Anwar, H., Ullah, F., Ullah, R., Rehman, M. A., Jan, N., & Zaman, F. (2021). Power prediction of combined cycle power plant (CCPP) using machine learning algorithm-based paradigm. *Wireless Communications and Mobile Computing*, 2021, 1-13.
- [2]. Kurz, R., & Brun, K. (2016). Gas turbine performance. In *Asia Turbomachinery & Pump Symposium. 2016 Proceedings.. Turbomachinery Laboratories, Texas A&M Engineering Experiment Station.*
- [3]. Bandić, L., Hasičić, M., & Kevrić, J. (2020). Prediction of power output for combined cycle power plant using random decision tree algorithms and ANFIS. In *Advanced technologies, systems, and applications IV-proceedings of the international symposium on innovative and interdisciplinary applications of advanced technologies (IAT 2019)* (pp. 406-416). Springer International Publishing.
- [4]. Duchesne, Laurine, Efthymios Karangelos, and Louis Wehenkel. "Recent developments in machine learning for energy systems reliability management." *Proceedings of the IEEE* 108.9 (2020): 1656-1676.
- [5]. Leni, D. (2023). PREDIKSI PELABELAN RATING AC EFISIENSI ENERGI MENGGUNAKAN PEMODELAN MACHINE LEARNING. *Majalah Ilmiah Momentum*, 19(1).
- [6]. Tüfökcü, P. (2014). Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods. *International Journal of Electrical Power & Energy Systems*, 60, 126-140.
- [7]. Alketbi, S., Nassif, A. B., Eddin, M. A., Shahin, I., & Elnagar, A. (2020, November). Predicting the power of a combined cycle power plant using machine learning methods. In *2020 International conference on communications, computing, cybersecurity, and informatics (CCCI)* (pp. 1-5). IEEE.
- [8]. Kaewprapha, P., Prempaneeerach, P., Singh, V., Tinikul, T., & Intarangsri, N. (2022, March). Machine Learning Approaches for Estimating the Efficiency of Combined Cycle Power Plant. In *2022 International Electrical Engineering Congress (iEECON)* (pp. 1-4). IEEE.
- [9]. Kaewprapha, P., Prempaneeerach, P., Singh, V., Tinikul, T., Intarangsri, N., & Kijkanjanarat, T. (2022, April). Predicting Full Load, Partial Load Efficiency of a Combined Cycle Power Plant Using Machine Learning Methods. In *2022 7th International Conference on Computer and Communication Systems (ICCCS)* (pp. 11-16). IEEE.
- [10]. Leni, D., Earnestly, F., Sumiati, R., Adriansyah, A., & Kusuma, Y. P. (2023). Evaluasi sifat mekanik baja paduan rendah berdasarkan komposisi kimia dan suhu perlakuan panas

- menggunakan teknik exploratory data analysis (EDA). *Dinamika Teknik Mesin*, 13(1), 74-83.
- [11]. Leni, D. (2023). Analisis Heatmap Korelasi dan Scatterplot untuk Mengidentifikasi Faktor-Faktor yang Mempengaruhi Pelabelan AC efisiensi Energi. *Jurnal Rekayasa Material, Manufaktur dan Energi*, 6(1).
- [12]. Maulud, D., & Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends*, 1(4), 140-147.
- [13]. Nakatsu, R. T. (2020). An evaluation of four resampling methods used in machine learning classification. *IEEE Intelligent Systems*, 36(3), 51-57.
- [14]. Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F., & Campbell, J. P. (2020). Introduction to machine learning, neural networks, and deep learning. *Translational vision science & technology*, 9(2), 14-14.
- [15]. Muhammad, L. J., Algehyne, E. A., & Usman, S. S. (2020). Predictive supervised machine learning models for diabetes mellitus. *SN Computer Science*, 1(5), 240.
- [16]. Pham, B. T., Jaafari, A., Avand, M., Al-Ansari, N., Dinh Du, T., Yen, H. P. H., ... & Tuyen, T. T. (2020). Performance evaluation of machine learning methods for forest fire modeling and prediction. *Symmetry*, 12(6), 1022.
- [17]. [Aliyu, M., AlQudaihi, A. B., Said, S. A., & Habib, M. A. (2020). Energy, exergy and parametric analysis of a combined cycle power plant. *Thermal Science and Engineering Progress*, 15, 100450.
- [18]. Dev, N., & Attri, R. (2015). Performance analysis of combined cycle power plant. *Frontiers in Energy*, 9, 371-386.
- [19]. Riady, M. I., Santoso, D., & Bustan, M. D. (2019, April). Thermodynamics Performance Evaluation in Combined Cycle Power Plant by Using Combined Pinch and Exergy Analysis. In *Journal of Physics: Conference Series* (Vol. 1198, No. 4, p. 042006). IOP Publishing.
- [20]. Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A comprehensive evaluation of ensemble learning for stock-market prediction. *Journal of Big Data*, 7(1), 1-40.