

BAB 2

TINJAUAN PUSTAKA

2.1. Penelitian yang Relevan

Penelitian yang relevan dilakukan guna menunjukkan kebaruan dan signifikansi penelitian ini. Untuk itu, dilakukan perbandingan dengan riset-riset terkait yang telah ada. Perbandingan tersebut menjabarkan secara rinci topik, metode, serta hasil dari penelitian sebelumnya untuk memperkuat konteks studi saat ini.

2.1.1. Ringkasan Penelitian Terdahulu

Penelitian pertama berjudul “Optimization of Heat Rate and Greenhouse Gas Emission Reduction at Coal-Fired Power Plants in Indonesia Through Machine Learning Modeling” (Setyawan & Sudiarto, 2024) bertujuan mengembangkan model prediksi heat rate pada PLTU batubara di Indonesia dengan menerapkan berbagai teknik machine learning, antara lain Linear Regression, Lasso Regression, Polynomial Regression, Ridge Regression, Support Vector Regression, Random Forest Regression, Gradient Boosting Regression, Elastic Net, AdaBoost, Neural Network Regression, Decision Tree Regression, dan Extra Trees Regression. Dataset yang digunakan terdiri dari 468 data hasil performance test PLTU batubara di Indonesia dengan parameter seperti tipe boiler, temperatur lingkungan, temperatur gas buang, dan unburned carbon. Hasil analisis menunjukkan bahwa model Extra Trees Regression memberikan performa terbaik dengan nilai R^2 sebesar 0,946, MAE sebesar 136,884, MSE sebesar 35053,258, dan RMSE sebesar 187,225 untuk pemodelan heat rate, serta mencapai R^2 sebesar 0,992 pada pemodelan emisi CO_2 . Penelitian ini menunjukkan bahwa penerapan machine learning berpotensi meningkatkan efisiensi energi pembangkit sekaligus mendukung target dekarbonisasi nasional melalui optimasi berbasis data.

Penelitian Kedua “Artificial Intelligence Modeling-Based Optimization of an Industrial-Scale Steam Turbine for Moving toward Net-Zero in the Energy Sector” (Ashraf et al., 2023). Penelitian ini mengembangkan model Artificial Neural Network (ANN) dan Support Vector Machine (SVM) untuk meningkatkan efisiensi turbin uap tekanan tinggi (HP turbine) pada PLTU supercritical 660 MW. Berdasarkan validasi

eksternal, model ANN menunjukkan performa prediksi yang lebih baik dibandingkan SVM dan kemudian digunakan untuk analisis sensitivitas berbasis Monte Carlo. Hasil optimasi menunjukkan bahwa efisiensi HP turbine dapat ditingkatkan sebesar 1,43% (half-load), 5,09% (mid-load), dan 3,40% (full-load) dibandingkan kondisi rata-rata operasi. Peningkatan ini menghasilkan penghematan bahan bakar sekitar 3,33 ton/jam (half-load), 7,05 ton/jam (*mid-load*), dan 4,04 ton/jam (*full-load*). Dampak lingkungannya signifikan, dengan estimasi penurunan emisi CO₂ tahunan sebesar 58,3 kt/tahun (half-load), 123,5 kt/tahun (*mid-load*), dan 70,8 kt/tahun (*full-load*), serta reduksi SO₂, CH₄, N₂O, dan Hg. Variabel paling berpengaruh terhadap efisiensi turbin adalah main steam pressure (18,8%), feed water temperature (17,6%), dan governing valve opening (12,7%). Studi ini menunjukkan bahwa pendekatan AI mampu meningkatkan efisiensi termal sekaligus mendukung target net-zero melalui optimasi berbasis data operasi aktual.

Penelitian Ketiga berjudul “Artificial Neural Network (ANN) for Boiler Efficiency Prediction in Indonesian Coal-Fired Power Plant” (Arferiandi et al., 2021a) menerapkan model Artificial Neural Network (ANN) untuk memprediksi efisiensi boiler pada PLTU batubara di Indonesia menggunakan data operasi aktual. Model dikembangkan berdasarkan parameter operasional pembangkit seperti temperatur dan tekanan uap serta variabel pembakaran lainnya. Hasil penelitian menunjukkan bahwa model ANN menghasilkan nilai koefisien determinasi (R^2) sebesar 0,84, yang menunjukkan kemampuan model dalam merepresentasikan hubungan non-linear antarparameter operasi. Studi ini menunjukkan bahwa pendekatan jaringan saraf dapat digunakan untuk pemodelan performa sistem pembakaran pada pembangkit listrik berbasis batubara.

Penelitian keempat berjudul “Applying Analytics in the Energy Industry: A Case Study of Heat Rate and Opacity Prediction in a Coal-Fired Power Plant” (Chongwatpol & Phurithitanapong, 2014) mengembangkan model prediksi heat rate dan opacity emisi pada PLTU batubara di Thailand menggunakan pendekatan data mining berbasis CRISP-DM. Dataset yang digunakan terdiri dari 105 variabel dengan 17.376 record untuk pemodelan heat rate dan 60 variabel dengan 29.952 record untuk pemodelan opacity, yang diperoleh dari sekitar enam bulan operasi pembangkit. Heat rate dalam studi ini memiliki nilai rata-rata 9.519,49 kJ/kWh, dengan rentang minimum 8.481,53 kJ/kWh dan

maksimum 12.535,13 kJ/kWh. Model yang dibandingkan meliputi stepwise regression, decision tree, neural network, dan polynomial regression dengan skema pembagian data 50% training dan 50% validation. Hasil evaluasi menunjukkan bahwa polynomial regression menghasilkan kinerja terbaik untuk prediksi heat rate dengan Average Squared Error (ASE) sebesar 4052,84, diikuti stepwise regression (ASE 5373,79) dan decision tree (ASE 5676,58). Nilai koefisien determinasi model terbaik dilaporkan sekitar $R^2 \approx 63,66\%$. Untuk model opacity, neural network menunjukkan performa terbaik dibanding metode lainnya. Studi ini menegaskan bahwa pendekatan data-driven mampu mengidentifikasi parameter operasional dan karakteristik batubara yang berpengaruh terhadap variasi heat rate dan emisi, sehingga dapat mendukung monitoring performa pembangkit secara lebih sistematis.

Penelitian Kelima “Ross-Veitia et al. (2024) – Machine learning regression algorithms to predict emissions from steam boilers” (Ross-Veitia et al., 2024). Penelitian ini bertujuan memprediksi emisi gas buang (CO, CO₂, NO_x) serta temperatur gas buang pada boiler uap industri menggunakan pendekatan machine learning berbasis data nyata dari 20 boiler fire-tube di Kuba. Variabel input meliputi temperatur lingkungan, tekanan kerja (6–13 kg/cm²), produksi uap (1277–2500 kg/jam), dan jenis bahan bakar (diesel/fuel oil). Dataset dibagi menjadi 70% training dan 30% testing, dengan validasi internal pada model DNN. Empat algoritma dibandingkan, yaitu Multiple Linear Regression (MLR), Deep Neural Network (DNN), Random Forest Regressor (RFR), dan Gradient Boosting Regressor (GBR). Hasil pengujian menunjukkan bahwa GBR memberikan performa terbaik dengan $R^2 = 0,998$ (99,80%), MAE = 0,51, dan RMSE = 2,86, mengungguli DNN ($R^2 = 0,979$; MAE = 2,09; RMSE = 5,52), RFR ($R^2 = 0,997$; MAE = 0,57; RMSE = 2,87), dan MLR ($R^2 = 0,784$; MAE = 11,03; RMSE = 24,76)**. Penelitian ini menegaskan keunggulan metode ensemble boosting dalam menangkap hubungan non-linear kompleks pada proses pembakaran serta menunjukkan bahwa pendekatan data-driven dapat menggantikan pengujian eksperimental berulang yang mahal dan memakan waktu.

Penelitian keenam berjudul “Design Optimization and Dynamic Simulation of Steam Cycle Power Plants: A Review” (Martelli et al., 2021) merupakan studi tinjauan komprehensif mengenai metode optimasi desain dan simulasi dinamis pada sistem siklus uap pembangkit listrik. Artikel ini mengulas pendekatan optimasi berbasis model termodinamika, teknik matematis, serta integrasi metode numerik untuk meningkatkan

efisiensi termal dan fleksibilitas operasional pembangkit. Selain itu, penelitian ini membahas pentingnya simulasi dinamis dalam mengevaluasi respons pembangkit terhadap variasi beban dan kondisi transien, terutama dalam konteks integrasi energi terbarukan yang meningkatkan fluktuasi sistem tenaga. Studi ini menegaskan bahwa pendekatan optimasi konvensional berbasis model fisik memiliki keterbatasan dalam menangani kompleksitas non-linear dan dinamika operasi nyata, sehingga membuka peluang pemanfaatan metode berbasis data dan machine learning untuk meningkatkan akurasi prediksi serta kemampuan monitoring performa secara real-time.

Penelitian ketujuh berjudul “A Dataset of the Operating Station Heat Rate for 806 Indian Coal Plant Units using Machine Learning” (Ding et al., 2024) mengembangkan model prediksi Station Heat Rate (SHR) pada 806 unit PLTU batubara di India dengan total kapasitas terpasang 226 GW, yang terdiri dari 704 unit subcritical dan 102 unit supercritical. SHR didefinisikan sebagai rasio antara energi panas input dan energi listrik yang dihasilkan, sehingga nilai SHR yang lebih rendah menunjukkan efisiensi yang lebih tinggi. Penelitian ini membandingkan berbagai algoritma machine learning, termasuk Gradient Boosting Machine (GBM), XGBoost, Random Forest, Decision Tree, Support Vector Regression (SVR), K-Nearest Neighbor (KNN), Ridge Regression, dan Linear Regression, dengan skema pembagian data 80% untuk pelatihan dan 20% untuk pengujian. Hasil evaluasi menunjukkan bahwa untuk unit subcritical, model GBM memberikan performa terbaik dengan nilai R^2 sebesar 0,810, MAE sebesar 0,105, dan MAPE sebesar 0,849%. Sementara itu, untuk unit supercritical, model KNN menunjukkan performa terbaik dengan R^2 sebesar 0,967, MAE sebesar 0,099, dan MAPE sebesar 0,927%. Rata-rata SHR yang diprediksi adalah 12,91 MMBtu/MWh untuk unit subcritical dan 10,72 MMBtu/MWh untuk unit supercritical. Studi ini menunjukkan bahwa pendekatan machine learning efektif dalam memodelkan heat rate pada skala nasional serta mengidentifikasi faktor-faktor penting seperti kapasitas unit, load factor, dan umur pembangkit melalui analisis feature importance berbasis SHAP.

Penelitian kedelapan berjudul “Heat Rate and Efficiency Evaluation of Cofiring Biomass in CFB Boiler Coal-Fired Power Plant” (Ibnu Prasajo & Darmawan Pasek, n.d.) mengevaluasi pengaruh cofiring biomassa terhadap performa PLTU tipe Circulating Fluidized Bed (CFB) berkapasitas 2×25 MW di Air Anyir, Bangka. Penelitian ini menggunakan data uji performa bulanan dari Oktober 2022 hingga Oktober 2024 dengan

variasi cofiring sebesar 0%, 5,04%, 7,83%, 11,45%, 14,46%, dan 18,97%. Parameter yang dianalisis meliputi Specific Fuel Consumption (SFC), Net Plant Heat Rate (NPHR), Turbine Heat Rate, Boiler Efficiency, dan Thermal Efficiency. Pada kondisi awal commissioning, nilai NPHR (heat balance method) tercatat sebesar 3.178,50 kcal/kWh dan NPHR (input-output method) sebesar 3.446,77 kcal/kWh. Hasil penelitian menunjukkan bahwa peningkatan persentase cofiring menyebabkan kenaikan NPHR, dengan nilai tertinggi metode heat balance mencapai 4.187,56 kcal/kWh pada 11,45% cofiring (kenaikan sekitar 1.009,06 kcal/kWh dari kondisi awal), serta NPHR metode input-output mencapai 4.875,91 kcal/kWh pada 18,97% cofiring. Thermal efficiency mengalami penurunan dari baseline 27,06% menjadi 20,54% pada 11,45% cofiring. Studi ini menunjukkan bahwa meskipun cofiring biomassa mendukung transisi energi dan penurunan emisi, peningkatan fraksi biomassa berpengaruh terhadap kenaikan heat rate dan penurunan efisiensi termal pembangkit.

Penelitian kesembilan, “A Comparison of Machine Learning Models for Predicting Coal-Fired Power Plant Performance” (Hidiyanto et al., 2023). Penelitian ini membandingkan beberapa algoritma machine learning untuk memprediksi thermal efficiency pada PLTU batubara 660 MW dengan pendekatan data-driven berbasis data operasi aktual DCS. Dataset terdiri dari 120 data point hasil audit energi dengan 10 parameter operasional utama, termasuk coal flow rate, air flow rate, temperatur uap utama, tekanan uap, temperatur reheat, kandungan O₂, dan tekanan kondensor. Empat algoritma dibandingkan yaitu Neural Network (NN), Random Forest (RF), AdaBoost (AB), dan Gradient Boosting (GB) menggunakan evaluasi MSE, RMSE, MAE, dan R² dengan skema 80:20 train-test split. Hasil menunjukkan bahwa Neural Network memberikan performa terbaik dengan R² sebesar 0,951 pada data prediksi, RMSE sebesar 0,033, dan MAE sebesar 0,026, sementara RF dan GB menghasilkan R² sebesar 0,933. Studi ini menegaskan bahwa pendekatan machine learning, khususnya neural network, mampu memodelkan efisiensi termal PLTU secara akurat dan berpotensi diterapkan pada skenario co-firing biomassa melalui transfer learning untuk mendukung target net-zero emission.

Penelitian kesepuluh berjudul “Driving towards net-zero from the energy sector: Leveraging machine intelligence for robust optimization of coal and combined cycle gas power stations” (Muhammad Ashraf & Dua, 2024) mengembangkan pendekatan machine

intelligence berbasis neural network untuk meningkatkan efisiensi termal dan menurunkan heat rate pada pembangkit listrik batubara supercritical 660 MW serta combined cycle gas power plant 1180 MW. Studi ini menggunakan pendekatan Data Information integrated Neural Network (DINN), yaitu pengembangan dari Artificial Neural Network (ANN) yang mengintegrasikan informasi korelasi variabel ke dalam proses pelatihan model. Dataset yang digunakan terdiri dari 1.278 observasi untuk pembangkit batubara dan 577 observasi untuk pembangkit siklus gabungan. Model dikembangkan untuk memprediksi parameter performa utama seperti thermal efficiency, power output, dan turbine heat rate (THR). Hasil evaluasi menunjukkan bahwa model mampu mencapai nilai R^2 validasi sebesar 0,76 untuk efisiensi termal, 0,99 untuk daya keluaran, dan 0,72 untuk heat rate pada pembangkit batubara, serta R^2 di atas 0,90 untuk sebagian besar parameter pada pembangkit siklus gabungan. Selanjutnya dilakukan robust optimization berbasis dua tahap pada kondisi beban rendah, menengah, dan penuh, yang menghasilkan potensi penurunan emisi CO₂ hingga 200 ± 10 kt/tahun pada pembangkit batubara. Studi ini menunjukkan bahwa integrasi neural network dengan strategi optimasi robust mampu meningkatkan performa pembangkit secara signifikan serta mendukung transisi menuju sistem energi rendah karbon.

2.1.2. Matriks Penelitian

Tabel 2-1 Matriks Penelitian

No	Peneliti (Tahun)	Fokus Penelitian	Pendekatan / Metode	Parameter Utama	Hasil Utama	Relevansi Penelitian
1	Setyawan & Sudiarto (2024)	Optimasi heat rate dan emisi CO ₂ PLTU Indonesia	LR, Ridge, SVR, RF, GBR, NN, Extra Trees	Boiler type, ambient temp, flue gas temp, unburned carbon	Extra Trees R ² = 0.946 (heat rate)	pembelajaran mesin efektif untuk heat rate Indonesia; belum fokus NPHR unit- level dan belum uji generalisasi temporal

2	Ashraf et al. (2020)	Optimasi efisiensi PLTU 660 MW berbasis Industry 4.0	AI-based optimization	Thermal efficiency, operational parameters	Peningkatan efisiensi termal signifikan	Menguatkan pendekatan data-driven untuk optimasi performa
3	Arferiandi et al. (2021)	Prediksi heat rate CCPP	Artificial Neural Network (ANN)	Parameter uap dan pembakaran	$R^2 = 0.84$	Neural network efektif untuk sistem non-linear
4	Chongwatpol & Phurithititanapong (2014)	Prediksi heat rate dan opacity PLTU	Stepwise regression, Decision Tree, NN, Polynomial Regression	105 variabel operasional	Polynomial regression $R^2 \approx 63.66\%$	Studi awal data mining heat rate; performa masih moderat
5	Ross-Veitía et al. (2024)	Prediksi emisi boiler	MLR, DNN, RF, Gradient Boosting	Temperatur, tekanan, fuel type	GBR $R^2 = 0.998$	Ensemble unggul pada sistem pembakaran non-linear
6	Martelli et al. (2021)	Review optimasi dan simulasi siklus uap	Model fisik & optimasi numerik	Steam cycle parameters	Menyoroti kompleksitas sistem	Menguatkan kebutuhan pembelajaran mesin pada sistem dinamis
7	Ding et al. (2024)	Prediksi Station Heat Rate 806 PLTU India	GBM, XGB, RF, KNN, SVR	Capacity, load factor, plant age	$R^2 = 0.810 - 0.967$	Ensemble efektif untuk heat rate skala besar

8	Prasojo & Pasek (2024)	Evaluasi cofiring biomassa terhadap NPHR	Heat balance & input-output analysis	Fraksi biomassa	NPHR meningkat seiring % biomassa	Bukti kuantitatif pengaruh biomassa terhadap heat rate
9	Hidiyanto et al. (2023)	Komparasi algoritma untuk efisiensi termal PLTU	NN, RF, AdaBoost, Gradient Boosting	Coal flow, air flow, steam temp	NN R2 = 0.951	Komparasi neural vs ensemble; belum spesifik NPHR
10	Ashraf & Dua (2024)	Optimasi heat rate & efisiensi menuju net-zero	DINN (Neural Network)	Thermal efficiency, THR	R2 validation 0.72–0.99	Neural network + robust optimization untuk performa pembangkit

Berdasarkan sepuluh penelitian yang telah dirangkum pada Tabel 2-1, dapat disimpulkan bahwa pendekatan machine learning telah banyak diterapkan dalam pemodelan performa pembangkit listrik, khususnya untuk prediksi heat rate, efisiensi termal, emisi, dan parameter operasional lainnya. Sebagian besar penelitian menunjukkan bahwa algoritma berbasis neural network dan metode *ensemble learning* seperti Gradient Boosting dan Random Forest mampu memberikan tingkat akurasi yang tinggi dalam memodelkan hubungan non-linear pada sistem pembangkit yang kompleks dan dinamis.

Beberapa studi berfokus pada optimasi efisiensi dan penurunan emisi pada PLTU batubara maupun *combined cycle gas turbine*, sementara penelitian lainnya memanfaatkan dataset skala besar untuk memprediksi station heat rate atau mengevaluasi dampak cofiring biomassa terhadap nilai NPHR. Hasil-hasil tersebut secara umum menunjukkan bahwa pendekatan berbasis data (*data-driven modeling*) lebih adaptif

dibandingkan metode konvensional berbasis model fisik dalam menangani kompleksitas sistem dan variasi kondisi operasi.

Namun demikian, terdapat beberapa perbedaan mendasar antara penelitian terdahulu dan penelitian yang dilakukan dalam tesis ini, yaitu:

- a. Sebagian besar penelitian terdahulu berfokus pada optimasi efisiensi atau emisi, bukan secara spesifik pada prediksi Net Plant Heat Rate (NPHR) sebagai indikator efisiensi bersih pada level unit pembangkit.
- b. Penelitian sebelumnya umumnya menggunakan satu pendekatan algoritma dominan, seperti Artificial Neural Network (ANN) atau metode ensemble tertentu, tanpa melakukan perbandingan langsung dan terstruktur antara dua algoritma dengan karakteristik berbeda, khususnya Multi-Layer Perceptron (MLP) dan Gradient Boosting Regressor (GBR).
- c. Sebagian penelitian dilakukan pada skala nasional atau lintas unit pembangkit, sehingga belum merepresentasikan dinamika operasional unit PLTU tertentu secara detail, terutama dalam konteks data aktual berbasis Distributed Control System (DCS).
- d. Kajian yang memasukkan variabel cofiring biomassa dalam pemodelan berbasis machine learning masih terbatas, meskipun telah ada studi kuantitatif mengenai pengaruh biomassa terhadap NPHR melalui pendekatan heat balance.
- e. Evaluasi kemampuan generalisasi temporal model masih jarang dilakukan, khususnya pengujian terhadap periode operasi berbeda (*cold test*) untuk menilai ketahanan model terhadap perubahan pola beban dan kondisi operasi.

2.1.3. Kesenjangan Penelitian

Berdasarkan tinjauan pustaka terhadap sepuluh penelitian yang telah dibahas sebelumnya, teridentifikasi beberapa kesenjangan (*research gap*) yang menjadi landasan utama dilakukannya penelitian ini. Meskipun pemodelan performa pembangkit listrik berbasis machine learning telah banyak dikembangkan, terdapat sejumlah aspek yang masih belum dieksplorasi secara mendalam, khususnya dalam konteks prediksi NPHR pada PLTU Indonesia.

a. Kesenjangan Metodologi

Sebagian besar penelitian terdahulu menggunakan pendekatan neural network atau ensemble learning secara terpisah untuk memodelkan heat rate, efisiensi termal, maupun emisi pembangkit. Namun, analisis komparatif yang secara langsung dan sistematis membandingkan dua algoritma dengan karakteristik berbeda, yaitu Multi-Layer Perceptron (MLP) sebagai representasi jaringan saraf non-linear dan Gradient Boosting Regressor (GBR) sebagai metode ensemble berbasis boosting, masih sangat terbatas. Selain itu, sebagian studi lebih menekankan pada optimasi performa atau emisi, bukan pada evaluasi prediktif berbasis metrik kesalahan seperti RMSE, MAE, dan R^2 secara komprehensif. Kesenjangan ini menunjukkan perlunya penelitian yang membandingkan kedua pendekatan tersebut dalam kerangka evaluasi yang terstruktur dan objektif.

b. Kesenjangan Kontekstual

Beberapa penelitian dilakukan pada pembangkit di luar Indonesia atau menggunakan dataset skala nasional lintas unit pembangkit. Pendekatan tersebut belum sepenuhnya merepresentasikan dinamika operasional unit PLTU tertentu yang memiliki karakteristik desain, kualitas batubara domestik, serta pola operasi beban yang spesifik. Studi yang secara eksplisit menggunakan data operasi aktual berbasis Distributed Control System (DCS) dari unit PLTU Indonesia dan melakukan analisis komparatif algoritma masih belum banyak ditemukan. Oleh karena itu, diperlukan penelitian yang lebih kontekstual pada level unit pembangkit untuk meningkatkan relevansi implementatif.

c. Kesenjangan Fokus Variabel (Biomassa & NPHR)

Meskipun telah terdapat penelitian yang mengevaluasi pengaruh cofiring biomassa terhadap heat rate melalui pendekatan heat balance, integrasi variabel biomassa ke dalam model prediksi berbasis machine learning masih terbatas. Selain itu, sebagian studi berfokus pada Station Heat Rate (SHR) atau efisiensi termal, bukan secara spesifik pada Net Plant Heat Rate (NPHR) yang merepresentasikan efisiensi bersih pada level pembangkit. Hal ini

menunjukkan perlunya pemodelan NPHR yang secara eksplisit memasukkan fraksi biomassa sebagai variabel input.

d. Kesenjangan Evaluasi Generalisasi Model (Temporal Generalization Gap)

Sebagian besar penelitian terdahulu menggunakan pembagian data konvensional tanpa menguji ketahanan model terhadap perubahan periode operasi. Evaluasi kemampuan generalisasi temporal, seperti pengujian pada periode waktu berbeda (*cold test*), masih jarang dilakukan. Padahal, dalam sistem pembangkit yang dinamis dan dipengaruhi fluktuasi beban serta variasi kualitas bahan bakar, kemampuan model untuk mempertahankan akurasi pada periode operasi yang berbeda menjadi aspek krusial dalam implementasi nyata.

Berangkat dari kesenjangan tersebut, penelitian ini secara khusus:

- a. Membandingkan secara langsung kinerja MLP dan GBR dalam memprediksi NPHR pada PLTU Indonesia.
- b. Menggunakan data operasi aktual berbasis DCS dengan interval waktu tetap.
- c. Memasukkan variabel biomassa sebagai faktor relevan dalam estimasi NPHR.
- d. Melakukan evaluasi menggunakan RMSE, MAE, dan R^2 secara komprehensif.
- e. Menguji kemampuan generalisasi temporal model melalui beberapa skema pembagian data serta pengujian periode operasi baru.

Dengan demikian, penelitian ini tidak hanya memperkaya kajian pemodelan performa pembangkit berbasis machine learning, tetapi juga memberikan kontribusi metodologis dan kontekstual dalam pengembangan sistem monitoring efisiensi PLTU Indonesia yang lebih adaptif, akurat, dan implementatif.

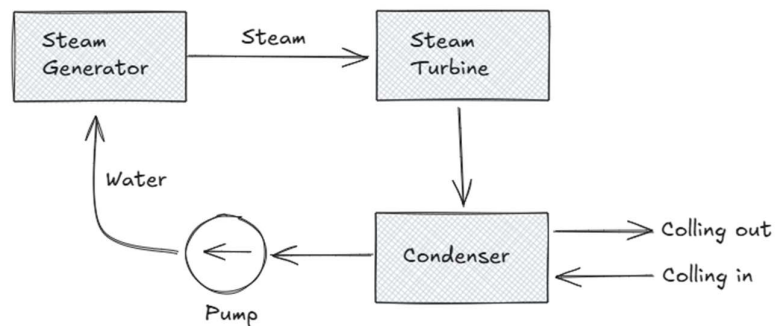
2.2. Landasan Teori

Landasan Teori membahas teori-teori yang berkaitan atau berhubungan dengan penelitian yang dilakukan. Berikut dijelaskan teori pada penulisan ini:

2.2.1. Pembangkit Listrik Tenaga Uap (PLTU)

Pembangkit Listrik Tenaga Uap (PLTU) merupakan jenis pembangkit listrik termal yang bekerja dengan mengubah energi panas hasil pembakaran bahan bakar menjadi energi listrik. Proses konversi energi tersebut dilakukan melalui suatu siklus termodinamika tertutup, di mana fluida kerja mengalami proses pemanasan, ekspansi, dan kondensasi secara berulang.

Menurut Sarkar (2015), terdapat beberapa jenis siklus termodinamika yang digunakan pada pembangkit tenaga termal, seperti siklus Rankine, Brayton, dan Kalina, yang berbeda dalam hal fluida kerja dan mekanisme konversi energinya. Di antara ketiganya, siklus Rankine merupakan yang paling umum diterapkan pada PLTU karena kesederhanaan rancangan dan efisiensinya yang tinggi untuk sistem berbasis uap air (Sarkar, 2015).



Gambar 2-1 Siklus uap paling sederhana (simple Rankine cycle) (Sarkar, 2015)

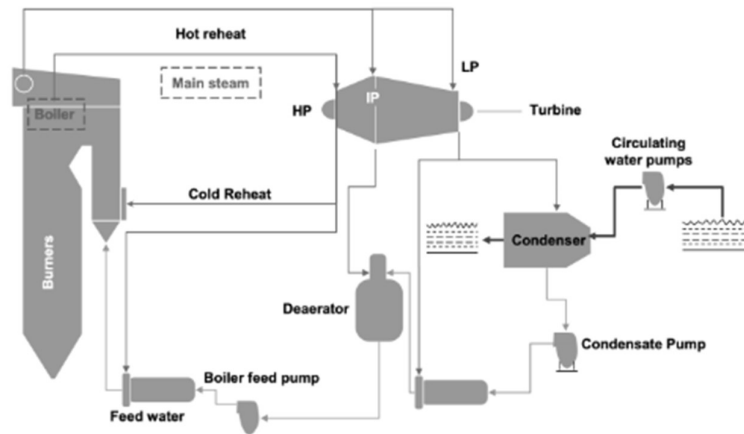
2.2.1.1. Landasan konseptual termodinamika

Siklus Rankine merupakan siklus termodinamika yang mendemonstrasikan proses konversi energi panas menjadi kerja mekanis melalui sirkulasi tertutup fluida kerja, umumnya air (Sarkar, 2015). Siklus ini dinamai dari ilmuwan Skotlandia, William John Mac Eown Rankine, yang berkontribusi pada pengembangan teori termodinamika siklus tenaga uap. Siklus Rankine paling sederhana seperti pada Gambar 2-1 yang menggunakan uap jenuh kering (dry saturated steam) sebagai fluida kerja, terdiri atas empat komponen utama: steam generator (boiler), steam turbine (turbin uap), kondenser (kondensor), dan pump (pompa air umpan). Keempat komponen ini terhubung dalam sirkulasi termodinamika tertutup. Dalam siklus ini, air dipanaskan hingga menjadi uap di dalam boiler, kemudian uap tersebut melakukan ekspansi pada turbin untuk menghasilkan

kerja mekanis sebelum didinginkan dan dikondensasikan kembali menjadi cairan untuk disirkulasikan ulang.

2.2.1.2. Prinsip Kerja PLTU

PLTU bekerja dengan mengandalkan siklus tertutup fluida kerja berupa air atau uap. Untuk memberikan gambaran lebih nyata mengenai sistem kerja PLTU, Gambar 2-2 menunjukkan skema aliran utama (main steam cycle) pada pembangkit listrik tenaga uap batubara berkapasitas 400 MW dengan konfigurasi subkritis yang menggunakan boiler tipe pulverized coal (PC) dan turbin uap bertingkat tekanan tinggi (HP), menengah (IP), dan rendah (LP).



Gambar 2-2 Diagram aliran utama pada PLTU batubara subkritis 400 MW (diadaptasi dari diagram proses yang tersedia di PLTU Suralaya).

Proses konversi energi berlangsung dalam beberapa tahap (Pambudi et al., 2023; Sarkar, 2015).

1. Pembakaran dan Pemanasan Air di Boiler

Energi kimia bahan bakar dibakar dan menghasilkan panas yang digunakan untuk memanaskan air di boiler hingga menjadi uap bertekanan dan suhu tinggi. Proses ini meliputi pemompaan air feedwater ke tekanan tinggi sebelum pemanasan, yang secara signifikan meningkatkan efisiensi pembangkit.

2. Ekspansi Uap di Turbin

Uap bertekanan dan bersuhu tinggi kemudian diarahkan ke turbin, di mana ekspansi uap mengubah energi potensial menjadi energi kinetik, memutar turbin

dan menghasilkan kerja mekanik yang sangat bergantung pada kondisi uap masuk.

3. Konversi Kerja Mekanik ke Listrik di Generator

Turbin yang berputar menggerakkan rotor generator yang kemudian menghasilkan energi listrik yang dapat digunakan secara komersial.

4. Kondensasi Uap Bekas Turbin

Uap bekas turbin didinginkan secara efisien di kondensor untuk kembali menjadi air (kondensat). Cairan ini kemudian dipanaskan melalui heat exchanger seperti feedwater heater atau economizer sebelum dikirim kembali ke boiler untuk memulai siklus ulang.

2.2.1.3. Faktor-Faktor yang Mempengaruhi Efisiensi Operasional

PLTU bekerja dengan mengandalkan siklus tertutup fluida kerja berupa air atau uap. Proses konversi energi berlangsung dalam beberapa tahap (Sarkar, 2015):

1. Temperatur dan Tekanan Uap Masuk Turbin

Semakin tinggi temperatur dan tekanan (dalam batas material dan desain), semakin besar potensi kerja yang dapat dihasilkan dari uap. Studi menunjukkan bahwa peningkatan tekanan dan temperatur uap secara signifikan meningkatkan efisiensi turbin (Yanti et al., 2024).

2. Tekanan Kondensor dan Suhu Air Pendingin

Ketika suhu air pendingin naik atau tekanan kondensor naik, gradien temperatur uap ke ambient berkurang, sehingga efisiensi termal menurun. Sebagai contoh, sebuah studi menunjukkan bahwa peningkatan suhu air masuk dari 12 °C ke 42 °C menyebabkan penurunan efisiensi termal dari ~36 % menjadi ~34,5 % (Prihananto, 2024).

3. Kualitas Bahan Bakar dan Kontrol Pembakaran

Efisiensi boiler sangat dipengaruhi oleh kualitas bahan bakar (kalori, kelembapan, abu) serta kontrol udara pembakaran, efisiensi heat-transfer dan pengurangan losses pada boiler. Studi terkait menggarisbawahi pentingnya optimasi kualitas bahan bakar dan kontrol pembakaran untuk menjaga efisiensi (Aditya et al., 2025; Mubarak et al., 2024).

4. Beban Turbin dan Generator

Efisiensi operasional turbin dan generator terbaik dicapai ketika beroperasi pada beban optimal yang sesuai dengan desain. Pada beban yang jauh di bawah desain, efisiensi menurun karena rasio kerugian tetap seperti auxiliary dan pompa menjadi lebih dominan. Beberapa studi menunjukkan bahwa operasi pada beban optimal dapat menjaga efisiensi kerja turbin dan pembangkit listrik secara keseluruhan (Setyawan & Sudiarto, 2024).

5. Instrumentasi dan Kontrol

Sistem monitoring dan kontrol real-time telah sangat meningkatkan kemampuan pembangkit listrik untuk memantau dan mengendalikan emisi dengan akurasi dan efisiensi tinggi, memungkinkan manajemen emisi yang proaktif dan kepatuhan regulasi yang lebih baik (Zuo & Niu, 2024).

2.2.1.4. Karakteristik PLTU Batubara Indonesia

PLTU batubara mendominasi sistem kelistrikan nasional Indonesia dengan kontribusi sekitar 60% dari total kapasitas terpasang, sebagaimana tercantum dalam RUPTL PLN 2025-2034. Karakteristik operasional PLTU Indonesia unik karena ketergantungan pada batubara domestik low rank coal (LRC) yang mencakup ~65% cadangan nasional, berupa lignite dan sub-bituminous dengan kalori rendah 4.000-4.700 kcal/kg, kadar air tinggi 25-30%, serta abu 3-5%. Kualitas batubara ini menyebabkan heat rate lebih tinggi dan efisiensi operasional lebih rendah dibandingkan PLTU berbahan bakar batubara impor berkalori tinggi (Hermawan et al., 2023).

Pola operasi PLTU Indonesia juga khas dengan load following untuk menyesuaikan fluktuasi beban, penerapan co-firing biomassa bertahap (target 10% sesuai RUKN/RUPTL), serta banyak unit generasi lama berumur >20 tahun dengan efisiensi <35% yang akan dipensiunkan hingga 2045 (PT PLN (Persero), 2025).

Karakteristik LRC, pola load following, dan co-firing biomassa ini menjadikan prediksi NPHR pada PLTU Indonesia berbeda signifikan dari studi luar negeri, sehingga memerlukan model machine learning berbasis data operasi domestik yang mempertimbangkan variabilitas lokal.

2.2.2. Heat Rate sebagai Indikator Kinerja

Heat Rate merupakan salah satu parameter utama dalam mengukur kinerja PLTU. Heat Rate didefinisikan sebagai jumlah energi panas yang diperlukan untuk menghasilkan sejumlah energi listrik tertentu, biasanya dinyatakan dalam satuan kJ/kWh atau Btu/kWh. Semakin kecil nilai Heat Rate, semakin efisien pembangkit tersebut, karena berarti konsumsi energi untuk menghasilkan listrik lebih rendah.

Secara umum, hubungan antara energi input (panas) dan energi output (listrik) dapat dinyatakan dengan persamaan:

$$\text{Heat Rate} = \frac{\text{Total Heat Input (kCal atau BTU)}}{\text{Net Power Output (kWh)}} \quad (2-1)$$

Nilai Heat Rate biasanya dinyatakan dalam satuan kCal/kWh atau BTU/kWh. Sementara itu, efisiensi termal pembangkit listrik (η) berkaitan terbalik dengan Heat Rate, yang dapat dinyatakan sebagai:

$$\eta = \frac{860}{\text{Heat Rate (kCal/kWh)}} = \frac{3412}{\text{Heat Rate (BTU/kWh)}} \quad (2-2)$$

Rasio ini memperlihatkan bahwa peningkatan efisiensi termal akan secara langsung menurunkan nilai Heat Rate (Moran & Shapiro, 2006). Penurunan Heat Rate sebesar 1% akan meningkatkan efisiensi termal dalam proporsi yang hampir sama. Dengan demikian, Heat Rate menjadi indikator langsung yang mencerminkan seberapa efisien energi panas bahan bakar dikonversi menjadi energi listrik.

2.2.2.1. Faktor-faktor yang Mempengaruhi Heat Rate

Faktor-faktor yang memengaruhi efisiensi operasional sebagaimana telah dijelaskan pada Subbab 2.1.3 juga berperan langsung terhadap nilai Heat Rate. Namun, pada bagian ini pembahasan difokuskan pada aspek kuantitatif dari efisiensi, yaitu bagaimana variasi kondisi operasional dan karakteristik sistem pembangkit tercermin dalam perubahan nilai Heat Rate. Indikator ini digunakan secara luas dalam industri pembangkitan untuk menilai kinerja termal secara numerik dan menjadi dasar analisis lebih lanjut dalam model prediksi efisiensi berbasis data.

Nilai Heat Rate dipengaruhi oleh berbagai kondisi operasional dan karakteristik teknis pembangkit, di antaranya:

Tabel 2-2 Faktor-faktor yang memengaruhi nilai Heat Rate pada PLTU (diadaptasi dari (Çengel et al., 2019; Sarkar, 2015)).

Faktor	Deskripsi Pengaruh terhadap Heat Rate
Beban generator (load)	Heat Rate cenderung meningkat saat beban rendah karena efisiensi sistem menurun akibat losses tetap (fixed losses).
Kualitas bahan bakar (coal quality)	Kandungan air, abu, dan nilai kalor batubara sangat memengaruhi jumlah energi panas yang tersedia per satuan massa bahan bakar.
Tekanan dan temperatur uap utama	Semakin tinggi tekanan dan temperatur uap, semakin besar energi entalpi yang dapat diekstraksi oleh turbin sehingga Heat Rate menurun.
Efisiensi pembakaran dan perpindahan panas di boiler	Ketidaksempurnaan pembakaran atau kerak pada pipa boiler dapat meningkatkan kehilangan panas.
Kinerja kondensor dan sistem pendingin	Tekanan kondensor yang terlalu tinggi menurunkan efisiensi ekspansi turbin dan meningkatkan Heat Rate.
Kehilangan energi pada sistem bantu (auxiliary systems)	Konsumsi daya pada fan, pompa, dan peralatan bantu lainnya turut memengaruhi nilai Heat Rate bersih (Net Plant Heat Rate).

2.2.2.2. Peran Heat Rate dalam Evaluasi dan Optimisasi PLTU

Dalam praktik industri, Heat Rate digunakan oleh operator dan manajemen pembangkit untuk memantau efisiensi operasional dan menjaga kinerja unit pembangkit. Peningkatan nilai Heat Rate umumnya mengindikasikan penurunan efisiensi yang dapat disebabkan oleh berbagai faktor, seperti degradasi atau kerusakan peralatan, penurunan kualitas bahan bakar, ketidakseimbangan beban operasi, maupun perubahan kondisi lingkungan. Oleh karena itu, pemantauan nilai Heat Rate secara berkelanjutan menjadi sangat penting untuk mendeteksi dan mengantisipasi potensi penyimpangan performa yang dapat menurunkan efisiensi pembangkit.

Dalam perhitungan praktis, nilai Heat Rate diperoleh dari perbandingan antara konsumsi energi bahan bakar dan produksi energi listrik selama interval waktu tertentu. Dengan demikian, analisis Heat Rate merupakan bagian krusial dalam manajemen energi dan strategi peningkatan efisiensi PLTU.

Pada era modern, pemantauan dan estimasi Heat Rate tidak lagi dilakukan secara manual melalui uji performa periodik (performance test) saja, melainkan juga melalui

pemanfaatan data operasi aktual yang dikumpulkan secara otomatis dari berbagai sistem kontrol dan sensor, seperti Plant Information System (PIS), Supervisory Control and Data Acquisition (SCADA), maupun Distributed Control System (DCS). Integrasi teknologi ini memungkinkan analisis Heat Rate secara real-time, sehingga operator dapat mengoptimalkan kondisi operasi pembangkit dan mempertahankan efisiensi termal pada tingkat yang optimal (Ding et al., 2024).

2.2.2.3. Net Plant Heat Rate (NPHR) sebagai Indikator Utama

Net Plant Heat Rate (NPHR) merupakan indikator efisiensi operasional pembangkit secara keseluruhan yang mengukur energi panas bersih dari bahan bakar yang dibutuhkan untuk menghasilkan 1 kWh listrik bersih yang disalurkan ke jaringan (Sarkar, 2015). Berbeda dengan *Gross Heat Rate* yang hanya mempertimbangkan input bahan bakar dan output turbin, NPHR memasukkan *parasitic load* atau konsumsi listrik internal pembangkit (*auxiliary power consumption*) (Burnett & Kiesling, 2019).

Rumus NPHR:

$$\text{NPHR} = \frac{\text{Net Electricity Output to Grid (kWh)}}{\text{Total Heat Input from Fuel (MJ)}} \quad (2-3)$$

di mana Net Electricity Output = Gross Generation - Auxiliary Consumption

Keunggulan NPHR untuk Monitoring Real-time:

- a. Representatif kinerja unit lengkap (bukan hanya boiler/turbin)
- b. Sensitif terhadap co-firing biomassa (perubahan calorific value + auxiliary load)

Karena NPHR mencakup *parasitic load* yang dinamis (dipengaruhi beban, co-firing, kondisi peralatan), prediksi NPHR lebih kompleks dibanding heat rate parsial, tetapi lebih relevan untuk pengambilan keputusan operator PLTU [Martelli et al., 2021].

2.2.3. Variabel Operasional PLTU sebagai Data Input Model

Dalam analisis dan pemodelan kinerja pembangkit listrik tenaga uap (PLTU), pemilihan variabel operasional merupakan aspek fundamental yang menentukan kualitas representasi sistem. Variabel-variabel ini mencerminkan kondisi operasi aktual dari subsistem utama pembangkit, seperti boiler, turbin, generator, sistem udara-gas

pembakaran, serta sistem pendingin. Oleh karena itu, karakteristik dan dinamika variabel operasional memiliki keterkaitan langsung dengan performa termal dan nilai Heat Rate yang dihasilkan.

Data operasional tersebut umumnya diperoleh dari sistem kontrol dan pemantauan pembangkit, seperti Supervisory Control and Data Acquisition (SCADA), Distributed Control System (DCS), atau Plant Information System (PIS), yang merekam parameter proses secara berkala dan kontinu. Sistem ini memungkinkan tersedianya data historis dengan resolusi waktu yang tinggi, sehingga dapat digunakan untuk analisis berbasis data dan pengembangan model prediktif.

Secara konseptual, variabel-variabel yang digunakan dalam penelitian ini merepresentasikan fenomena termodinamika pada siklus Rankine, yang menjadi dasar kerja PLTU. Pemilihan variabel dilakukan berdasarkan relevansi teoretis terhadap efisiensi termal serta potensi pengaruhnya terhadap perubahan nilai Heat Rate. Dengan demikian, setiap variabel tidak hanya dipilih berdasarkan ketersediaan data, tetapi juga atas pertimbangan hubungan fisik dan proses energi di dalam sistem pembangkit.

Secara umum, variabel operasional dapat dikelompokkan ke dalam tiga kategori utama, yaitu:

- a. Parameter Beban dan Pembangkitan, yang mencerminkan tingkat daya keluaran dan kondisi pembebanan unit.
- b. Parameter Udara, Gas, dan Pembakaran, yang berkaitan dengan proses pembakaran bahan bakar serta kualitas suplai udara dan aliran gas buang.
- c. Parameter Boiler, Steam, dan Cooling, yang merepresentasikan kondisi uap utama, tekanan, temperatur, serta efektivitas sistem pendingin.

Pengelompokan ini bertujuan untuk memberikan struktur konseptual terhadap faktor-faktor yang memengaruhi Heat Rate, sekaligus menjadi dasar dalam analisis hubungan sebab-akibat yang akan dibahas pada bagian berikutnya.

2.2.3.1. Parameter Beban dan Pembangkitan

Kelompok pada Tabel 2-3 ini menggambarkan tingkat keluaran daya dan kondisi kerja generator, yang merupakan faktor langsung terhadap efisiensi keseluruhan unit. Beban generator memengaruhi laju aliran bahan bakar, tekanan uap, serta proporsi energi yang digunakan untuk mengatasi kehilangan tetap atau *fixed losses*.

Tabel 2-3 Parameter beban dan pembangkitan listrik yang dapat digunakan sebagai variabel input dalam model prediksi Heat Rate PLTU.

Parameter	Simbol /Satuan	Keterangan
Daya keluaran generator	MW	Output listrik bersih yang dihasilkan unit.
Arus dan tegangan generator	A, kV	Indikator kestabilan pembangkitan listrik.
Frekuensi sistem	Hz	Menunjukkan kestabilan operasi sistem kelistrikan.
Faktor daya (power factor)	–	Rasio antara daya aktif dan daya semu; terkait efisiensi konversi energi.

Kinerja optimal PLTU umumnya dicapai saat beroperasi mendekati beban penuh, karena efisiensi termal meningkat dan kehilangan energi relatif menurun dibanding kondisi beban parsial.

2.2.3.2. Parameter Udara, Gas, dan Pembakaran

Sistem udara dan gas berperan penting dalam proses pembakaran batubara serta perpindahan panas di boiler seperti pada Tabel 2-4. Ketidakseimbangan aliran udara, tekanan fan, atau temperatur gas buang dapat memengaruhi efisiensi pembakaran dan meningkatkan nilai Heat Rate.

Tabel 2-4 Parameter udara, gas, dan pembakaran yang memengaruhi proses konversi energi dan efisiensi boiler pada PLTU.

Parameter	Sistem Terkait	Keterangan
Laju aliran udara primer dan sekunder	PA/SA Flow (t/h)	Menentukan suplai oksigen untuk proses pembakaran.
Tekanan udara di saluran masuk dan keluar fan	FD/ID/PA (kPa)	Mempengaruhi distribusi udara di ruang bakar dan sistem gas buang.
Temperatur udara keluar air heater	°C	Mempengaruhi efisiensi penyerapan panas di boiler.
Konsentrasi O ₂ gas buang	%	Indikator efisiensi pembakaran; kelebihan udara meningkatkan kehilangan panas.
Tekanan gas buang	kPa	Berhubungan dengan kerja fan dan kehilangan energi di sistem udara-gas.

Sistem udara dan gas juga berkontribusi pada pembentukan emisi dan efisiensi keseluruhan pembakaran, sehingga pemantauan parameter-parameter ini menjadi penting dalam menjaga efisiensi termal dan kepatuhan lingkungan.

2.2.3.3. Parameter Boiler, Steam, dan Cooling

Kelompok parameter pada Tabel 2-5 ini mencakup variabel yang secara langsung menentukan efisiensi siklus termodinamika Rankine, karena berkaitan dengan tekanan, temperatur, dan laju aliran uap serta performa kondensasi. Perubahan kecil pada parameter-parameter ini dapat berdampak signifikan terhadap nilai Heat Rate.

Tabel 2-5 Parameter sistem boiler, uap, dan pendinginan yang dapat digunakan dalam pemodelan efisiensi termal dan Heat Rate PLTU.

Parameter	Sistem Terkait	Keterangan
Tekanan uap utama (main steam pressure)	MPa	Menentukan energi potensial uap untuk ekspansi di turbin.
Temperatur uap utama (main steam temperature)	°C	Mempengaruhi efisiensi ekspansi dan kerja turbin.
Tekanan & temperatur reheater	MPa, °C	Menjaga efisiensi turbin tahap menengah dan bawah.
Laju aliran air umpan (feedwater flow)	t/h	Mempengaruhi keseimbangan energi antara boiler dan kondensor.
Temperatur air umpan (feedwater temperature)	°C	Dipengaruhi oleh efektivitas economizer dan pemanas regeneratif.
Kondisi vakum kondensor (condenser vacuum)	kPa	Semakin rendah tekanan kondensor, semakin tinggi efisiensi termal.
Temperatur air pendingin masuk & keluar	°C	Indikator kinerja sistem pendingin; memengaruhi tekanan kondensor.

Parameter-parameter tersebut menjadi fitur utama dalam model Machine Learning karena memiliki korelasi langsung dengan efisiensi termal dan konsumsi energi bahan bakar. Kombinasi parameter ini memungkinkan model untuk menangkap dinamika perubahan Heat Rate secara real-time berdasarkan kondisi operasi aktual (Ashraf et al., 2025; Ding et al., 2024).

2.2.3.4. Parameter Co-firing Biomassa

Co-firing biomassa merupakan pencampuran biomassa dengan batubara pada PLTU untuk mengurangi emisi karbon dan meningkatkan porsi energi terbarukan, sesuai target RUKN 2060 dan RUPTL PLN yang mewajibkan co-firing bertahap hingga 10% pada PLTU eksisting. Program ini sudah diimplementasikan di 20 lebih PLTU Indonesia dengan biomassa dari limbah pertanian seperti bagasse tebu, sawdust, dan serbuk kayu (Kementerian ESDM, 2025).

Data empiris dari 4 PLTU pada Tabel 2-6 menunjukkan persentase rendah (1-3%) dengan dampak minimal pada NPHR jika pre-mixing homogen:

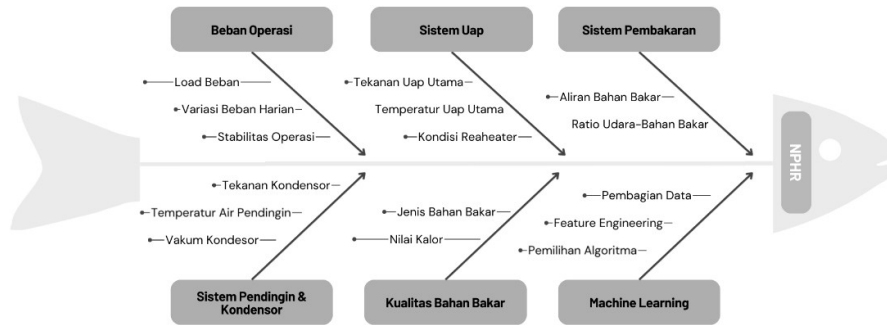
Tabel 2-6 Dampak Co-firing terhadap NPHR

PLTU	Biomassa	% dari Total Bahan Bakar	NPHR/SFC	Efisiensi	Sumber
Suralaya	Sawdust	1.62%	Optimal	85.55%	(Sidiq, 2022)
Rembang	Bagasse	2.5%	0.641 kg/kWh	Stabil	(Riris Destalia & Enny Aryanny, 2024)
Jeranjang	Sawdust	2-5%	↓0.94→0.91 kg/kWh	↑75.23%	(Legowo et al., 2025)
Pangkalan Susu	Sawdust	3%	↑3.56→3.70 kcal/kWh	↓ ringan	(Putri Lubis & Umma Tambunan, 2025)

Meskipun target nasional 10%, data empiris hanya 1-3%. Belum tersedia model prediksi NPHR pada fraksi biomassa 5-10% untuk transisi bertahap. Machine learning diperlukan untuk ekstrapolasi aman dari data historis (0-3%) ke target nasional, mengantisipasi non-linear impact terhadap efisiensi.

2.2.4. Fishbone (Ishikawa Diagram)

Menurut Li et al. (2024), Fishbone Diagram merupakan alat konseptual yang digunakan untuk mengidentifikasi, mengelompokkan, dan memetakan berbagai faktor penyebab terhadap suatu fenomena secara terstruktur. Diagram ini membantu peneliti menyusun faktor-faktor yang kompleks ke dalam kategori utama dan sub-faktor, sehingga hubungan sebab-akibat dapat divisualisasikan secara sistematis dan logis.



Gambar 2-3 Fishbone Kerangka Pemikiran Penelitian

Hubungan antara variabel operasional dan NPHR bersifat kompleks dan non-linear. Nilai NPHR tidak dipengaruhi oleh satu faktor tunggal, melainkan merupakan hasil interaksi berbagai parameter teknis dalam sistem pembangkit. Untuk memetakan hubungan sebab-akibat tersebut secara sistematis, penelitian ini menggunakan pendekatan diagram Fishbone (Cause-and-Effect Diagram).

Pada penelitian ini, NPHR diposisikan sebagai variabel akibat (effect), sedangkan faktor-faktor operasional PLTU sebagai variabel penyebab (cause), yang meliputi:

- Faktor Beban Operasi (daya keluaran dan stabilitas beban)
- Faktor Sistem Uap (tekanan dan temperatur uap)
- Faktor Sistem Pembakaran (aliran bahan bakar dan rasio udara-bahan bakar)
- Faktor Sistem Pendingin dan Kondensor (tekanan kondensor dan temperatur pendingin)
- Faktor Kualitas Bahan Bakar (nilai kalor dan kadar air)
- Machine Learning

Kelompok faktor tersebut menunjukkan bahwa perubahan NPHR merupakan hasil interaksi multidimensi yang tidak sepenuhnya dapat dijelaskan melalui pendekatan linier sederhana. Oleh karena itu, diperlukan metode machine learning untuk memodelkan hubungan kompleks antar variabel secara lebih komprehensif.

2.2.5. Machine Learning

Perubahan nilai Heat Rate yang dipengaruhi oleh banyak parameter operasional menjadikan analisis efisiensi PLTU bersifat kompleks dan non-linear. Hubungan antarvariabel seperti tekanan, temperatur, beban, dan laju aliran udara tidak selalu dapat dijelaskan secara eksplisit melalui persamaan termodinamika konvensional. Oleh karena itu, dibutuhkan pendekatan berbasis data yang mampu mengenali pola tersembunyi dan memodelkan hubungan kompleks antarvariabel operasi secara adaptif. Pendekatan tersebut difasilitasi oleh kemajuan teknologi Machine Learning dan Data Mining yang kini banyak digunakan dalam sistem analitik pembangkit modern (Alpaydin, 2004; Han et al., 2023).

2.2.5.1. Konsep Dasar dan Jenis

Machine Learning dan Data Mining merupakan dua pendekatan komputasional yang saling terkait dalam analisis data modern. Keduanya memiliki tujuan untuk memanfaatkan data untuk memperoleh pola, pengetahuan, dan prediksi yang berguna, namun berbeda dalam fokus utama:

- a. Machine Learning berfokus pada pengembangan model prediktif untuk mempelajari hubungan input–output dari data. Menurut (Alpaydin, 2004), Machine Learning merupakan bagian dari kecerdasan buatan yang memungkinkan sistem belajar dari pengalaman.
- b. Data Mining, menurut (Han et al., 2023), adalah bagian dari Knowledge Discovery in Databases (KDD), yaitu proses sistematis untuk mengekstraksi pola bermakna dari kumpulan data besar. Data Mining lebih menekankan pada penemuan pola tersembunyi, asosiasi, dan hubungan antarvariabel tanpa harus selalu membuat model prediktif.

Pendekatan komputasional berbasis data telah banyak diterapkan untuk mengekstraksi informasi dari data yang memiliki karakteristik kompleks dan sulit dianalisis secara manual. Agtriadi et al. (2023) menunjukkan bahwa metode komputasional mampu merepresentasikan pola tersembunyi dari data kompleks, yang menegaskan pentingnya pendekatan berbasis data dalam memahami sistem dengan tingkat kompleksitas tinggi (Agtriadi et al., 2023).

Sejalan dengan perkembangan tersebut, penerapan machine learning dalam bentuk supervised learning telah banyak digunakan pada berbagai permasalahan prediksi dan klasifikasi. Contohnya adalah penerapan jaringan syaraf tiruan berbasis backpropagation untuk memodelkan hubungan nonlinier antarvariabel pada data nyata, yang menunjukkan kemampuan Artificial Neural Network dalam menangkap pola kompleks yang sulit direpresentasikan oleh model linier (Rohayani & Umam, 2022). Selain pendekatan berbasis jaringan syaraf tiruan, model supervised learning berbasis probabilistik seperti Naive Bayes juga digunakan dalam sistem prediksi berbasis data historis untuk menangani ketidakpastian dan variasi data, sebagaimana ditunjukkan pada penerapan credit scoring untuk penilaian risiko (Afandi et al., 2024). Meskipun berada pada domain yang berbeda, prinsip pemodelan hubungan input–output pada pendekatan tersebut relevan secara konseptual dengan penelitian ini.

Secara umum, metode Machine Learning dibagi menjadi tiga kategori utama:

- a. Supervised Learning: model dilatih menggunakan data berlabel, di mana input dan output diketahui. Contoh: regresi, klasifikasi.
- b. Unsupervised Learning: digunakan untuk menemukan struktur atau kelompok tersembunyi dalam data tanpa label, misalnya clustering atau dimensionality reduction.
- c. Reinforcement Learning: sistem belajar melalui interaksi dengan lingkungan dan memperoleh umpan balik berupa reward atau penalty.

Untuk kasus estimasi Heat Rate, pendekatan yang digunakan termasuk dalam kategori supervised learning dengan model regresi, karena target yang diprediksi berupa nilai kontinu (nilai Heat Rate dalam kCal/kWh atau BTU/kWh).

2.2.5.2. Penerapan Machine Learning dalam Industri Energi

Dalam industri energi, Machine Learning telah digunakan secara luas untuk mendukung optimisasi operasi, pemeliharaan prediktif, dan efisiensi energi. Beberapa contoh penerapan nyata meliputi:

- a. Prediksi performa peralatan seperti boiler, turbin, dan kondensor berdasarkan data sensor.

- b. Deteksi anomali operasi melalui analisis pola data real-time dari sistem kontrol (SCADA atau DCS).
- c. Optimisasi konsumsi bahan bakar dengan mengidentifikasi kombinasi parameter operasi yang menghasilkan efisiensi tertinggi.
- d. Estimasi emisi dan degradasi peralatan untuk mendukung kebijakan operasi ramah lingkungan.

2.2.5.3. Model Prediksi Berbasis Regresi

Pendekatan regresi berbasis data historis telah banyak digunakan dalam konteks sistem energi untuk memprediksi parameter kinerja secara kuantitatif. Luqman et al. (2023) menerapkan model regresi polinomial untuk memprediksi potensi nilai kalor refuse-derived fuel di Indonesia, yang menunjukkan bahwa hubungan nonlinier antara variabel input dan output energi dapat dimodelkan secara efektif menggunakan pendekatan regresi berbasis data. Studi tersebut menegaskan bahwa metode regresi masih relevan sebagai pendekatan prediktif, khususnya sebagai model dasar (baseline) dalam analisis kinerja energi.

Dalam konteks penelitian ini, pendekatan Machine Learning difokuskan pada pemodelan regresi untuk memprediksi nilai kontinu berupa Heat Rate PLTU berdasarkan variabel operasi aktual. Pemodelan regresi digunakan untuk mempelajari hubungan antara sekumpulan variabel input berupa parameter operasional pembangkit dan output yang ingin diestimasi (nilai Heat Rate atau efisiensi termal). Dengan memanfaatkan data historis dari sistem kontrol pembangkit, model regresi mampu mengidentifikasi pola dan keterkaitan yang kompleks termasuk hubungan non-linear yang sering kali sulit dijelaskan melalui model termodinamika konvensional (Alpaydin, 2004; Han et al., 2023).

Tabel 2-7 Perbandingan beberapa algoritma regresi yang dapat digunakan dalam pemodelan performa PLTU.

Model Regresi	Karakteristik Utama	Kelebihan	Keterbatasan
Linear Regression	Mengasumsikan hubungan linear antara variabel input dan output.	Sederhana, mudah diinterpretasi.	Tidak mampu menangkap hubungan non-linear yang kompleks.

Model Regresi	Karakteristik Utama	Kelebihan	Keterbatasan
Support Vector Regression (SVR)	Menggunakan fungsi kernel untuk menangani data non-linear.	Akurat untuk dataset kecil dan variabel sedikit.	Kurang efisien untuk dataset besar, memerlukan tuning parameter yang sensitif.
Random Forest / Gradient Boosting Regressor (GBR)	Pendekatan ensemble berbasis pohon keputusan.	Mampu menangkap hubungan non-linear dan interaksi antarvariabel.	Interpretabilitas relatif rendah, waktu komputasi lebih tinggi.
Artificial Neural Network (ANN / MLP)	Berbasis neuron dengan beberapa lapisan tersembunyi.	Dapat mempelajari pola non-linear kompleks dan dinamis.	Memerlukan jumlah data besar dan pengaturan hiperparameter yang hati-hati.

Beberapa algoritma regresi yang umum digunakan dalam analisis performa pembangkit antara lain disajikan pada Tabel 2-7 diatas. Secara umum, tujuan regresi dalam Machine Learning adalah menemukan fungsi $f(x)$ yang memetakan sekumpulan variabel input $x = (x_1, x_2, \dots, x_n)$ terhadap output kontinu y .

$$y \approx f(x), \text{ dengan } x = (x_1, x_2, \dots, x_n) \quad (2-4)$$

Fungsi ini digunakan untuk melakukan prediksi terhadap nilai baru berdasarkan data operasi aktual. Dalam konteks PLTU, variabel input dapat berupa beban generator, tekanan dan temperatur uap utama, laju aliran udara primer dan sekunder, serta aliran bahan bakar batubara. Output yang diprediksi berupa Heat Rate (kCal/kWh) atau efisiensi termal pembangkit.

Dalam penelitian ini, dua model utama digunakan karena mewakili dua paradigma berbeda dalam pembelajaran regresi, yaitu:

- a. Multi-Layer Perceptron (MLP) – model berbasis jaringan saraf tiruan dengan beberapa lapisan tersembunyi (hidden layers) yang mampu menangkap hubungan non-linear kompleks antarvariabel operasi. MLP efektif untuk mengenali pola dinamis pada sistem multi-input seperti PLTU, namun membutuhkan proses pelatihan yang intensif serta penyesuaian parameter (epoch, learning rate, dan optimizer).

- b. Gradient Boosting Regressor (GBR) – model ensemble learning berbasis pohon keputusan yang bekerja dengan membangun model secara bertahap dan memperbaiki kesalahan model sebelumnya. Metode ini unggul dalam menangani data heterogen dan menghasilkan akurasi tinggi, serta lebih stabil terhadap outlier dibanding model neural network tunggal.

Pemilihan model MLP dan GBR didasarkan pada pertimbangan keseimbangan antara akurasi, kompleksitas, dan interpretabilitas, di mana GBR unggul dalam stabilitas dan efisiensi komputasi, sedangkan MLP lebih fleksibel dalam menangkap pola non-linear dari data operasional PLTU. Kombinasi dan evaluasi keduanya diharapkan menghasilkan sistem prediksi Heat Rate yang adaptif, akurat, dan dapat diterapkan secara real-time dalam mendukung pengambilan keputusan operasional di pembangkit listrik tenaga uap modern.

2.2.6. Pra-Pemrosesan Data

Tahap pra-pemrosesan data (data preprocessing) merupakan langkah penting sebelum data digunakan dalam pelatihan model Machine Learning. Kualitas model prediksi sangat dipengaruhi oleh kualitas data yang digunakan. Data operasi PLTU umumnya diperoleh dari sistem SCADA, DCS, atau PIS, yang menyimpan ribuan titik data sensor dengan interval waktu tertentu.

Data mentah sering mengandung noise, missing value, atau anomali, sehingga perlu dilakukan pembersihan dan standarisasi agar model dapat belajar secara optimal (Han et al., 2023)

2.2.6.1. Data Cleaning dan Normalisasi

Tahap data cleaning merupakan proses penting untuk mendeteksi serta memperbaiki data yang tidak valid, inkonsisten, atau tidak relevan agar kualitas analisis data meningkat. Pada data operasi pembangkit listrik, langkah ini harus memastikan keakuratan dan reliabilitas sensor dan histori operasional sebelum dianalisis secara lebih lanjut (Elhafiz & Ahmed, 2019). Dalam konteks data operasi PLTU, beberapa masalah umum yang ditemukan meliputi:

- a. Data hilang (missing values) akibat gangguan komunikasi sensor atau kegagalan sistem akuisisi data.

- b. Noise dan spike berupa lonjakan nilai ekstrem yang tidak mencerminkan kondisi sebenarnya.
- c. Data duplikat atau tidak sinkron antar-sumber data (misalnya antara SCADA dan histori manual).

Teknik data cleaning dan preprocessing yang sering digunakan:

- a. Imputasi nilai hilang dengan cara mean, median, regresi, atau algoritma Multiple Imputation (MI) dan MICE (Multiple Imputation by Chained Equations), serta pendekatan deep learning.
- b. Penghapusan data ekstrem/outlier berdasarkan ambang batas realistis dan validasi domain.
- c. Filtering sinyal untuk mengurangi noise pada data sensor, misalnya dengan moving average atau algoritma robust seperti RANSAC.

Setelah pembersihan, langkah berikutnya adalah normalisasi atau standardisasi agar seluruh variabel berada dalam skala yang sebanding. Hal ini penting karena perbedaan skala antarparameter (misalnya tekanan dalam MPa dan temperatur dalam °C) dapat menyebabkan bias dalam proses pelatihan model.

Dua metode umum digunakan:

- a. Min-Max Normalization

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (2-5)$$

Mengubah nilai variabel ke dalam rentang [0, 1].

- b. Z-Score Standardization

$$X' = \frac{X - \mu}{\sigma} \quad (2-6)$$

Mengubah nilai berdasarkan distribusi normal dengan rata-rata nol dan deviasi standar satu.

Model seperti Neural Network (MLP) sensitif terhadap skala input, sehingga normalisasi Min–Max sering lebih sesuai, sedangkan algoritma berbasis tree ensemble seperti Gradient Boosting Regressor tidak selalu memerlukan normalisasi, tetapi tetap diuntungkan dari konsistensi data yang bersih (de Amorim et al., 2022).

2.2.6.2. Pembagian Data Latih dan Uji

Agar model machine learning dapat dilatih secara optimal dan dievaluasi dengan akurat, dataset umumnya dibagi menjadi dua bagian utama yaitu data pelatihan (training set) dan data pengujian (testing set). Data pelatihan berfungsi untuk mengembangkan model melalui pembelajaran pola dan hubungan antar variabel dari data yang tersedia, sedangkan data pengujian digunakan untuk menilai kemampuan model dalam melakukan generalisasi terhadap data baru yang belum pernah dikenali sebelumnya. Pemisahan ini merupakan tahap penting untuk menghindari bias dan memastikan penilaian yang objektif pada kinerja model terhadap data yang tidak terobservasi (Sivakumar et al., 2024).

Rasio pembagian dataset yang lazim adalah antara 70:30 hingga 80:20, di mana sebagian besar data digunakan untuk pelatihan model dan sisanya untuk pengujian. Strategi ini sering diterapkan untuk memperoleh generalisasi yang baik sekaligus mencegah terjadinya overfitting, yaitu ketika model terlalu menyesuaikan data pelatihan sehingga kurang mampu mengolah data baru secara akurat (Joseph, 2022).

Untuk data operasi PLTU yang bersifat time series, pembagian data dilakukan secara berurutan berdasarkan waktu (chronological split) agar mencerminkan kondisi aktual operasi. Ini bertujuan agar model yang dibangun dapat dievaluasi menggunakan data periode terbaru untuk menggambarkan aplikasi aktual di lapangan. Pembagian semacam ini merepresentasikan situasi nyata di mana model harus melakukan prediksi berdasarkan informasi sensor terkini, serta memudahkan penilaian generalisasi dan akurasi model dalam skenario operasional yang sebenarnya (Wakjira et al., 2022).

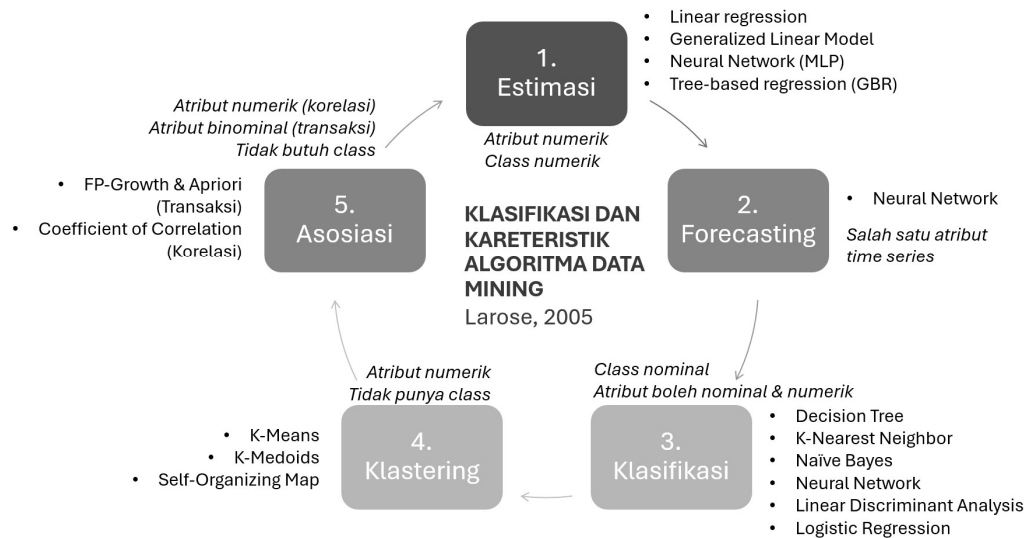
Selain membagi data secara sederhana, banyak penelitian di industri energi juga menggunakan validasi silang seperti k-fold cross-validation. Metode ini membantu memastikan model tidak hanya cocok pada satu bagian data, tetapi tetap akurat di berbagai kondisi operasional pembangkit, termasuk variasi beban dan musim (Boshoma et al., 2025).

2.2.7. Algoritma Model Machine Learning

Tahapan pemodelan merupakan inti dari proses analisis Machine Learning, di mana algoritma dilatih untuk mengenali pola hubungan antara variabel input (parameter operasional PLTU) dan output (nilai Heat Rate).

Pemilihan algoritma dalam penelitian ini didasarkan pada peran analitik data mining serta karakteristik tipe data dan variabel target yang dianalisis. Mengacu pada

klasifikasi peran utama data mining (Larose, 2005) seperti pada Gambar 2-4, permasalahan dalam penelitian ini termasuk ke dalam kategori estimasi dan forecasting, karena variabel target yang diprediksi, yaitu Net Plant Heat Rate (NPHR), bersifat numerik kontinu dan dianalisis berdasarkan data historis operasi PLTU. Oleh karena itu, pendekatan yang digunakan harus berupa supervised learning untuk tugas estimasi.



Gambar 2-4 Karakteristik dan klasifikasi algoritma berdasarkan peran utama data mining diadaptasi dari Larose (2005).

Data operasi PLTU yang digunakan dalam penelitian ini terdiri dari atribut numerik multivariat dengan hubungan yang bersifat non-linear dan dipengaruhi oleh berbagai kondisi operasi, seperti beban unit, parameter uap, serta karakteristik pembakaran. Kondisi tersebut menuntut penggunaan algoritma yang mampu menangkap pola non-linear kompleks sekaligus memiliki kemampuan generalisasi yang baik. Berdasarkan pertimbangan tersebut, penelitian ini memilih Multi-Layer Perceptron (MLP) dan Gradient Boosting Regressor (GBR) sebagai representasi dua paradigma pembelajaran non-linear yang berbeda, namun berada dalam peran analitik yang sama, yaitu estimasi dan forecasting nilai numerik.

MLP merupakan jaringan saraf tiruan berlapis yang mampu memetakan hubungan non-linear kompleks antara input dan output, sedangkan GBR merupakan metode ansambel berbasis pohon keputusan yang membangun model prediktif secara bertahap melalui penggabungan sejumlah pohon keputusan lemah. Pendekatan komparatif ini tidak bertujuan untuk menentukan algoritma yang unggul secara universal, melainkan untuk

mengevaluasi perbedaan karakteristik kinerja dan kemampuan generalisasi kedua model ketika diterapkan pada data operasi aktual PLTU.

Kinerja kedua model dievaluasi menggunakan tiga metrik utama, yaitu Root Mean Squared Error (RMSE) untuk mengukur rata-rata kesalahan kuadrat akar, Mean Absolute Error (MAE) untuk mengukur rata-rata selisih absolut, dan Koefisien Determinasi (R^2) sebagai ukuran seberapa baik variabilitas output dijelaskan oleh model

2.2.7.1. Multi-Layer Perceptron (MLP)

Multi-Layer Perceptron (MLP), yang juga dikenal sebagai Feed Forward Neural Networks (FFNN), adalah jenis arsitektur jaringan saraf tiruan (Artificial Neural Network, ANN) yang terdiri dari beberapa lapisan. Struktur MLP mencakup satu lapisan input, satu atau lebih lapisan tersembunyi (hidden layers), dan satu lapisan output. Setiap lapisan terdiri dari sejumlah neuron yang menerima input dari neuron pada lapisan sebelumnya dan mengirimkan output ke neuron pada lapisan berikutnya. Dalam MLP, tidak terdapat koneksi antar neuron dalam satu lapisan yang sama.

$$y = f(Wx + b) \quad (2-7)$$

Dengan:

- a. W = *matrix bobot*,
- b. x = *vektor input*,
- c. b = *bias*,
- d. $f(.)$ = fungsi aktivasi non-linear, misal ReLU (Rectified Linear Unit), tanh, atau sigmoid.

Proses pembelajaran dilakukan menggunakan algoritma backpropagation, yang menyesuaikan bobot agar kesalahan antara nilai aktual (y_i) dan hasil prediksi ($\hat{y}_i$) dapat diminimalkan.

Hyperparameter atau parameter utama yang memengaruhi kinerja MLP antara lain:

- a. Jumlah hidden layers dan neuron per layer,
- b. Learning rate dan jumlah epoch,
- c. jenis optimizer (misalnya Adam, SGD).

Kelebihan MLP terletak pada kemampuannya menangkap hubungan non-linear kompleks antarvariabel operasi PLTU. Namun, model ini membutuhkan data dalam jumlah besar dan sangat sensitif terhadap skala input, sehingga normalisasi (Min–Max) diperlukan pada tahap pra-pemrosesan data (Han et al., 2023)

2.2.7.2. Gradient Boosting Regressor (GBR)

Gradient Boosting Regressor (GBR) adalah metode ensemble learning berbasis pohon keputusan (decision tree) yang bekerja dengan prinsip boosting. Model ini membangun sejumlah pohon secara bertahap, di mana setiap pohon baru berusaha memperbaiki kesalahan (residual error) dari pohon sebelumnya. Dengan pendekatan ini, GBR mampu menghasilkan model yang lebih akurat dan tahan terhadap overfitting.

Secara umum, model GBR dapat ditulis sebagai:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \quad (2-8)$$

Dengan:

- a. $F_m(x)$ = model pada iterasi ke- m ,
- b. $F_{m-1}(x)$ = model sebelumnya.
- c. $h_m(x)$ = pohon keputusan baru yang ditambahkan,
- d. γ_m = koefisien pembobot hasil *gradient descent*.

Model akhir $F_m(x)$ diperoleh setelah m iterasi pembelajaran.

Pendekatan ini efektif karena meminimalkan fungsi kerugian secara bertahap, mengikuti arah gradien dari kesalahan model sebelumnya.

Beberapa hyperparameter utama dalam GBR meliputi:

- a. learning rate (η) – mengontrol kontribusi tiap pohon baru,
- b. n_estimators – jumlah total pohon yang dibangun,
- c. max_depth – kedalaman maksimum setiap pohon,
- d. subsample ratio – proporsi data yang digunakan pada setiap iterasi.

Kelebihan utama GBR adalah stabilitas tinggi dan ketahanan terhadap overfitting, bahkan ketika data mengandung noise. Selain itu, model ini memiliki kemampuan interpretasi yang baik melalui analisis feature importance (Ashraf et al., 2025), yang membantu mengidentifikasi parameter operasi dominan terhadap Heat Rate.

2.2.8. Pengukuran dan Evaluasi Model

Pengukuran (*measurement*) merupakan proses pemberian nilai numerik terhadap suatu karakteristik berdasarkan aturan yang sistematis dan terdefinisi. Dalam penelitian kuantitatif, pengukuran digunakan untuk merepresentasikan tingkat kesesuaian antara nilai observasi dan hasil estimasi model (Montgomery & Runger, 2005).

Dalam pemodelan regresi, evaluasi kinerja model didasarkan pada konsep kesalahan (*error*), yaitu selisih antara nilai aktual dan nilai prediksi. Besarnya error menunjukkan tingkat penyimpangan model terhadap kondisi nyata. Selain itu, konsep deviasi dan variansi digunakan untuk menilai sejauh mana model mampu menjelaskan penyebaran data terhadap rata-rata (Kutner et al., 2005).

Pengukuran berbasis error menjadi dasar dalam menentukan kualitas model prediktif. Oleh karena itu, dalam analisis regresi numerik, digunakan ukuran kuantitatif untuk menilai akurasi dan konsistensi model secara objektif. Pendekatan ini memungkinkan peneliti mengevaluasi apakah model mampu merepresentasikan pola data secara memadai.

Dengan demikian, teori pengukuran dalam penelitian ini menekankan bahwa evaluasi model prediktif harus didasarkan pada ukuran kuantitatif yang mencerminkan tingkat kesalahan dan kemampuan model dalam menjelaskan variasi data.

2.2.8.1. Evaluasi Kinerja Model

Untuk mengukur akurasi dan reliabilitas model prediksi Heat Rate, digunakan tiga metrik evaluasi utama, yaitu Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan Koefisien Determinasi (R^2).

Ketiga metrik ini umum digunakan dalam analisis regresi dan mampu memberikan gambaran menyeluruh tentang kualitas model.

- a. Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2-9)$$

Mengukur rata-rata akar kuadrat selisih antara nilai aktual (y_i) dan nilai prediksi ($\hat{y}_i$). Semakin kecil nilai RMSE, semakin baik akurasi model.

- b. Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2-10)$$

Mengukur rata-rata kesalahan absolut tanpa memperhatikan arah deviasi. MAE lebih mudah diinterpretasi dan tidak terlalu sensitif terhadap nilai ekstrem dibanding RMSE.

- c. Koefisien Determinasi (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2-11)$$

Menunjukkan proporsi variansi data yang dapat dijelaskan oleh model. Nilai R^2 mendekati 1 menandakan bahwa model memiliki kemampuan prediksi yang baik.

Target performa model prediksi Heat Rate pada penelitian serupa mencapai nilai $R^2 \geq 0.85$ dengan dengan nilai RMSE dan MAE yang rendah (Arferiandi et al., 2021a; Ding et al., 2024).

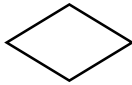
2.2.9. Representasi Alur Sistem (Flowchart)

Flowchart menggunakan berbagai simbol standar untuk merepresentasikan jenis proses yang berbeda dalam suatu sistem atau algoritma. Simbol-simbol tersebut telah distandarisasi dalam praktik rekayasa sistem dan pemrograman sehingga dapat dipahami secara universal. Setiap simbol memiliki makna spesifik dalam menggambarkan aktivitas tertentu, mulai dari awal proses, pengolahan data, pengambilan keputusan, hingga akhir proses (Kendall & Kendall, 2010; Pressman & Maxim, 2020)

Beberapa simbol yang umum digunakan dalam flowchart disajikan pada Tabel 2.8 berikut.

Tabel 2-8 Simbol-simbol flowchart

No	Nama Simbol	Bentuk	Fungsi	Keterangan Penggunaan
1	Terminator (Oval)		Menunjukkan awal dan akhir proses	Digunakan untuk menandai titik mulai dan selesai dalam suatu alur sistem

2	Process (Persegi Panjang)		Merepresentasikan proses atau langkah komputasi	Digunakan untuk menunjukkan aktivitas pengolahan data atau instruksi
3	Decision (Belah Ketupat)		Menunjukkan pengambilan keputusan	Memiliki dua atau lebih cabang keluaran berdasarkan kondisi tertentu
4	Input/Output (Jajar Genjang)		Menunjukkan proses input atau output data	Digunakan untuk menerima atau menampilkan informasi
5	Flow Line (Garis Panah)		Menunjukkan arah aliran proses	Menghubungkan simbol satu dengan lainnya untuk menunjukkan urutan langkah

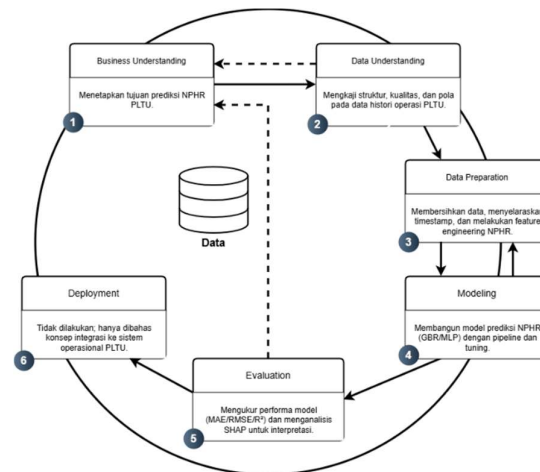
Penggunaan simbol-simbol tersebut memungkinkan alur logika suatu sistem atau algoritma dapat divisualisasikan secara jelas, terstruktur, dan mudah dipahami. Dalam penelitian ini, flowchart digunakan untuk menggambarkan alur proses penelitian mulai dari tahap pengumpulan data, pra-pemrosesan, pemodelan menggunakan algoritma Multi-Layer Perceptron dan Gradient Boosting Regressor, hingga evaluasi kinerja model. Dengan demikian, setiap tahapan metodologis dapat dipetakan secara sistematis dan transparan.

2.2.10. Metodologi Data Mining: CRISP-DM

Cross Industry Standard Process for Data Mining (CRISP-DM) merupakan salah satu kerangka kerja yang paling banyak digunakan dalam pengembangan proyek data mining dan machine learning. CRISP-DM diakui sebagai model proses yang sistematis dan terstruktur dalam berbagai penelitian data science modern (Mariscal et al., 2010).

Metodologi ini bersifat iteratif dan memungkinkan adanya umpan balik antar tahapan sehingga proses pengembangan model dapat disesuaikan dengan dinamika data dan tujuan analisis (Saltz & Shamshurin, 2016).

CRISP-DM dikembangkan untuk menjembatani kebutuhan analisis data dengan tujuan operasional atau bisnis yang ingin dicapai. Kerangka ini tidak hanya berfokus pada aspek teknis pemodelan, tetapi juga menekankan pentingnya pemahaman konteks permasalahan dan kualitas data sebelum model dibangun. Dengan demikian, CRISP-DM menyediakan pendekatan yang menyeluruh mulai dari identifikasi masalah hingga evaluasi hasil analisis.



Gambar 2-5 Diagram CRISP-DM dan keterkaitan antar tahapannya (diadaptasi dari Fitria (2023))

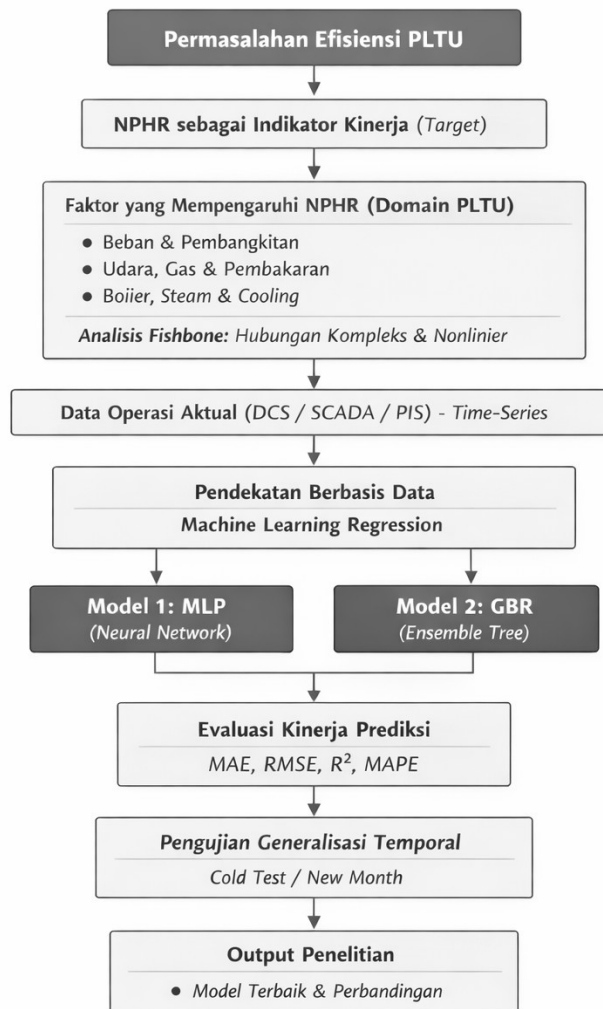
CRISP-DM seperti pada Gambar 2-4 terdiri atas enam tahapan utama, yaitu:

- a. Business Understanding, pada tahap ini dilakukan penyelarasan antara kebutuhan analitis dan konteks operasional sehingga arah pengembangan model menjadi jelas dan terukur.
- b. Data Understanding, tahap ini bertujuan untuk mengidentifikasi pola, anomali, serta potensi keterbatasan data sebelum dilakukan pemrosesan lebih lanjut.
- c. Data Preparation, yaitu tahapan pembersihan data (data cleaning), transformasi variabel, integrasi dataset, serta seleksi fitur yang relevan.
- d. Modeling, yaitu pembangunan dan pelatihan model menggunakan algoritma yang dipilih.
- e. Evaluation, yaitu pengukuran kinerja model berdasarkan metrik yang telah ditentukan serta penilaian kesesuaian hasil model terhadap tujuan awal proyek. Tahap ini memastikan bahwa model tidak hanya akurat secara numerik, tetapi juga relevan secara konseptual.
- f. Deployment, yaitu implementasi model atau perumusan integrasi hasil analisis ke dalam sistem operasional. Pada konteks penelitian akademik, tahap ini dapat berupa rekomendasi atau rancangan pemanfaatan model dalam sistem nyata.

Pendekatan ini banyak digunakan karena memberikan alur kerja yang terstruktur, namun tetap fleksibel dalam menangani permasalahan berbasis data.

2.3. Kerangka Pemikiran

Dengan mempertimbangkan hasil kajian terhadap penelitian-penelitian terdahulu, diperoleh pemahaman mengenai konsep-konsep kunci, pendekatan metodologis, serta temuan empiris yang relevan dengan permasalahan prediksi Net Plant Heat Rate (NPHR) pada PLTU. Sintesis dari landasan teori dan kesenjangan penelitian yang telah diidentifikasi tersebut menjadi dasar dalam penyusunan kerangka pemikiran penelitian ini. Kerangka pemikiran ini menggambarkan hubungan logis antara permasalahan, variabel penelitian, pendekatan pemodelan, serta metode evaluasi yang digunakan, sebagaimana ditunjukkan pada Gambar 2-6.



Gambar 2-6 Kerangka Pemikiran Penelitian