

PROPOSAL TUGAS AKHIR

**PERAMALAN BEBAN LISTRIK NEGARA MALAYSIA
MENGUNAKAN XGBOOST RECURSIVE DENGAN
SHAPLEY VALUE**



Disusun Oleh:

EVAN HANIF WIDIATAMA

M0722037

**Diajukan untuk menyusun skripsi
PROGRAM STUDI STATISTIKA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN
ALAM
UNIVERSITAS SEBELAS MARET
SURAKARTA
2025**

BAB I PENDAHULUAN

1.1 LATAR BELAKANG

Peramalan beban listrik merupakan komponen fundamental dalam operasional sistem tenaga listrik modern yang memainkan peran kritis dalam memastikan keandalan dan efisiensi biaya operasi penyediaan Listrik (Hopf et al., 2023). Karakteristik unik energi listrik yang tidak dapat disimpan dalam jangka waktu yang lama mengharuskan operator sistem untuk secara kontinyu menyeimbangkan antara produksi dan konsumsi energi secara *real-time* (Su & Gamal, 2011). Ketidakseimbangan antara pasokan dan permintaan energi listrik dapat menimbulkan konsekuensi ekonomi dan operasional yang signifikan bagi semua *stakeholder* dalam rantai nilai industri ketenagalistrikan.

Akurasi peramalan beban listrik memiliki implikasi finansial yang substansial terhadap perusahaan pembangkit listrik dan operator sistem. Overestimasi beban listrik jangka panjang mengakibatkan investasi yang terbuang sia-sia dalam pembangunan fasilitas pembangkit berlebih, sementara underestimasi beban masa depan berpotensi menghasilkan pembangkitan yang tidak mencukupi dan permintaan yang tidak terpenuhi (Khuntia et al., 2017). Ketidakakuratan prakiraan harga listrik dapat menyebabkan kerugian *profit* yang signifikan bagi pemasok energi akibat kesalahan dalam penjadwalan skema pembangkitan Listrik (Ugurlu et al., 2018). Dari perspektif lingkungan, peningkatan akurasi prediksi permintaan listrik memberikan manfaat ekologis melalui minimalisasi pemborosan energi yang dihasilkan dari pembangkitan listrik berlebihan (Wu et al., 2023).

Kompleksitas peramalan beban listrik semakin meningkat seiring dengan transisi global menuju sumber energi terbarukan yang terdistribusi dan elektrifikasi berbagai sektor ekonomi (Hopf et al., 2023). Variabilitas tinggi dari sumber energi terbarukan menghadirkan tantangan signifikan terhadap operasi jaringan listrik, di mana generator konvensional dapat digunakan untuk memitigasi variabilitas ini namun memiliki biaya operasional tinggi dan menghasilkan emisi karbon (Su & Gamal, 2011). Aktivitas ekonomi hampir secara universal memerlukan listrik

sebagai input yang sulit disubstitusi, khususnya dalam jangka pendek, yang menjadikan konsumsi listrik sebagai indikator *real-time* yang akurat untuk aktivitas ekonomi (Fezzi & Fanghella, 2021).

Perkembangan metodologi peramalan beban listrik telah mengalami evolusi signifikan dengan munculnya teknik *machine learning* dan *deep learning* yang mampu menangkap pola kompleks dan *nonlinear* dalam data konsumsi listrik. Namun, tantangan utama dalam peramalan beban listrik mencakup *multicollinearity* dan non-stasioneritas pada data deret waktu, yang memerlukan pendekatan metodologis yang *sophisticated* untuk menghasilkan prediksi yang akurat dan *robust*. Meta-analisis terhadap 421 model peramalan dari 59 studi menunjukkan bahwa faktor-faktor seperti level grid, horizon peramalan, dan metodologi yang digunakan secara signifikan mempengaruhi akurasi peramalan beban listrik jangka pendek (Hopf et al., 2023).

Sistem penyimpanan energi menyediakan alternatif yang lebih ramah lingkungan untuk memitigasi variabilitas energi terbarukan, namun memiliki biaya *deployment* yang tinggi dalam skala besar (Su & Gamal, 2011). Kapasitas *value* dari sistem penyimpanan energi bergantung pada strategi kontrol yang diterapkan, di mana berbagai strategi kontrol dapat dimanfaatkan untuk mencapai tujuan yang beragam dalam sistem penyimpanan energi (Shi & Luo, 2017). Kondisi ini menggarisbawahi pentingnya pengembangan model peramalan yang akurat untuk mengoptimalkan utilisasi infrastruktur energi yang ada tanpa mengandalkan investasi masif dalam teknologi penyimpanan.

Dalam konteks ini, pengembangan model XGBoost rekursif dengan integrasi *Shapley Additive Explanations* (SHAP) menawarkan pendekatan inovatif untuk mengatasi kompleksitas peramalan beban listrik. Pendekatan ini tidak hanya bertujuan untuk meningkatkan akurasi prediksi melalui optimasi hiperparameter berbasis Bayesian, tetapi juga memberikan interpretabilitas model melalui analisis kontribusi variabel prediktor dalam kerangka teori *game-theoretic*, yang sangat penting untuk pengambilan keputusan strategis dalam manajemen sistem tenaga listrik.

1.2 RUMUSAN MASALAH

Berdasarkan latar belakang yang telah diuraikan, dapat diidentifikasi beberapa permasalahan penelitian sebagai berikut:

1. Bagaimana mengembangkan model XGBoost rekursif yang mampu mengintegrasikan variabel temporal multi-skala, variabel kalender struktural, dan variabel eksogen lingkungan untuk mengatasi kompleksitas dinamika permintaan beban listrik pada data deret waktu?
2. Bagaimana mengoptimalkan kinerja sistem peramalan beban listrik jangka menengah melalui implementasi *Bayesian Optimization* dan validasi *sliding window cross-validation* untuk meningkatkan akurasi prediksi dan kemampuan generalisasi model?
3. Bagaimana mengidentifikasi dan mengkuantifikasi kontribusi faktor-faktor penentu utama dalam permintaan beban listrik melalui pendekatan *Shapley Additive Explanations* (SHAP) untuk memberikan interpretabilitas model dan pemahaman mendalam terhadap kompleksitas interaksi *nonlinear* antar variabel prediktor?

1.3 TUJUAN PENELITIAN

Berdasarkan rumusan masalah yang telah diformulasikan, penelitian ini bertujuan untuk:

1. Mengembangkan model XGBoost rekursif yang mengintegrasikan variabel temporal multi-skala (*lag* 1-24 jam, *moving average* 3 hari, 1 minggu, 1 bulan, dan 3 bulan), variabel kalender struktural (pola harian, jam operasional, variasi musiman bulanan, dan indikator hari libur nasional), serta variabel eksogen suhu lingkungan untuk memodelkan dinamika permintaan beban listrik dengan efektif.
2. Membangun sistem peramalan beban listrik jangka menengah dengan horizon prediksi 168 jam melalui implementasi optimasi hiperparameter berbasis *Bayesian Optimization* yang dikombinasikan

dengan teknik validasi *sliding window cross-validation* untuk meningkatkan akurasi prediksi dan kemampuan generalisasi model secara signifikan.

3. Menganalisis dan mengkuantifikasi faktor-faktor penentu utama dalam permintaan beban listrik melalui implementasi pendekatan *Shapley Additive Explanations* (SHAP) untuk mengukur kontribusi relatif setiap variabel prediktor dan mengungkap kompleksitas interaksi *nonlinear* antar variabel dalam kerangka teoritis *game-theoretic*, sehingga memberikan interpretabilitas model yang komprehensif untuk mendukung pengambilan keputusan strategis dalam manajemen sistem tenaga listrik.

1.4 MANFAAT PENELITIAN

1.4.1 Manfaat Teoritis

Penelitian ini memberikan kontribusi teoretis yang signifikan dalam pengembangan metodologi peramalan beban listrik melalui beberapa aspek berikut:

1. **Pengembangan Framework Metodologis:** Penelitian ini menghasilkan kerangka kerja inovatif yang mengintegrasikan algoritma XGBoost rekursif dengan teknik interpretabilitas SHAP untuk peramalan deret waktu, yang dapat menjadi referensi metodologis bagi pengembangan model prediktif dalam domain energi dan aplikasi deret waktu lainnya.
2. **Kontribusi pada Teori Machine Learning:** Implementasi *Bayesian Optimization* untuk optimasi hiperparameter dalam konteks peramalan beban listrik memberikan wawasan baru tentang efektivitas teknik optimasi *advanced* dalam meningkatkan performa model *ensemble learning* pada data dengan karakteristik *temporal* yang kompleks.
3. **Advancement dalam Interpretabilitas Model:** Penerapan *Shapley Additive Explanations* dalam konteks peramalan energi memperkaya pemahaman teoretis tentang *explainable AI* dalam aplikasi *critical*

infrastructure, khususnya dalam mengungkap interaksi nonlinear antar variabel temporal dan eksogen.

1.4.2 Manfaat Praktis

1.4.2.1 Bagi Industri Ketenagalistrikan

1. **Optimasi Operasional:** Model yang dikembangkan memungkinkan operator sistem tenaga listrik untuk melakukan perencanaan operasional yang lebih akurat, mengurangi biaya operasional melalui optimasi *dispatching* pembangkit listrik, dan meminimalkan risiko *blackout* atau *brownout*.
2. **Manajemen Risiko Finansial:** Akurasi peramalan yang ditingkatkan membantu perusahaan pembangkit listrik dalam mengurangi *exposure* terhadap volatilitas harga energi dan mengoptimalkan strategi *hedging* dalam pasar energi.
3. **Perencanaan Investasi Infrastruktur:** Hasil peramalan jangka menengah yang akurat mendukung pengambilan keputusan investasi yang lebih *informed* dalam pengembangan kapasitas pembangkit dan infrastruktur transmisi.

1.4.2.2 Bagi Pemerintah dan Regulator

1. **Kebijakan Energi:** Insight dari analisis SHAP memberikan *evidence-based foundation* untuk formulasi kebijakan energi nasional dan strategi ketahanan energi jangka panjang.
2. **Regulasi Tarif:** Pemahaman mendalam tentang faktor-faktor penentu konsumsi listrik mendukung penetapan struktur tarif yang lebih reflektif terhadap pola konsumsi aktual dan mendorong efisiensi energi.
3. **Perencanaan Pembangunan:** Hasil penelitian dapat diintegrasikan dalam *master plan* pembangunan sektor energi untuk mendukung target elektrifikasi dan transisi energi berkelanjutan.

1.4.2.3 Bagi Masyarakat dan Konsumen

1. **Stabilitas Pasokan Listrik:** Peningkatan akurasi peramalan berkontribusi pada stabilitas pasokan listrik yang lebih baik, mengurangi frekuensi pemadaman bergilir dan gangguan layanan listrik.
2. **Efisiensi Biaya:** Optimasi operasional sistem ketenagalistrikan berpotensi menurunkan biaya pokok penyediaan tenaga listrik yang pada akhirnya dapat direfleksikan dalam tarif listrik yang lebih kompetitif.
3. **Kualitas Layanan:** Prediksi yang lebih akurat memungkinkan penyediaan layanan listrik yang lebih *reliable* dan *consistent*, mendukung aktivitas ekonomi dan *social welfare* masyarakat.

1.4.2.4 Bagi Pengembangan Ilmu Pengetahuan

1. **Referensi Metodologis:** Penelitian ini menyediakan *baseline methodology* yang dapat diadaptasi dan dikembangkan lebih lanjut untuk aplikasi peramalan pada sektor infrastruktur kritikal lainnya seperti air, gas, dan telekomunikasi.
2. **Dataset dan Benchmark:** Hasil evaluasi kinerja model dapat menjadi *benchmark* untuk pengembangan algoritma prediksi beban listrik yang lebih advanced di masa depan.
3. **Cross-disciplinary Application:** *Framework* yang dikembangkan dapat diadaptasi untuk aplikasi *forecasting* dalam domain lain yang memiliki karakteristik data temporal yang serupa, seperti *financial markets*, *supply chain management*, dan *urban planning*.

BAB II LANDASAN TEORI

2.1 TINJAUAN PUSTAKA

2.1.1 Time Series Forecasting

Time series forecasting merupakan salah satu bidang fundamental dalam analisis data yang memiliki aplikasi luas dalam berbagai domain praktis (Adhikari & Agrawal, 2013). Time series didefinisikan sebagai serangkaian observasi yang dikumpulkan secara berurutan dalam interval waktu tertentu, di mana setiap observasi mempertimbangkan komponen waktu sebagai elemen fundamental dalam struktur dan analisisnya (Garg et al., 2022).

Secara matematis, time series dapat direpresentasikan sebagai:

$$Y_t = f(t) + \epsilon_t$$

dengan Y_t adalah nilai observasi pada waktu t , $f(t)$ adalah fungsi deterministik dari waktu, dan ϵ_t adalah komponen *error* atau *noise*.

2.1.2 Extreme Gradient Boosting (XGBoost)

XGBoost (*Extreme Gradient Boosting*) merupakan implementasi *optimized* dari algoritma *gradient boosting* yang dirancang untuk meningkatkan performa dan efisiensi komputasi. Algoritma ini menggunakan pendekatan *ensemble learning* yang menggabungkan *multiple weak learners* untuk membentuk *strong learner* (Pavlyshenko, 2017).

Objective function XGBoost didefinisikan sebagai:

$$\mathcal{L}(\Phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k)$$

dengan l adalah *loss function* yang dapat didiferensiasi, $\Omega(f_k)$ adalah *regularization term* untuk tree k , dan Φ merepresentasikan parameter model.

Regularization term dinyatakan sebagai:

Commented [e1]: Di mana diubah menjadi dengan

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

dengan T adalah jumlah *leaves*, w_j adalah *weight* pada *leaf* j , γ dan λ adalah *hyperparameter regularization*.

Untuk iterasi ke- t , prediksi model adalah:

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

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

dengan f_t adalah *tree* yang ditambahkan pada iterasi t .

Commented [e2]: Hat di ynya saja, indexnya jangan

2.1.3 Recursive Multi-step Forecasting

Recursive forecasting merupakan strategi prediksi *multi-step* yang menggunakan prediksi sebelumnya sebagai input untuk prediksi berikutnya (Jing et al., 2022). Metode ini sangat efektif untuk *forecasting horizon* jangka menengah hingga panjang.

Untuk *horizon forecasting* h , *recursive strategy* dapat diformulasikan sebagai:

$$\hat{y}_{(t+1)} = f(y_t, y_{t-1}, \dots, y_{t-p+1}, MA_3(t), \dots, h(t), \dots, x_{t+1})$$

$$\hat{y}_{(t+2)} = f(\hat{y}_{(t+1)}, y_t, \dots, y_{t-p+2}, x_{t+2})$$

⋮

$$\hat{y}_{(t+h)} = f(\hat{y}_{(t+h-1)}, \hat{y}_{(t+h-2)}, \dots, \hat{y}_{(t+h-p+1)}, x_{t+h})$$

dengan f adalah fungsi prediksi, p adalah *lag order*, dan x_t adalah variabel eksogen.

2.1.4 Variabel Input

Commented [e3]: Diganti menjadi variabel input, gunakan istilah yang umum

2.1.4.1 Variabel Lag

Lag features merupakan nilai historis dari variabel target yang digunakan sebagai prediktor. Untuk *electrical load forecasting*, *lag features* dapat diformulasikan sebagai:

$$Y_{lag,i} = y_{t-i} \quad \text{untuk } i = 1, 2, \dots, L$$

dengan L adalah maksimum *lag* yang dipertimbangkan.

2.1.4.2 Variabel Moving Average

Variabel *Moving average* menangkap trend jangka pendek hingga menengah dalam data:

$$MA_k(t) = \frac{1}{k} \sum_{i=0}^{k-1} y_{t-i}$$

dengan k adalah *window size*.

2.1.4.3 Variabel Kalender

Variabel kalender menangkap pola temporal struktural:

- **Hour of day:** $h(t) \in \{0, 1, \dots, 23\}$
- **Day of week:** $d(t) \in \{0, 1, \dots, 6\}$
- **Month of year:** $m(t) \in \{1, 2, \dots, 12\}$
- **Holiday indicator:** $H(t) \in \{0, 1\}$

2.1.4.4 Variabel Encoding Siklis

Untuk menangkap sifat siklis dari variabel kalender:

$$\text{hour_sin}(t) = \sin\left(\frac{2\pi \times h_t}{23}\right)$$

$$\text{hour_cos}(t) = \cos\left(\frac{2\pi \times h_t}{23}\right)$$

2.1.4.5 Variabel Eksogen Suhu

2.1.5 Optimasi Bayesian Hiperparameter Tuning

Bayesian Optimization merupakan metode optimasi global yang efisien untuk *hyperparameter tuning*, terutama ketika evaluasi *objective function* membutuhkan komputasi yang mahal (Arslan et al., 2023).

$$f(x) \sim \mathcal{GP}(\mu(x), k(x, x'))$$

dengan $\mu(x)$ adalah *mean function* dan $k(x, x')$ adalah *covariance function*.

Acquisition function yang umum digunakan adalah *Expected Improvement (EI)*:

$$EI(x) = E[\max(f(x) - f(x^+), 0)]$$

dengan $f(x^+)$ adalah nilai terbaik yang telah ditemukan.

Untuk *Gaussian Process*, EI dapat dihitung secara analitis:

$$EI(x) = (\mu(x) - f(x^+))\Phi(Z) + \sigma(x)\phi(Z)$$

dengan $Z = \frac{\mu(x) - f(x^+)}{\sigma(x)}$, Φ adalah CDF normal standar, dan ϕ adalah PDF normal standar.

2.1.6 Time Series Cross-Validation

2.1.6.1 Expanding Window Cross-Validation

Expanding window menggunakan semua data historis yang tersedia untuk *training*:

$$\text{Train}_i = \{(x_1, y_1), (x_2, y_2), \dots, (x_{n_i}, y_{n_i})\}$$

$$\text{Test}_i = \{(x_{n_i+1}, y_{n_i+1}), \dots, (x_{n_i+h}, y_{n_i+h})\}$$

dengan n_i meningkat setiap *fold*.

2.1.6.2 Sliding Window Cross-Validation

Sliding window menggunakan *window size* tetap untuk *training*:

$$\text{Train}_i = \{(x_i, y_i), (x_{i+1}, y_{i+1}), \dots, (x_{i+w-1}, y_{i+w-1})\}$$

$$\text{Test}_i = \{(x_{i+w}, y_{i+w}), \dots, (x_{i+w+h-1}, y_{i+w+h-1})\}$$

dengan w adalah *window size* yang konstan.

2.1.7 Shapley Additive Explanations (SHAP)

SHAP values memberikan *unified framework* untuk *model interpretability* berdasarkan game theory. Untuk feature i , *SHAP value* didefinisikan sebagai:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|! (|F| - |S| - 1)!}{|F|!} [f(S \cup \{i\}) - f(S)]$$

dengan F adalah *set* semua *features*, S adalah *subset features*, dan $f(S)$ adalah *expected prediction* ketika hanya menggunakan *features* dalam S .

SHAP values memenuhi tiga axiom penting:

1. **Efficiency:** $\sum_{i=1}^n \phi_i = f(x) - f(\emptyset)$
2. **Symmetry:** Jika $f(S \cup \{i\}) = f(S \cup \{j\})$ untuk semua S , maka $\phi_i = \phi_j$
3. **Dummy:** Jika $f(S \cup \{i\}) = f(S)$ untuk semua S , maka $\phi_i = 0$

Untuk *tree-based models* seperti XGBoost, *SHAP values* dapat dihitung secara efisien menggunakan *TreeSHAP algorithm* yang memiliki kompleksitas *polynomial time*.

2.1.8 Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error (MAPE) merupakan salah satu metrik evaluasi yang paling banyak digunakan dalam praktik peramalan di dunia bisnis dan organisasi untuk mengukur akurasi model prediksi. MAPE didefinisikan secara matematis sebagai:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%$$

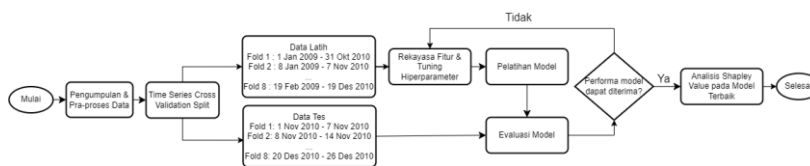
dengan y_i adalah nilai aktual, \hat{y}_i adalah nilai prediksi, dan n adalah jumlah observasi.

2.2 KERANGKA PEMIKIRAN

XGBoost (*Extreme Gradient Boosting*) diakui sebagai salah satu metode *gold-standard* dan *state-of-the-art* untuk pemodelan data tabular (Chen, 2020; Schwerter *et al.*, 2024), yang seringkali menunjukkan performa lebih unggul dibandingkan model-model *deep learning* dengan kebutuhan *tuning* yang lebih minimal (Fayaz *et al.*, 2022). Keunggulan ini menjadi sangat relevan untuk peramalan data deret waktu (*time series*), karena data yang bersifat temporal dapat ditransformasi secara efektif ke dalam format tabular melalui *feature engineering*, sehingga memungkinkan pemanfaatan penuh potensi prediktif XGBoost yang berkinerja tinggi pada data terstruktur (Sumalatha, 2025). Lebih lanjut, untuk mengatasi sifat *black-box* dari model dan meningkatkan interpretabilitas, metode *Shapley Additive Explanations* (SHAP) dapat diintegrasikan untuk memberikan penjelasan yang transparan dan dapat ditafsirkan mengenai dampak serta kontribusi setiap variabel terhadap hasil prediksi, sehingga meningkatkan kredibilitas dan pemahaman terhadap model peramalan (Sumalatha, 2025).

BAB III METODE PENELITIAN

3.1 METODE PENELITIAN



Penelitian ini mengadopsi pendekatan metodologis yang sistematis dan terstruktur untuk mengembangkan model XGBoost rekursif dalam peramalan beban listrik dengan integrasi analisis *Shapley value*. Kerangka metodologi yang diimplementasikan terdiri dari beberapa tahapan utama yang dirancang untuk memastikan rigor ilmiah dan reproduktibilitas hasil penelitian.

1. Tahap Pengumpulan dan Preprocessing Data

Tahap awal penelitian dimulai dengan pengumpulan data historis beban listrik Malaysia periode 2009-2011 beserta variabel eksogen terkait. Proses preprocessing data difokuskan pada memastikan bahwa tidak terdapat null value dalam dataset untuk menjamin integritas dan kelengkapan data yang akan digunakan dalam tahap pemodelan selanjutnya.

2. Implementasi Time Series Cross Validation Split

Untuk memastikan validitas temporal dalam evaluasi model, penelitian ini mengimplementasikan teknik time series cross validation dengan pembagian data secara kronologis. Dataset dibagi menjadi multiple folds berdasarkan periode temporal yang spesifik, yaitu:

- Fold 1: 1 Januari 2009 - 31 Oktober 2010
- Fold 2: 8 Januari 2009 - 7 November 2010
- Fold 3: 15 Januari 2009 - 19 Desember 2010

Strategi pembagian ini dilanjutkan secara konsisten untuk periode selanjutnya, dengan Fold 4 (1 November 2010 - 7 November 2010), Fold 5 (8 November 2010 - 14 November 2010), hingga Fold 8 (20 Desember 2010 - 26 Desember 2010).

Setiap fold training data memiliki corresponding test set yang merupakan periode temporal segera setelah training period, dengan contoh sebagai berikut:

- Test Set 1: 1 November 2010 - 7 November 2010 (untuk evaluasi model yang dilatih pada Fold 1)
- Test Set 2: 8 November 2010 - 14 November 2010 (untuk evaluasi model yang dilatih pada Fold 2)
- Test Set 3: 20 Desember 2010 - 26 Desember 2010 (untuk evaluasi model yang dilatih pada Fold 3)

Pendekatan sliding window cross-validation ini memungkinkan evaluasi yang robust terhadap kemampuan generalisasi model pada data unseen dengan mempertahankan struktur temporal yang inherent dalam data deret waktu.

3. Penentuan Area Hiperparameter

Tahap ini berfokus pada pendefinisian ruang pencarian (*search space*) hiperparameter yang akan dieksplorasi oleh algoritma *Bayesian Optimization*. Penentuan rentang nilai yang tepat untuk setiap hiperparameter merupakan langkah krusial untuk memastikan proses optimasi berjalan efisien dan mampu menemukan konfigurasi yang mendekati optimal global, sehingga dapat meningkatkan akurasi dan kemampuan generalisasi model XGBoost. Ruang pencarian ini dirancang untuk mencakup parameter-parameter yang paling berpengaruh terhadap kompleksitas model, kecepatan konvergensi, dan regularisasi untuk mencegah *overfitting*.

Hiperparameter utama yang dioptimalkan beserta rentang pencariannya adalah sebagai berikut:

- *n_estimators*: Jumlah pohon (*boosting rounds*) yang akan dibangun. Rentang yang lebih besar memungkinkan model yang lebih kompleks, namun meningkatkan risiko *overfitting* dan waktu komputasi.
Ruang Pencarian: [0, 1000] (Integer)
- *learning_rate* (*eta*): Laju pembelajaran yang mengontrol bobot kontribusi dari setiap pohon. Nilai yang lebih kecil membuat proses training lebih robust terhadap *overfitting*.
Ruang Pencarian: [0.01, 0.3] (Log-uniform)

- *max_depth*: Kedalaman maksimum dari setiap pohon. Parameter ini secara langsung mengontrol kompleksitas model.
Ruang Pencarian: [3, 10] (Integer)
- *subsample*: Fraksi dari data training yang digunakan untuk membangun setiap pohon. Ini adalah teknik regularisasi untuk mengurangi varians.
Ruang Pencarian: [0.5, 1.0] (Continuous)
- *colsample_bytree*: Fraksi dari fitur (kolom) yang digunakan saat membangun setiap pohon.
Ruang Pencarian: [0.5, 1.0] (Continuous)
- *reg_alpha* (*L1 Regularization*): Term regularisasi L1 pada bobot. Berguna untuk menangani fitur yang sangat banyak.
Ruang Pencarian: [0.0, 1] (Continuous)
- *reg_lambda* (*L2 Regularization*): Term regularisasi L2 pada bobot. Merupakan metode regularisasi standar untuk mencegah overfitting.
Ruang Pencarian: [0.0, 1] (Continuous)

4. Perencanaan Input Model

Untuk mengevaluasi kontribusi relatif dari berbagai jenis variabel prediktor, penelitian ini merancang serangkaian skenario input model yang sistematis. Perancangan ini dilakukan secara inkremental, dimulai dari model dasar yang menggunakan prediktor endogen berupa fitur kalender (jam, hari, bulan) dan fitur lag dari beban historis (t-1 hingga t-24). Skenario selanjutnya diperkaya dengan menambahkan informasi struktural seperti indikator hari libur dan tren beban jangka menengah (moving average). Setelah itu, model diuji dengan memasukkan variabel eksogen berupa data suhu, termasuk nilai lag dan trennya. Tahap terakhir adalah eksplorasi rekayasa fitur lanjutan untuk menangkap volatilitas (standar deviasi, min/max) serta interaksi non-linear melalui transformasi siklus (sin/cos). Setiap kombinasi fitur ini digunakan untuk melatih dan mengevaluasi model XGBoost, memungkinkan perbandingan performa yang objektif guna mengidentifikasi arsitektur fitur yang paling optimal.

Ringkasan Tabel Skenario Input Model:

ID Model	Deskripsi Singkat	Komponen Fitur Utama
A1	Baseline Kalender	Kalender
A2	Baseline Lag	Lag Beban (1 – 24 Jam)
A3	Lag + Kalender	Lag Beban, Kalender
B1	+ Hari Libut	Lag Beban, Kalender, Hari Libur
B2	+ Tren Beban	Variabel B1 + MA Beban
C1	+ Suhu Dasar	Variabel B2 + Lag Suhu
C2	+ Tren Suhu	Variabel C1 + MA Suhu
D1	+ Volatilitas	Variabel C2 + Rolling Std Dev/Min/Max
D2	+ Interaksi & Siklus	Variabel D1 + Fitur Siklus/Interaksi

5. Rekayasa Fitur dan Optimasi Hiperparameter

Tahap rekayasa fitur melibatkan ekstraksi dan konstruksi variabel prediktor multi-skala yang mencakup komponen temporal (lag features 1-24 jam, moving averages dengan window 3 hari, 1 minggu, 1 bulan, dan 3 bulan), variabel kalender struktural (pola harian, jam operasional, variasi musiman bulanan, indikator hari libur nasional), dan variabel eksogen suhu lingkungan. Proses optimasi hiperparameter dilakukan menggunakan Bayesian Optimization untuk menemukan konfigurasi parameter optimal yang memaksimalkan performa prediksi model XGBoost.

6. Pelatihan dan Evaluasi Model

Model XGBoost dilatih secara iteratif pada setiap fold training data dengan menggunakan konfigurasi hiperparameter yang telah dioptimasi. Evaluasi model dilakukan pada corresponding validation set untuk setiap fold menggunakan Mean Absolute Percentage Error (MAPE) sebagai metrik evaluasi utama. Pemilihan MAPE sebagai single metric evaluasi dimaksudkan untuk memberikan ukuran persentase error yang dapat diinterpretasikan secara intuitif dan memungkinkan perbandingan yang konsisten across different scales of electrical load demand. Proses ini memastikan bahwa model yang

dikembangkan memiliki kemampuan prediksi yang konsisten across different time periods dengan standard evaluasi yang unified dan meaningful.

7. Iterative Model Improvement

Metodologi penelitian ini mengimplementasikan loop iteratif untuk continuous improvement terhadap performa model. Jika hasil evaluasi menunjukkan bahwa performa model belum mencapai threshold yang ditetapkan, proses akan kembali ke tahap rekayasa fitur dan tuning hiperparameter untuk optimasi lebih lanjut. Iterasi ini dilakukan hingga diperoleh model dengan performa prediksi yang optimal dan memenuhi kriteria akurasi yang telah ditetapkan.

8. Analisis Interpretabilitas dengan Shapley Value

Setelah model terbaik diperoleh, tahap final metodologi melibatkan implementasi Shapley Additive Explanations (SHAP) untuk menganalisis kontribusi dan importansi setiap variabel prediktor terhadap hasil prediksi. Analisis SHAP memberikan insight mendalam tentang mekanisme prediksi model XGBoost dan mengungkap interaksi kompleks antar variabel yang mempengaruhi pola konsumsi beban listrik. Hasil analisis ini menyediakan interpretabilitas model yang comprehensive untuk mendukung understanding terhadap faktor-faktor determinan dalam forecasting beban listrik.

3.2 JADWAL PENELITIAN

Kegiatan	September	Oktober	November	Desember
Ujian Proposal TA				
Revisi Proposal TA				
Olah Data				
Penyusunan Skripsi				
Penyusunan Artikel				
Seminar Hasil				
Revisi Artikel				
Ujian Skripsi				
Revisi Skripsi				

DAFTAR PUSTAKA

- Adhikari, R., & Agrawal, R. K. (2013). *An Introductory Study on Time Series Modeling and Forecasting*.
- Arslan, İ., Dağıdır, C. H., & Işlak, Ü. (2023). *An overview of time series point and interval forecasting based on similarity of trajectories, with an experimental study on traffic flow forecasting*.
- Chen, Y. (2020). *Attention augmented differentiable forest for tabular data*.
- Fayaz, S. A., Zaman, M., Kaul, S., & Butt, M. A. (2022). Is Deep Learning on Tabular Data Enough? An Assessment. *International Journal of Advanced Computer Science and Applications*, 13(4). <https://doi.org/10.14569/IJACSA.2022.0130454>
- Fezzi, C., & Fanghella, V. (2021). Tracking GDP in real-time using electricity market data: Insights from the first wave of COVID-19 across Europe. *European Economic Review*, 139, 103907. <https://doi.org/10.1016/j.eurocorev.2021.103907>
- Garg, R., Barpanda, S., S, G. R. S. N., & S, R. (2022). *Machine Learning Algorithms for Time Series Analysis and Forecasting*.
- Hopf, K., Hartstang, H., & Staake, T. (2023). *Meta-Regression Analysis of Errors in Short-Term Electricity Load Forecasting*. <https://doi.org/10.1145/3599733.3600248>
- Jing, B., Zhang, S., Zhu, Y., Peng, B., Guan, K., Margenot, A., & Tong, H. (2022). *Retrieval Based Time Series Forecasting*.
- Khuntia, S. R., Rueda, J. L., & van der Meijden, M. A. M. M. (2017). *Long-Term Load Forecasting Considering Volatility Using Multiplicative Error Model*. <https://doi.org/10.3390/en1123308>
- Pavlyshenko, B. M. (2017). *Linear, Machine Learning and Probabilistic Approaches for Time Series Analysis*.

- Schwerter, J., Romero, A., Dumpert, F., & Pauly, M. (2024). *Which Imputation Fits Which Feature Selection Method? A Survey-Based Simulation Study*.
- Shi, N., & Luo, Y. (2017). Capacity value of energy storage considering control strategies. *PLOS ONE*, 12(5), e0178466. <https://doi.org/10.1371/journal.pone.0178466>
- Su, H.-I., & Gamal, A. El. (2011). *Limits on the Benefits of Energy Storage for Renewable Integration*.
- Sumalatha, Dr. P. (2025). Enhancing House Price Prediction Using Hybrid Model Approach. *International Scientific Journal of Engineering and Management*, 04(05), 1–7. <https://doi.org/10.55041/ISJEM03420>
- Ugurlu, U., Tas, O., Kaya, A., & Oksuz, I. (2018). The Financial Effect of the Electricity Price Forecasts' Inaccuracy on a Hydro-Based Generation Company. *Energies*, 11(8), 2093. <https://doi.org/10.3390/en11082093>
- Wu, J., Levi, N., Araujo, R., & Wang, Y.-G. (2023). An evaluation of the impact of COVID-19 lockdowns on electricity demand. *Electric Power Systems Research*, 216, 109015. <https://doi.org/10.1016/j.epsr.2022.109015>