PROPOSAL TUGAS AKHIR

PERAMALAN BEBAN LISTRIK NEGARA MALAYSIA MENGGUNAKAN 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

# PENDAHULUAN

## 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.

## 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?

## 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.

## MANFAAT PENELITIAN

### 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.

### Manfaat Praktis

#### 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.

#### 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.

#### 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.

#### 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*.

# LANDASAN TEORI

## TINJAUAN PUSTAKA

### 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:

di mana adalah nilai observasi pada waktu , adalah fungsi deterministik dari waktu, dan adalah komponen *error* atau *noise*.

*Time series* umumnya terdiri dari empat komponen utama (Dama & Sinoquet, 2021):

* **Trend:** pola jangka panjang dalam data
* **Seasonality:** pola berulang dalam periode tertentu
* **Cyclical:** fluktuasi jangka panjang tanpa periode tetap
* **Irregular/Random:** komponen acak yang tidak dapat diprediksi

Dekomposisi *time series* dapat dinyatakan dalam bentuk aditif atau multiplikatif:

Model Aditif:

Model Multiplikatif:

di mana , , , dan berturut-turut merepresentasikan komponen *trend*, *seasonal*, *cyclical*, dan *irregular* pada waktu .

### 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:

## di mana adalah *loss function* yang dapat didifferensiasi, adalah *regularization term* untuk tree , dan merepresentasikan parameter model.

*Regularization term* dinyatakan sebagai:

di mana adalah jumlah *leaves*, adalah *weight* pada *leaf* , dan adalah *hyperparameter regularization*.

Untuk iterasi ke-, prediksi model adalah:

di mana adalah *tree* yang ditambahkan pada iterasi .

### 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* , *recursive strategy* dapat diformulasikan sebagai:

di mana adalah fungsi prediksi, adalah *lag order*, dan adalah variabel eksogen.

### Variabel Input

#### Variabel Lag

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

di mana adalah maksimum *lag* yang dipertimbangkan.

#### Variabel Moving Average

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

di mana adalah *window size*.

#### Variabel Kalendar

*Calendar features* menangkap pola temporal struktural:

* **Hour of day:**
* **Day of week:**
* **Month of year:**
* **Holiday indicator:**

#### Variabel Encoding Siklis

Untuk menangkap sifat *cyclical* dari *calendar features*:

### 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).

di mana adalah *mean function* dan adalah *covariance function*.

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

di mana adalah nilai terbaik yang telah ditemukan.

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

di mana , adalah CDF normal standar, dan adalah PDF normal standar.

### Time Series Cross-Validation

#### Expanding Window Cross-Validation

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

di mana meningkat setiap *fold*.

#### Sliding Window Cross-Validation

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

di mana adalah *window size* yang konstan.

### Shapley Additive Explanations (SHAP)

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

di mana adalah *set* semua *features*, adalah *subset features*, dan adalah *expected prediction* ketika hanya menggunakan *features* dalam .

*SHAP values* memenuhi tiga axiom penting:

1. **Efficiency:**
2. **Symmetry:** Jika untuk semua , maka
3. **Dummy:** Jika untuk semua , maka

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

### 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:

di mana adalah nilai aktual, adalah nilai prediksi, dan adalah jumlah observasi.

## KERANGKA PEMIKIRAN

Model XGBoost (*Extreme Gradient Boosting*) telah diakui sebagai salah satu algoritma *machine learning* terdepan untuk pemodelan data tabular, yang dibuktikan dengan dominasinya dalam berbagai kompetisi *machine learning* di *platform* Kaggle (Bentéjac *et al*., 2019; Shwartz-Ziv & Armon, 2021). Keunggulan XGBoost terletak pada kemampuannya dalam memodelkan pola *non-linear* yang kompleks melalui teknik *gradient boosting* yang secara iteratif mempelajari residual dari prediksi sebelumnya hingga mencapai konvergensi optimal (Bentéjac *et al*., 2019; Chen & Guestrin, 2016; März, 2019). *Tree boosting* yang menjadi dasar XGBoost merupakan metode *machine learning* yang sangat efektif dan banyak digunakan oleh *data scientists* untuk mencapai hasil *state-of-the-art* dalam berbagai tantangan *machine learning* (Chen & Guestrin, 2016).

Data time series merupakan jenis data yang memiliki karakteristik temporal, di mana setiap observasi mempertimbangkan komponen waktu sebagai elemen fundamental dalam struktur dan analisisnya (Azorin *et al*., 2024; Shankaranarayana & Runje, 2021). Meskipun data *time series inherently* bersifat sekuensial dan temporal, data tersebut dapat ditransformasi ke dalam format tabular melalui teknik *feature engineering* yang memungkinkan ekstraksi informasi temporal menjadi variabel-variabel *predictor* (Azorin *et al*., 2024; Padhi *et al*., 2020).

Transformasi data *time series* ke dalam bentuk tabular membuka peluang untuk memanfaatkan kekuatan XGBoost dalam konteks *forecasting* (Grinsztajn *et al*., 2022). Pendekatan ini memungkinkan integrasi berbagai variabel eksogen sebagai fitur tambahan yang dapat meningkatkan akurasi prediksi secara signifikan. Dalam konteks *forecasting* beban Listrik (März, 2019), variabel eksternal seperti suhu, kelembaban, hari dalam seminggu, dan faktor musiman dapat diinkorporasikan sebagai prediktor yang memperkaya model dan meningkatkan kemampuan prediktifnya.

Meskipun XGBoost menunjukkan performa prediksi yang superior, algoritma ini memiliki karakteristik *black-box* yang membatasi interpretabilitas model, khususnya dalam memahami kontribusi individual setiap variabel prediktor terhadap hasil prediksi (Lundberg & Lee, 2017). Tantangan interpretabilitas ini menjadi krusial dalam aplikasi *forecasting* beban listrik, di mana pemahaman mendalam tentang faktor-faktor yang mempengaruhi konsumsi energi diperlukan untuk pengambilan keputusan strategis. Untuk mengatasi limitasi ini, *Shapley Additive Explanations* (SHAP) telah dikembangkan sebagai metode atribusi fitur yang konsisten dan akurat, yang mampu mengdekomposisi prediksi model menjadi kontribusi individual dari setiap fitur secara *fair* dan matematis *rigorous* (Lundberg & Lee, 2017; Mayer, 2022).

SHAP values menyediakan *framework* teoritis yang kuat berdasarkan teori game cooperatif untuk menjelaskan output model *machine learning* dengan mengkuantifikasi kontribusi setiap fitur terhadap prediksi dalam bentuk nilai numerik yang dapat dianalisis berdasarkan arah dan magnitude-nya (Yi *et al*., 2023). Implementasi SHAP pada model XGBoost telah terbukti efektif dalam memberikan interpretabilitas yang komprehensif, di mana metode ini dapat mengungkap *not only* kontribusi linear dari setiap variabel, tetapi juga interaksi kompleks antar variabel yang seringkali tidak terdeteksi oleh metode analisis konvensional (Bjerregaard, 2023; Huang *et al*., 2023). Integrasi XGBoost-SHAP *framework* memungkinkan peneliti untuk memperoleh *explicit explanations* terhadap hasil prediksi, yang sangat valuable dalam konteks *critical infrastructure* seperti sistem ketenagalistrikan (Huang *et al*., 2023; Yi *et al*., 2023).

Penelitian terdahulu menunjukkan bahwa meskipun *deep learning* telah mencapai kemajuan luar biasa pada data teks dan gambar, superioritas model tree-based seperti XGBoost pada *data tabular* masih terbukti konsisten (Grinsztajn *et al*., 2022; Shwartz-Ziv & Armon, 2021). Hal ini menjadikan XGBoost sebagai pilihan yang tepat untuk *forecasting* beban listrik, terutama ketika dikombinasikan dengan variabel eksogen yang relevan untuk meningkatkan performa prediksi. Kombinasi kekuatan prediktif XGBoost dengan kemampuan interpretabilitas SHAP menciptakan *framework* yang ideal untuk aplikasi *forecasting* dalam domain energi, di mana akurasi prediksi dan *explainability model* menjadi *equally important* untuk mendukung *decision-making process* yang *informed* dan *accountable*.

# METODE PENELITIAN

## METODE PENELITIAN

A diagram of a flowchart

AI-generated content may be incorrect.

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.

1. **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.

1. **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.

1. **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.

1. **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.

1. **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.

## JADWAL PENELITIAN

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Kegiatan | Bulan 1 | Bulan 2 | Bulan 3 | Bulan 4 |
| 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*.

Azorin, R., Houidi, Z. Ben, Gallo, M., Finamore, A., & Michiardi, P. (2024). *Fine-grained Attention in Hierarchical Transformers for Tabular Time-series*.

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2019). *A Comparative Analysis of XGBoost*. https://doi.org/10.1007/s10462-020-09896-5

Bjerregaard, S. S. (2023). Exploring predictors of welfare dependency 1, 3, and 5 years after mental health-related absence in danish municipalities between 2010 and 2012 using flexible machine learning modelling. *BMC Public Health*, *23*(1), 224. https://doi.org/10.1186/s12889-023-15106-y

Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*. https://doi.org/10.1145/2939672.2939785

Dama, F., & Sinoquet, C. (2021). *Time Series Analysis and Modeling to Forecast: a Survey*.

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.euroecorev.2021.103907

Garg, R., Barpanda, S., S, G. R. S. N., & S, R. (2022). *Machine Learning Algorithms for Time Series Analysis and Forecasting*.

Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). *Why do tree-based models still outperform deep learning on tabular data?*

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

Huang, J., Chen, H., Deng, J., Liu, X., Shu, T., Yin, C., Duan, M., Fu, L., Wang, K., & Zeng, S. (2023). Interpretable machine learning for predicting 28-day all-cause in-hospital mortality for hypertensive ischemic or hemorrhagic stroke patients in the ICU: a multi-center retrospective cohort study with internal and external cross-validation. *Frontiers in Neurology*, *14*. https://doi.org/10.3389/fneur.2023.1185447

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/en11123308

Lundberg, S. M., & Lee, S.-I. (2017). *Consistent feature attribution for tree ensembles*.

März, A. (2019). *XGBoostLSS -- An extension of XGBoost to probabilistic forecasting*.

Mayer, M. (2022). *SHAP for additively modeled features in a boosted trees model*.

Padhi, I., Schiff, Y., Melnyk, I., Rigotti, M., Mroueh, Y., Dognin, P., Ross, J., Nair, R., & Altman, E. (2020). *Tabular Transformers for Modeling Multivariate Time Series*.

Pavlyshenko, B. M. (2017). *Linear, Machine Learning and Probabilistic Approaches for Time Series Analysis*.

Shankaranarayana, S. M., & Runje, D. (2021). *Attention Augmented Convolutional Transformer for Tabular Time-series*.

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

Shwartz-Ziv, R., & Armon, A. (2021). *Tabular Data: Deep Learning is Not All You Need*.

Su, H.-I., & Gamal, A. El. (2011). *Limits on the Benefits of Energy Storage for Renewable Integration*.

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

Yi, F., Yang, H., Chen, D., Qin, Y., Han, H., Cui, J., Bai, W., Ma, Y., Zhang, R., & Yu, H. (2023). XGBoost-SHAP-based interpretable diagnostic framework for alzheimer’s disease. *BMC Medical Informatics and Decision Making*, *23*(1), 137. https://doi.org/10.1186/s12911-023-02238-9