

Pengaruh Panjang Data pada Holt-Winters dan LSTM dalam Knowledge Discovery Peramalan Konsumsi Listrik

The Effect of Data Length on Holt-Winters and LSTM in Knowledge Discovery for Electricity Consumption Forecasting

¹Dimas Agil Kusuma, ²Ken Ditha Tania*, ³Yadi Utama

^{1,2,3}Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Sriwijaya

^{1,2,3}Jl. Srijaya Negara, Bukit Besar, Palembang, Sumatera Selatan 30139 - Indonesia

*e-mail: kenya.tania@gmail.com

(received: 17 August 2026, revised: 28 September 2026, accepted: 29 September 2026)

Abstrak

Peramalan konsumsi energi listrik merupakan penerapan *knowledge discovery* yang dipengaruhi panjang data historis, meskipun penambahan volume data tidak selalu meningkatkan performa dan justru membebani komputasi. Penelitian ini menganalisis pengaruh variasi panjang data latih (1, 3, 5, 7, dan 9 tahun) terhadap akurasi peramalan dengan membandingkan Holt-Winters dan *Long Short-Term Memory* (LSTM) pada *dataset* PJME harian, menggunakan data uji 90 hari tetap, dievaluasi melalui MAE, RMSE, MAPE, dan waktu komputasi. Hasil menunjukkan Holt-Winters mencapai titik optimal pada data latih sembilan tahun (MAPE 9,37%), sedangkan LSTM optimal pada tiga tahun (MAPE 9,69%) namun membutuhkan waktu komputasi 51 kali lebih lama. Pola *crossover* ini mengindikasikan performa Holt-Winters ditentukan representativitas estimasi awal komponen musiman, sedangkan LSTM dipengaruhi akumulasi galat pada skema *recursive multi-step*, menegaskan kebutuhan panjang data historis bergantung pada karakteristik metode. Kontribusi penelitian adalah desain eksperimen pembandingan adil yang mengisolasi pengaruh panjang data latih dari efek horizon peramalan, desain yang belum diterapkan pada domain konsumsi listrik skala regional, sehingga mengisi kesenjangan metodologis pada penelitian sejenis. Secara praktis, temuan ini memberikan pedoman konkret bagi penyedia layanan energi dalam menentukan strategi pengelolaan data historis yaitu Holt-Winters direkomendasikan bagi penyedia dengan riwayat data panjang namun infrastruktur komputasi terbatas, sedangkan LSTM lebih sesuai bagi penyedia dengan data terbatas namun sumber daya komputasi memadai, sehingga investasi data dan komputasi dapat diarahkan lebih efisien.

Kata kunci: holt-winters, *knowledge discovery*, konsumsi energi listrik, *long short-term memory*, panjang data historis

Abstract

Electricity consumption forecasting is an application of knowledge discovery influenced by the length of historical data, even though adding more data does not always improve performance and can instead increase computational burden. This study analyzes the effect of varying training data lengths (1, 3, 5, 7, and 9 years) on forecasting accuracy by comparing Holt-Winters and Long Short-Term Memory (LSTM) on the daily PJME electricity consumption dataset, using a fixed 90-day test set, evaluated through MAE, RMSE, MAPE, and computation time. Results show that Holt-Winters achieves its optimal point at nine years of training data (MAPE 9.37%), while LSTM is optimal at three years (MAPE 9.69%) but requires 51 times longer computation time. This crossover pattern indicates that Holt-Winters performance is determined by the representativeness of the initial seasonal component estimate, whereas LSTM is affected by error accumulation in the recursive multi-step forecasting scheme, confirming that the required length of historical data depends on the method's characteristics. The contribution of this study is a fair comparative experimental design that isolates the effect of training data length from the effect of the forecasting horizon, a design not previously applied to regional-scale electricity consumption forecasting, thereby filling a methodological gap in similar studies. Practically, these findings provide concrete guidance for energy providers in determining historical data management strategies Holt-Winters is recommended

<http://sistemasi.ftik.unisi.ac.id>

for providers with long historical data but limited computational infrastructure, while LSTM is more suitable for providers with limited historical data but adequate computational resources, allowing data and computational investments to be directed more efficiently.

Keywords: electricity consumption, historical data length, holt-winters, knowledge discovery, long short-term memory

1 Pendahuluan

Konsumsi energi listrik adalah variabel deret waktu krusial dalam pengelolaan sistem tenaga, karena akurasi peramalannya menentukan efektivitas perencanaan pembangkitan, distribusi, dan kebijakan efisiensi energi. Dalam konteks ini, peramalan konsumsi energi listrik dapat dipandang sebagai salah satu penerapan utama *knowledge discovery in databases* (KDD), yaitu proses sistematis mengekstraksi pola pengetahuan potensial yang bermanfaat dari data mentah untuk mendukung pengambilan keputusan organisasi maupun instansi terkait kelistrikan [1]. Sifat data yang dipengaruhi cuaca, ekonomi, dan faktor geografis membuatnya nonlinier dan tidak stabil, sehingga peningkatan akurasi peramalan tetap menjadi tantangan penelitian hingga saat ini [2]. Estimasi yang terlalu tinggi berisiko menimbulkan pembangunan infrastruktur pembangkit yang tidak efisien, sementara estimasi terlalu rendah dapat menyebabkan kekurangan pasokan dan ketidakpuasan pelanggan [3].

Persoalan mendasar dalam peramalan deret waktu energi listrik tidak hanya terletak pada pemilihan algoritma, tetapi juga pada konfigurasi panjang data historis yang digunakan untuk melatih model. Metode statistik klasik seperti Holt-Winters menangkap komponen level, tren, dan musiman melalui tiga parameter pemulusan sehingga unggul pada data dengan pola musiman stabil [4], [5], sedangkan metode *deep learning* seperti *Long Short-Term Memory* (LSTM) mengatasi keterbatasan jaringan saraf konvensional dalam mempelajari dependensi jangka panjang melalui mekanisme *forget gate*, *input gate*, dan *output gate* [3]. Namun, penambahan volume data pelatihan tidak selalu berbanding lurus dengan akurasi; pada titik tertentu, data tambahan hanya meningkatkan beban komputasi tanpa perbaikan performa yang sepadan, bahkan dapat menurunkan generalisasi model terhadap pola terbaru [6]. Hyndman & Athanasopoulos [7] turut mencatat bahwa model peramalan cenderung tidak optimal pada deret waktu sangat panjang karena proses yang mendasari data berubah seiring waktu. Zhang *et al.* [8] menambahkan bahwa ukuran sampel data berperan penting bagi performa model *machine learning* maupun *deep learning*, namun hubungan keduanya tidak selalu linear.

Sejumlah penelitian terdahulu telah menguji hubungan panjang data pelatihan dengan akurasi peramalan pada berbagai metode dan domain, mulai dari perbandingan Holt-Winters dan LSTM pada data curah hujan maupun penjualan musiman [4], [5], hingga pengujian khusus pengaruh durasi data pelatihan pada beban bangunan dan pasar harga listrik [6], [9] (dibahas lebih rinci pada Tinjauan Literatur). Namun, penelitian yang secara khusus menguji pengaruh panjang data pelatihan pada domain konsumsi listrik skala regional dengan desain pengujian yang mengisolasi murni efek panjang data latih terlepas dari efek horizon peramalan masih jarang ditemukan. Studi terdekat seperti Royal *et al.* [10] pada *district energy* kampus masih memiliki keterbatasan desain pengujian (dirinci pada Tinjauan Literatur), padahal pemahaman *trade-off* ini penting bagi penyedia layanan energi dalam merancang strategi peramalan yang efisien.

Merespons kesenjangan penelitian yang ada, studi ini mengkaji dampak perbedaan panjang data historis terhadap akurasi peramalan dan beban komputasi konsumsi listrik menggunakan algoritma Holt-Winters dan LSTM. Fokus utamanya adalah menemukan titik keseimbangan paling ideal antara volume data, tingkat keakuratan, dan efisiensi komputasi. Secara teoretis, penelitian ini memperjelas tingkat sensitivitas metode statistik dan *deep learning* terhadap ukuran data. Secara praktis, hasilnya diharapkan menjadi panduan bagi penyedia energi dalam mengoptimalkan strategi pengelolaan serta kebijakan retensi data (*data governance*). Studi ini juga memetakan kesesuaian metode berdasarkan profil organisasi yaitu membandingkan solusi bagi subunit yang memiliki data melimpah namun komputasi terbatas, dengan subunit yang datanya sedikit namun memiliki komputasi memadai. Pada akhirnya, riset ini tidak sekadar mengukur performa teknis, tetapi menawarkan landasan praktis untuk mendukung pengambilan keputusan berbasis data (*data-driven decision making*) di sektor sistem informasi kelistrikan.

2 Tinjauan Literatur

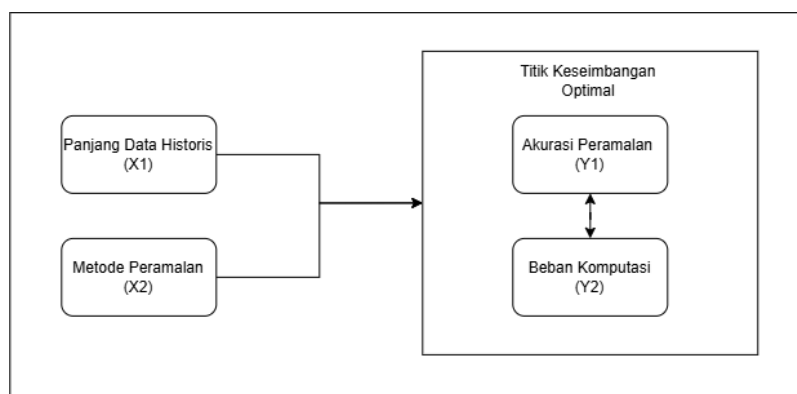
Studi komparatif antara metode statistik klasik dan *deep learning* dalam peramalan deret waktu menunjukkan hasil yang tidak konsisten, bergantung pada karakteristik data yang diuji. Citrayanti *et al.* [4] melaporkan bahwa Holt-Winters Additive (MAPE 47,69%) mengungguli LSTM (MAPE 79,78%) pada data curah hujan dengan riwayat historis terbatas, sementara Firdaus *et al.* [5] menemukan SARIMA mengungguli kedua varian Holt-Winters pada data penjualan musiman (MASE 0,216). Pada domain lain, Rachmatullah [11] menunjukkan *Multilayer Perceptron* (MLP) lebih stabil dibandingkan LSTM pada *dataset* berkompleksitas rendah-sedang, sedangkan Zhou *et al.* [2] justru mencatat bahwa LSTM yang dikombinasikan dengan dekomposisi sinyal dan optimasi *hyperparameter* mampu mencapai R^2 0,98 pada peramalan beban jangka pendek. Keempat temuan ini memperlihatkan bahwa keunggulan salah satu pendekatan bersifat kontekstual: metode statistik klasik cenderung kompetitif pada data berpola musiman jelas namun riwayat terbatas, sementara *deep learning* lebih unggul ketika didukung teknik pra-pemrosesan tambahan dan volume data yang memadai namun belum ada dari keempat studi ini yang secara khusus mengisolasi pengaruh panjang data historis sebagai variabel yang dimanipulasi.

Beberapa penelitian menguji secara lebih spesifik bagaimana durasi data pelatihan memengaruhi akurasi. Simani *et al.* [6] menemukan bahwa galat model LSTM untuk peramalan beban bangunan *grid-interactive* meluruh secara eksponensial terhadap ukuran data pelatihan, dan setelah durasi sekitar 90 hari, penambahan data hanya memberi peningkatan akurasi marjinal. Michalakopoulos *et al.* [9] menunjukkan pola serupa pada peramalan harga listrik *day-ahead*, menemukan bahwa *training window* berdurasi menengah (45–60 hari) memberi keseimbangan lebih baik dibandingkan *window* 90 hari mengindikasikan bahwa data pelatihan yang lebih panjang tidak otomatis lebih baik. Pandangan yang sejalan juga tersirat dari Moosbrugger *et al.* [12], yang secara eksplisit mempertanyakan apakah kompleksitas tambahan model *deep learning* untuk peramalan beban rumah tangga dan komunitas energi sepadan dengan perbaikan akurasi yang dihasilkan, dengan menerapkan *test set* berdurasi satu kuartal secara konsisten di seluruh variasi panjang data latih yang diuji. Ketiga studi ini secara konsisten mengindikasikan adanya titik jenuh atau bahkan potensi penurunan performa akibat penambahan data pelatihan yang berlebihan, namun ketiganya dilakukan pada skala data yang jauh lebih kecil (bangunan tunggal, pasar harga tahunan, rumah tangga) dibandingkan skala regional yang menjadi fokus penelitian ini.

Dari sisi desain eksperimen, penelitian yang menguji variasi panjang data latih secara terkontrol masih terbatas. Royal *et al.* [10] pada studi kasus *district energy* Colorado School of Mines menguji panjang data latih 1–12 tahun menggunakan GAM-SARIMA dan LSTM, menemukan bahwa *training set* 3 dan 5 tahun justru menghasilkan galat uji lebih tinggi dibanding *training set* 1 tahun pada kedua metode namun penelitian tersebut memvariasikan panjang data latih dan data uji secara berbalik dengan total durasi tetap (~10–12 tahun), sehingga efek panjang data latih berpotensi tercampur dengan efek panjang horizon peramalan. Pendekatan pembagian data yang lebih terkontrol ditunjukkan oleh Bayram *et al.* [13], yang menetapkan data uji tetap terlebih dahulu sebelum membentuk data validasi sebagai pecahan dari data latih yang tersisa, sehingga proporsi pembagian data tetap konsisten di seluruh skenario pengujian. Di sisi arsitektur, tren pada literatur *load forecasting* terkini justru mengarah pada kompleksitas model yang meningkat, seperti jaringan *hybrid* LSTM-TCN dengan mekanisme *residual* dan atensi [14] maupun LSTM *multilayer* berjenis *dilated* dengan mekanisme *attention* [15], yang keduanya menambahkan komponen arsitektural demi perbaikan akurasi tanpa menjadikan panjang data pelatihan sebagai variabel penelitian utama.

Berdasarkan tinjauan tersebut, penelitian yang secara khusus menguji pengaruh panjang data pelatihan terhadap akurasi peramalan Simani *et al.* [6] pada bangunan *grid-interactive*, Michalakopoulos *et al.* [9] pada pasar harga *day-ahead*, dan Royal *et al.* [10] pada *district energy* kampus berbeda dari konteks penelitian ini dalam beberapa hal mendasar. Pertama, metode statistik pembandingan: Royal *et al.* [10] menggunakan GAM-SARIMA, bukan Holt-Winters, sehingga sensitivitas mekanisme pemulusan level-tren-musiman (α , β , γ) terhadap variasi panjang data belum diuji pada domain serupa. Kedua, domain dan skala data: ketiga studi berfokus pada skala bangunan tunggal [6], pasar harga berdurasi satu tahun kalender [9], dan *district energy* satu kampus [10], sedangkan konsumsi listrik skala regional/multi-wilayah dengan rentang historis jauh lebih panjang (PJME, $\pm 16,7$ tahun) belum diuji.

Ketiga, desain eksperimen: Royal *et al.* [10] memvariasikan panjang data latih dan uji secara berbalik dengan total durasi tetap, sehingga efek panjang data latih berpotensi tercampur dengan efek horizon peramalan, sedangkan penelitian ini mempertahankan panjang data uji tetap (90 hari) di seluruh skenario mengikuti prinsip data uji tetap yang juga diterapkan Bayram *et al.* [13] dan Moosbrugger *et al.* [12] agar efek panjang data latih dapat diisolasi lebih terkontrol. Keempat, kompleksitas arsitektur: berbeda dari tren literatur yang mengarah pada model LSTM semakin kompleks [14], [15], penelitian ini sengaja mempertahankan arsitektur LSTM tunggal yang sederhana dan konfigurasi tetap di seluruh skenario, agar perbedaan akurasi antar-skenario dapat diatribusikan murni pada pengaruh panjang data historis, bukan pada perbedaan *tuning*, evaluasi antar-metode, atau kompleksitas model. Keempat perbedaan ini membuat temuan-temuan terdahulu belum tentu tergeneralisasi ke konteks perbandingan Holt-Winters dan LSTM pada domain skala luas seperti PJME, sehingga *novelty* penelitian ini terletak pada desain eksperimen pembandingan yang adil (*apple-to-apple*) antara metode statistik klasik dan *deep learning* pada domain dan skala data yang belum diuji secara khusus oleh literatur sebelumnya.

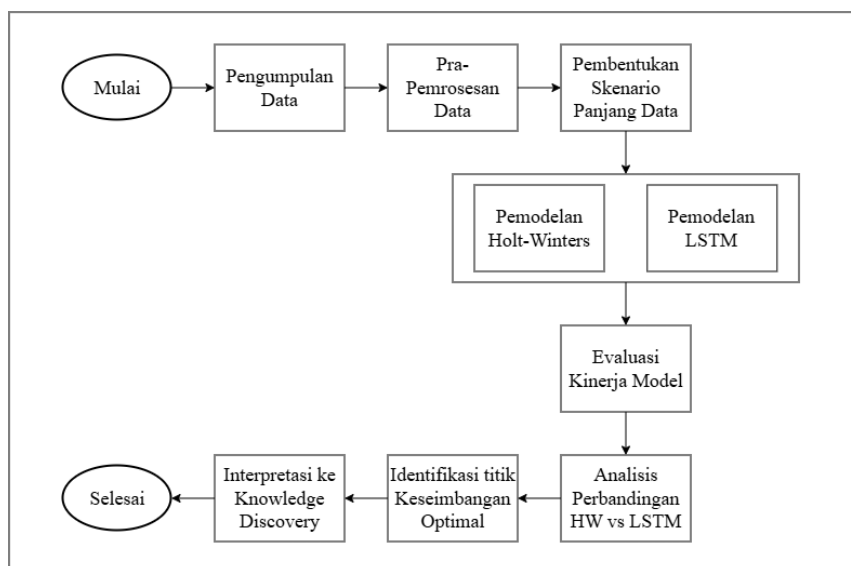


Gambar 1 Kerangka konseptual

Berdasarkan keempat kesenjangan tersebut, terlihat bahwa literatur yang ada belum secara jelas membahas hubungan antara panjang data historis, karakteristik metode peramalan, akurasi, dan beban komputasi secara bersamaan pada metode statistik maupun deep learning. Karena itu, penelitian ini membuat kerangka konseptual (Gambar 1) yang menghubungkan panjang data historis (X1) dan metode peramalan (X2) terhadap akurasi peramalan (Y1) dan beban komputasi (Y2). Hipotesisnya, metode peramalan tidak hanya menjadi faktor tambahan. Metode ini juga ikut menentukan arah dan besar pengaruh panjang data historis terhadap kedua variabel hasil tersebut. Holt-Winters diduga dipengaruhi oleh kualitas estimasi awal pada komponen musiman, yang membaik saat jumlah siklus musiman penuh dalam data latih bertambah. LSTM diduga bergantung pada kemampuannya dalam menghadapi akumulasi galat pada skema peramalan recursive multi-step. Kemampuan ini bisa menurun jika data latih semakin panjang. Interaksi antara panjang data historis dan karakteristik khusus metode peramalan inilah yang akhirnya menentukan titik keseimbangan optimal antara akurasi peramalan dan beban komputasi.

3 Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimental, di mana panjang data historis dimanipulasi ke dalam lima skenario (1, 3, 5, 7, dan 9 tahun) untuk mengamati pengaruhnya terhadap akurasi dan efisiensi komputasi model. Tahapan penelitian divisualisasikan pada Gambar 2 Tahapan diawali dengan pengumpulan data deret waktu *PJME_hourly* dari platform Kaggle, dilanjutkan pra-pemrosesan (penanganan anomali *Daylight Saving Time* dan agregasi ke granularitas harian), lalu pembentukan lima skenario panjang data historis yang masing-masing dipartisi menjadi data latih, validasi, dan uji.



Gambar 2 Tahapan Penelitian

Penerapan metode dijalankan pada dua jalur pemodelan paralel yang independen namun setara. Pada jalur pertama, Holt-Winters diterapkan melalui estimasi parameter pemulusan otomatis pada varian *additive* dan *multiplicative*, dilanjutkan pemilihan varian terbaik berdasarkan galat terkecil pada data validasi. Pada jalur kedua, LSTM diterapkan melalui normalisasi *Min-Max Scaling*, pengaturan *hyperparameter*, dan pelatihan dengan lima kali pengulangan *seed* acak untuk mengakomodasi variabilitas, diakhiri peramalan *recursive multi-step* sepanjang 90 hari. Kedua jalur kemudian dievaluasi secara eksklusif pada data uji dan dibandingkan untuk memperoleh titik keseimbangan optimal antara panjang data, akurasi, dan beban komputasi.

3.1 Data Penelitian

Data penelitian bersumber dari *dataset* "Hourly Energy Consumption" di *platform* Kaggle, secara spesifik berkas *PJME_hourly.csv* yang merekam konsumsi energi listrik agregat berskala regional (MW) dari PJM Interconnection LLC di belasan negara bagian Amerika Serikat[16]. Data *univariate* ini tersusun atas kolom waktu (*Datetime*) dan konsumsi daya (*PJME_MW*), bergranularitas per jam sejak 1 Januari 2002 hingga 3 Agustus 2018, menghasilkan total 145.366 observasi. Durasi historis 16 tahun 7 bulan ini jauh lebih panjang dibandingkan *dataset* yang lazim digunakan pada penelitian sejenis (Simani *et al.* [6], Li *et al.*[14], Michalakopoulos *et al.* [9], Zhou *et al.* [2]), sehingga lebih fleksibel untuk menguji variasi rentang data pelatihan. Perlu ditekankan bahwa karena data ini merupakan agregasi beban regional berskala makro tanpa variabel eksogen (seperti cuaca atau indikator ekonomi), temuan penelitian ini paling relevan untuk konteks perencanaan sistem tenaga regional; generalisasi terhadap beban berskala granular seperti rumah tangga belum tentu berlaku langsung mengingat perbedaan volatilitas data pada skala tersebut.

3.2 Pra Proses Data

Sebelum digunakan untuk pemodelan, data mentah terlebih dahulu melalui tahap pembersihan guna menjamin keandalan hasil peramalan [10], khususnya untuk mengatasi anomali interval waktu akibat penyesuaian *Daylight Saving Time* (DST). Pemeriksaan awal menemukan 30 penanda waktu yang hilang (mayoritas akibat *spring forward* di bulan Maret, ditambah satu gangguan pada 10 Desember 2010) serta 8 penanda waktu terduplikasi (akibat *fall back* di bulan November 2014 – 2017). Data duplikat ditangani dengan membuang observasi kedua, sementara data hilang diisi menggunakan interpolasi linear mengikuti pendekatan Michalakopoulos *et al.* [9], dipilih sebagai alternatif dari substitusi hari sebelumnya [5]. Deret waktu per jam yang telah bersih kemudian diagregasi menjadi granularitas harian melalui rata-rata (*mean*) 24 jam, menghasilkan deret waktu kontinu sebanyak 6.059 observasi (1 Januari 2002–3 Agustus 2018), guna menyeimbangkan kemampuan model menangkap pola musiman jangka menengah-panjang dengan efisiensi komputasi, sejalan dengan standar pra-pemrosesan pada penelitian sejenis [10].

3.3 Skema Eksperimen

Skema eksperimen dirancang untuk menguji sensitivitas metode Holt-Winters dan LSTM terhadap lima variasi panjang data historis: 1 tahun (365 hari), 3 tahun (1.095 hari), 5 tahun (1.825 hari), 7 tahun (2.555 hari), dan 9 tahun (3.285 hari) seluruh skenario berdurasi minimal satu tahun agar mencakup satu siklus musiman penuh ($m=365$) yang diperlukan Holt-Winters. Jendela input (*window*) LSTM tetap ditetapkan sepanjang 30 hari pada seluruh skenario mengikuti prinsip *sliding time window* [10]. Kombinasi rentang 1–9 tahun ini sejalan dengan desain panjang data latih yang diuji Royal *et al.* [10] dan Simani *et al.* [6], namun berbeda dari keduanya, penelitian ini mempertahankan panjang data uji tetap 90 hari di seluruh skenario agar efek panjang data latih dapat diisolasi lebih terkontrol dari efek horizon peramalan. Rincian kelima skenario beserta pembagian data latih, validasi, dan uji disajikan pada Tabel 1.

Tabel 1 Skema eksperimen

Skenario	Panjang Data Latih	Panjang Data Validasi	Data Uji
S1 (1 thn)	365 hari	36 hari	90 hari
S2 (3 thn)	1.095 hari	110 hari	90 hari
S3 (5 thn)	1.825 hari	182 hari	90 hari
S4 (7 thn)	2.555 hari	256 hari	90 hari
S5 (9 thn)	3.285 hari	328 hari	90 hari

Data uji ditetapkan tetap sepanjang 90 hari dengan periode kalender identik di seluruh skenario (6 Mei– 3 Agustus 2018) mengikuti pendekatan *test set* tetap Moosbrugger *et al.* [12]; yang bervariasi hanya titik awal data latih beserta panjang data validasi yang mengikutinya secara proporsional (10% dari panjang data latih), sehingga perbedaan akurasi antar-skenario dapat diatribusikan murni pada pengaruh panjang data latih. Validasi ditempatkan sebagai lapisan terpisah yang mengikuti data latih, alih-alih dipotong dari dalamnya (mengacu pada Bayram *et al.* [13]), agar seluruh skenario termasuk S1 tetap mempertahankan panjang data latih penuh sesuai durasi berlabel. Data validasi digunakan berbeda pada tiap metode untuk menjaga kesetaraan pengujian: pada Holt-Winters, validasi memilih varian (*additive/multiplicative*) dengan performa terbaik; pada LSTM, validasi memantau kurva *loss* untuk mendeteksi *overfitting/underfitting* akibat kebijakan *epoch* tetap.

Arsitektur dan *hyperparameter* LSTM ditentukan melalui *tuning* terpisah pada *subset* 180 hari pertama *dataset* (Januari–Juni 2002) yang tidak tumpang tindih dengan data latih maupun uji kelima skenario, guna menghindari sirkularitas *tuning* [6]. Konfigurasi ini dikunci dan diterapkan identik pada S1–S5, termasuk *epoch* tetap sebagai pengganti *early stopping*. Seluruh proses pelatihan dan peramalan kedua metode dijalankan pada spesifikasi perangkat keras dan lingkungan komputasi yang sama, dengan waktu dicatat menggunakan *timer*. Untuk mengatasi variabilitas inisialisasi bobot acak, setiap skenario LSTM dijalankan 5 kali pengulangan dengan *seed* berbeda (dilaporkan rata-rata $\pm$ standar deviasi); Holt-Winters yang deterministik dilaporkan dari satu kali estimasi per skenario per varian, namun waktu komputasinya tetap diukur 3 kali pengulangan.

Pemilihan nilai akhir tiap *hyperparameter* didasarkan pada kombinasi galat validasi terendah dan pertimbangan efisiensi komputasi selama proses *tuning* pada subset 180 hari tersebut. Window 30 hari dipilih karena dianggap cukup untuk menangkap pola dependensi jangka pendek pada data harian tanpa membuat dimensi input terlalu besar. Prinsip *sliding time window* ini juga digunakan oleh Royal *et al.* [10]. Epoch 50 ditetapkan karena pada titik ini training loss sudah stabil pada subset *tuning*. Tidak ada tanda *overfitting* yang berarti pada tahap tersebut. Oleh karena itu, epoch 50 dipilih sebagai pengganti *early stopping* supaya konfigurasi tetap sama di semua skenario. Batch size 32 dan learning rate 0,001 adalah kombinasi yang sering dipakai dan memberikan konvergensi stabil pada arsitektur LSTM berlapis tunggal [17]. Sementara itu, 64 unit pada hidden layer dengan aktivasi tanh dipilih sebagai titik tengah antara kapasitas model dan risiko *overfitting*. Arsitektur dalam penelitian ini sengaja dibuat sederhana supaya perbedaan akurasi antar-skenario murni berasal dari pengaruh panjang data historis, bukan dari kompleksitas arsitektur.

Peramalan pada 90 hari data uji dilakukan tanpa koreksi dari data aktual di tengah horizon. Mekanisme peramalan multi-langkah diterapkan berbeda pada kedua metode sesuai sifat alamnya [6]: Holt-Winters menghasilkan ramalan *direct* untuk periode $t+m$ mana pun dalam satu perhitungan tanpa iterasi, sedangkan LSTM diterapkan secara *recursive* 30 hari terakhir data aktual digunakan

<http://sistemasi.ftik.unisi.ac.id>

sebagai *input* awal, dan hasil ramalan (bukan nilai aktual) disertakan ke jendela *input* untuk meramalkan hari berikutnya, diulang hingga 90 hari terlampaui. Pendekatan ini dipilih agar kedua metode setara: sama-sama tidak pernah mengakses nilai aktual selama periode uji. Konsekuensinya, pendekatan *recursive* pada LSTM secara inheren berisiko mengalami akumulasi galat (*error accumulation*) sepanjang horizon [6], karakteristik yang menjadi salah satu fokus analisis penelitian ini.

3.4 Metode Holt Winters

Metode Holt-Winters, atau *Triple Exponential Smoothing* (TES), merupakan pengembangan *exponential smoothing* yang menangkap komponen level, tren, dan musiman secara simultan melalui tiga parameter pemulusan: α untuk level, β untuk tren, dan γ untuk musiman, yang diperbarui rekursif pada setiap periode mengikuti formulasi standar Holt-Winters [7]. Periode musiman (m) ditetapkan sebesar 365 mengikuti siklus musiman tahunan pada data konsumsi listrik harian [6]; pada skenario S1 (1 tahun), rasio antara jumlah observasi (365 hari) dan parameter musiman ($m=365$) berada pada titik paling ketat di antara seluruh skenario, sehingga keterbatasan ini sengaja dipertahankan sebagai bagian dari fenomena yang diamati (dibahas pada subbagian Evaluasi Kinerja Metode Holt-Winters), bukan dihindari.

Dua varian diuji pada setiap skenario: *additive*, yang lebih sesuai untuk variasi musiman relatif konstan, dan *multiplicative*, yang lebih sesuai ketika variasi musiman berubah proporsional terhadap level deret waktu [6]. Varian terbaik dipilih berdasarkan MAE, RMSE, dan MAPE terendah pada data validasi (10% dari panjang data latih), bukan performa data uji, agar proses pemilihan model tidak bocor ke tahap evaluasi akhir. Ketiga parameter pemulusan diestimasi secara otomatis melalui optimasi internal yang meminimalkan galat pada data latih tiap skenario, dengan nilai awal (*initial state*) komponen level, tren, dan musiman ditetapkan secara heuristik berdasarkan regresi linear dan rata-rata dekomposisi data latih. Perlu dicatat bahwa kombinasi inisialisasi heuristik dengan kriteria optimasi yang murni meminimalkan galat *training* berpotensi mengarahkan estimasi parameter ke solusi batas (*boundary solution*), terutama pada deret waktu berautokorelasi tinggi seperti data ini. Formulasi lengkap varian *additive* maupun *multiplicative* mengikuti Hyndman & Athanasopoulos [7].

3.5 Metode LSTM (Long Short Term Memory)

Berbeda dengan Holt-Winters yang mengandalkan pemulusan eksponensial, LSTM lahir dari ranah *deep learning* untuk mengatasi kelemahan *recurrent neural network* (RNN) konvensional dalam mempelajari dependensi jangka panjang pada data sekuensial akibat masalah *vanishing/exploding gradient* [13]. Kemampuan LSTM mengatasi masalah ini terletak pada struktur *memory cell*-nya: tiga gerbang (*gate*) yang mengendalikan arus informasi secara selektif *forget gate* mengatur informasi yang dihapus dari *cell state*, *input gate* mengontrol seberapa besar data baru dimasukkan, dan *output gate* menentukan bagian *cell state* yang dikeluarkan sebagai *output* pada setiap langkah waktu [2]. Ketiga gerbang diperbarui secara bertahap sepanjang pelatihan melalui propagasi balik galat (*back-propagation*) menggunakan *optimizer* Adam [17]. Formulasi matematis lengkap ketiga gerbang serta pembaruan *cell state* dan *hidden state* mengikuti Chollet [17] dan Zhou et al. [2].

Skema *sliding window*, *hyperparameter*, dan alasan penetapannya tetap di seluruh skenario mengikuti desain yang telah dijelaskan pada subbagian Skema Eksperimen. Sebelum dimasukkan ke model, nilai PJME_MW terlebih dahulu dinormalisasi ke rentang [0,1] menggunakan *Min-Max Scaling*, mengingat fungsi aktivasi sigmoid dan tanh pada LSTM bekerja optimal pada rentang nilai yang relatif kecil dan seragam. Parameter penskalaan (nilai minimum dan maksimum) dihitung khusus dari data latih pada setiap skenario, kemudian diterapkan ke data validasi dan data uji yang bersesuaian, guna mencegah kebocoran informasi (*data leakage*). Seluruh nilai ramalan LSTM, termasuk hasil skema *recursive* sepanjang 90 hari, dikembalikan ke satuan MW melalui *inverse transform* sebelum dihitung metrik evaluasinya, sehingga hasil MAE dan RMSE yang dilaporkan berada pada skala yang sama dengan Holt-Winters.

3.6 Metrik Evaluasi

Evaluasi kinerja model peramalan pada penelitian ini menggunakan tiga metrik kesalahan yang umum digunakan pada penelitian peramalan deret waktu, yaitu *Mean Absolute Error* (MAE), *Root*

<http://sistemasi.ftik.unisi.ac.id>

Mean Square Error (RMSE), dan *Mean Absolute Percentage Error* (MAPE) [17], [13], [1], dihitung eksklusif dari data uji sesuai prinsip bahwa akurasi peramalan yang andal hanya terukur pada data yang belum pernah dilihat model [6]. MAE mengukur rata-rata kesalahan absolut antara nilai prediksi dan nilai aktual, RMSE mengukur akar kuadrat dari rata-rata kesalahan kuadrat, sedangkan MAPE mengukur kesalahan dalam bentuk persentase relatif terhadap nilai aktual [2]. Ketiga metrik dirumuskan sebagaimana persamaan 1, 2 dan 3 berikut:

$$MAE = (1/n) \sum |y_i - F_i| \quad (1)$$

$$RMSE = \sqrt{(1/n) \sum (y_i - F_i)^2} \quad (2)$$

$$MAPE = (1/n) \sum |y_i - F_i| / y_i \times 100\% \quad (3)$$

dengan y_i adalah nilai aktual, F_i adalah nilai hasil prediksi, dan n adalah jumlah total data yang diramalkan; semakin kecil nilai ketiganya [18], semakin akurat model peramalan [18].

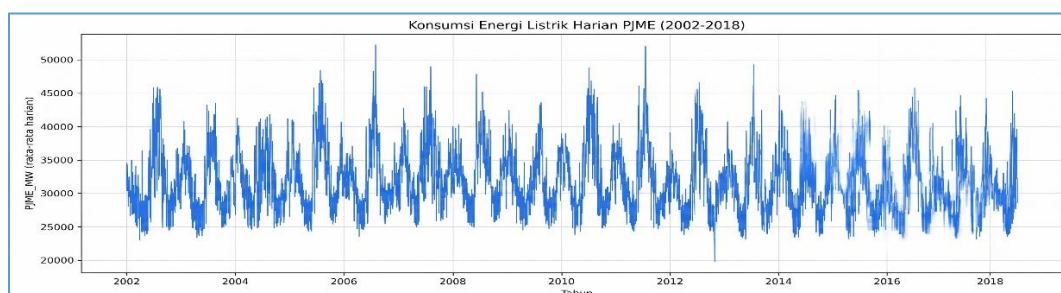
Penggunaan ketiga metrik secara bersamaan dipilih karena saling melengkapi: MAE dan RMSE dinyatakan dalam satuan yang sama dengan data asli sehingga memudahkan interpretasi besaran kesalahan, sedangkan MAPE dinyatakan dalam persentase yang tidak bergantung skala data sehingga memungkinkan perbandingan akurasi antar-skenario panjang data historis secara lebih objektif [13]. Sebagai acuan interpretasi, nilai MAPE di bawah 10% menandakan kemampuan model peramalan tergolong sangat baik [4], menjadi salah satu pertimbangan dalam menilai kelayakan hasil peramalan Holt-Winters dan LSTM pada setiap skenario. Selain ketiga metrik akurasi, penelitian ini turut mencatat waktu komputasi (*training time* dan *inference time*) sebagai metrik tambahan, guna mengukur secara empiris *trade-off* antara akurasi dan beban komputasi akibat penambahan panjang data historis [5], [7], [16]. Dengan demikian, titik keseimbangan optimal antara panjang data historis dan akurasi peramalan pada penelitian ini dinilai tidak hanya dari akurasi semata, melainkan juga dari efisiensi sumber daya komputasi yang diperlukan untuk mencapainya. Signifikansi statistik selisih akurasi diuji menggunakan dua pendekatan yang saling melengkapi: *one-sample t-test* (Holt-Winters vs LSTM per skenario, berdasarkan variabilitas antar-*seed*) dan uji Diebold-Mariano (perbandingan *loss* harian sepanjang horizon uji), dengan tingkat signifikansi $\alpha=0,05$.

4 Hasil dan Pembahasan

Bagian ini menyajikan hasil eksperimen peramalan konsumsi energi listrik menggunakan metode Holt-Winters dan LSTM pada kelima skenario panjang data historis, mencakup evaluasi kinerja kedua metode berdasarkan metrik akurasi dan waktu komputasi, analisis perbandingan guna mengidentifikasi titik keseimbangan optimal antara panjang data, akurasi, dan beban komputasi, serta interpretasi temuan tersebut menjadi pengetahuan (*knowledge discovery*) yang dapat dimanfaatkan penyedia layanan energi dalam merancang strategi pengelolaan data historis.

4.1 Deskripsi Pra Proses Data

Data pasca-praproses yang telah dijelaskan pada tahap Pra Proses Data menghasilkan deret waktu konsumsi energi listrik harian yang kontinu dan bebas dari nilai hilang, mencakup 6.059 observasi sepanjang 1 Januari 2002 - 3 Agustus 2018.



Gambar 3 Konsumsi energi harian 2002-2018

Gambar 3 memvisualisasikan trajektori harian tersebut, yang menunjukkan pola musiman tahunan yang persisten sepanjang periode pengamatan 16 tahun, dengan formasi puncak dan lembah

<http://sistemasi.ftik.unisi.ac.id>

konsumsi yang berulang secara konsisten pada tiap periode tahunan mengindikasikan tingkat volatilitas beban listrik regional yang cukup tinggi namun tetap mengikuti pola siklus yang dapat diprediksi.

Tabel 2 Statistik deskriptif pasca-praproses

Statistik	Nilai (MW)
Jumlah data (n)	6.059
Rata-rata (Mean)	32.079,46
Standar Deviasi	4.687,03
Minimum	19.769,04
Kuartil 1 (25%)	28.627,85
Median (50%)	31.356,92
Kuartil 3 (75%)	34.972,52
Maksimum	52.229,83
Skewness	0,68
Kurtosis	0,23

Tabel 2 menyajikan statistik deskriptif dari data pasca-praproses ini. Rata-rata beban harian tercatat sebesar 32.079,46 MW dengan standar deviasi 4.687,03 MW, sementara rentang observasi terbentang dari 19.769,04 MW hingga 52.229,83 MW. Nilai *skewness* sebesar 0,68 mengindikasikan distribusi yang menjulur ke kanan (*right-skewed*), menandakan adanya kecenderungan lonjakan konsumsi ekstrem pada hari-hari tertentu; sementara nilai *kurtosis* sebesar 0,23 yang mendekati nol menunjukkan distribusi berkarakter *mesokurtic*, mendekati bentuk normal tanpa ekor probabilitas yang tebal.

4.2 Evaluasi Kinerja Metode Holt Winters

Evaluasi Holt-Winters pada bagian ini mencakup estimasi parameter otomatis pada kedua varian di tiap skenario, pemilihan varian terbaik berdasarkan data validasi, dan pelaporan akurasi akhir dari data uji [6]. Tabel 3 merangkum hasil estimasi parameter dan waktu komputasi kedua varian di tiap skenario.

Tabel 3 Parameter, waktu komputasi, dan akurasi holt-winters

Skenario	Varian	α	β	γ	Fit (s)	Inf (s)
S1	Additive	1,0	0,0	0,0	0,1043	0,0056
S1	Multiplicative	1,0	0,0	0,0	0,1316	0,0086
S2	Additive	1,0	0,0	0,0	0,2062	0,0117
S2	Multiplicative	1,0	0,0	0,0	0,2121	0,0119
S3	Additive	1,0	0,0	0,0	0,3113	0,0189
S3	Multiplicative	1,0	0,0	0,0	0,3004	0,0184
S4	Additive	1,0	0,0	0,0	0,5531	0,0488
S4	Multiplicative	1,0	0,0	0,0	0,4898	0,0462
S5	Additive	1,0	0,0	0,0	0,5172	0,0310
S5	Multiplicative	1,0	0,0	0,0	0,4899	0,0314

Temuan utama Tabel 3: mekanisme optimasi secara konsisten mengestimasi $\alpha=1$, $\beta=0$, $\gamma=0$ pada seluruh sepuluh kombinasi skenario-varian tanpa pengecualian, mengindikasikan solusi batas (*boundary solution*), bukan titik optimum di bagian dalam ruang parameter [6]. Ini berbeda signifikan dari deskripsi umum Holt-Winters yang memperbarui ketiga komponen secara rekursif [3]: dengan $\beta=0$ dan $\gamma=0$, komponen tren dan musiman dibekukan pada nilai inisialisasi awal dan tidak pernah disesuaikan lagi, sementara $\alpha=1$ membuat *level* sepenuhnya mengikuti nilai aktual terakhir tanpa pemulusan historis. Akurasi peramalan secara efektif ditentukan oleh seberapa representatif estimasi awal komponen tren dan musiman dari keseluruhan data latih tiap skenario, bukan oleh adaptasi rekursif parameter. Ini juga menjelaskan mengapa kekhawatiran teoretis mengenai rasio observasi-parameter yang ketat pada S1 ($m=365$) tidak serta-merta membuat S1 menjadi skenario terburuk: persoalan utamanya bukan pembaruan parameter, melainkan kecukupan siklus musiman untuk membentuk estimasi awal yang stabil.

Dari sisi waktu komputasi, *fit time* (0,10–0,55 detik) dan *inference time* (0,006–0,049 detik) berada pada orde sangat kecil dan meningkat relatif teratur seiring panjang data; S4 *Additive* mencatat *fit time* tertinggi (0,5531 detik), sedikit melampaui S5 *Additive* (0,5172 detik) meski data latih S5 lebih panjang satu-satunya penyimpangan kecil dari tren tersebut. Pola ini mengonfirmasi estimasi Holt-Winters tetap komputasional ringan (di bawah satu detik pada seluruh skenario) terlepas dari volume data, berbeda dari metode *deep learning* yang lazimnya sensitif terhadap ukuran data pelatihan [5], [7]. Perbandingan langsung dibahas pada subbagian Perbandingan Akurasi Holt-Winters vs LSTM. Hasil pemilihan varian terbaik beserta akurasinya pada tiap skenario disajikan pada Tabel 4.

Tabel 4 Pemilihan varian dan akurasi holt-winters

Skenario	Varian	Terpilih	Val MAPE (%)	Test MAE	Test RMSE	Test MAPE (%)
S1	Additive	×	10,455	4.191,5	5.793,7	13,41
S1	Multiplicative	✓	10,22	4.418,6	6.171,96	14,09
S2	Additive	×	22,28	7.018,2	8.107,77	22,71
S2	Multiplicative	✓	20,26	7.054,68	8.333,07	22,58
S3	Additive	✓	12,09	4.173,57	5.174,69	12,25
S3	Multiplicative	×	12,47	4.380,29	5.438,76	12,71
S4	Additive	×	25,27	8.090,72	8.836,21	26,12
S4	Multiplicative	✓	22,77	7.967,12	8.845,13	25,47
S5	Additive	×	9,349	3.061,92	4.064,75	9,44
S5	Multiplicative	✓	9,33	3.043,80	4.044,44	9,37

Pemilihan varian berdasarkan MAPE validasi konsisten dengan hasil data uji pada empat dari lima skenario (S2–S5). Pengecualian terjadi pada S1: *Multiplicative* terpilih berdasarkan validasi (MAPE 10,22% vs 10,46% pada *Additive*), namun pada data uji *Additive* justru lebih rendah (13,41% vs 14,09%). Selisih validasi yang sangat tipis (0,24 poin) mengindikasikan kedua varian nyaris setara, sehingga pembalikan ini lebih mencerminkan variabilitas sampel wajar dibandingkan kegagalan prosedur seleksi justru menjadi bukti proses pemilihan model bebas dari kebocoran data uji [6].

Berdasarkan model terpilih, MAPE data uji menunjukkan pola tidak monoton: S1 14,09%, S2 22,58%, S3 12,25%, S4 25,47%, dan S5 9,37%. S2 dan S4 justru mencatat MAPE tertinggi, lebih tinggi dari S1 yang data latihnya jauh lebih pendek memperkuat argumen bahwa penambahan panjang data historis tidak selalu berbanding lurus dengan akurasi [7]. Pola ini sejalan secara konseptual dengan Royal *et al.* [10], yang juga melaporkan *training set* lebih panjang (3 dan 5 tahun) tidak mengungguli *training set* 1 tahun; namun tahun spesifik yang bermasalah berbeda (S2/S4 di sini vs 3/5 tahun pada penelitian Royal *et al.* [10]), kemungkinan mencerminkan perbedaan domain dan skala data. S5 mencatat performa terbaik (MAPE 9,37%), satu-satunya skenario di bawah ambang 10% yang menandakan model sangat baik [4]. S3 menempati posisi kedua (12,25%), sementara S1, meski secara teoretis paling rentan karena keterbatasan satu siklus musiman, justru mengungguli S2 dan S4 (14,09%). Dibandingkan MAPE Holt-Winters *Additive* sebesar 47,69% pada data curah hujan terbatas [4], seluruh skenario di sini jauh lebih rendah, kemungkinan karena karakteristik data konsumsi listrik yang lebih stabil dan berulang dibanding data curah hujan yang inheren lebih bergejolak. Dari sisi MW, rentang *Test MAE* model terpilih (3.043,80–7.967,12 MW) sebanding dengan rentang galat maksimum Holt-Winters (~4.000 MW) dan Prophet (<2.500 MW) pada beban puncak Houston yang dilaporkan Waheed *et al.* [18], meski perbandingan ini bersifat indikatif mengingat perbedaan metrik dan karakteristik data.

4.3 Evaluasi Kinerja Metode Long Short Term Memory (LSTM)

Evaluasi LSTM (Tabel 5) mengikuti arsitektur dan *hyperparameter* yang dikunci dari *tuning* pada subset 180 hari pertama data, diterapkan identik pada S1–S5, dengan lima pengulangan *seed* (42, 123, 456, 789, 999) per skenario untuk mengakomodasi variabilitas inisialisasi bobot. Konfigurasi final: *window* 30 hari, *epoch* 50, *batch size* 32, *learning rate* 0,001, arsitektur 1 *layer* LSTM (64 unit, aktivasi *tanh*) + 1 *dense output layer* struktur yang relatif ringkas dibanding arsitektur

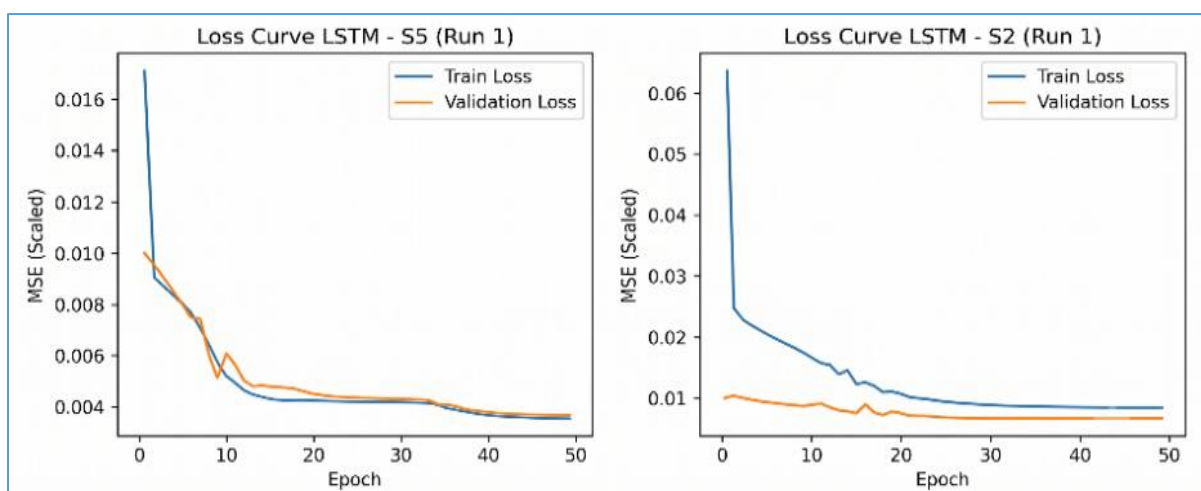
<http://sistemasi.ftik.unisi.ac.id>

multilayer/hybrid pada penelitian sejenis, seperti *multilayer dilated LSTM* dengan *attention* [15] atau kombinasi EMD-SSA-SAM-LSTM [2], konsisten dengan fokus penelitian ini pada pengaruh panjang data, bukan optimasi arsitektur.

Tabel 5 Waktu komputasi dan akurasi LSTM (Mean $\pm$ SD, 5 Run)

Skenario	Fit (s)	Inf (s)	MAE (MW)	RMSE (MW)	MAPE (%)
S1	11,27 $\pm$ 2,59	7,6837 $\pm$ 1,0621	4.671,03 $\pm$ 723,02	6.066,45 $\pm$ 861,11	13,02 $\pm$ 1,96
S2	18,67 $\pm$ 4,15	8,1938 $\pm$ 1,8750	3.195,67 $\pm$ 77,08	4.052,33 $\pm$ 245,53	9,69 $\pm$ 0,31
S3	24,39 $\pm$ 0,82	7,6965 $\pm$ 0,9776	6.822,77 $\pm$ 2.404,59	8.454,97 $\pm$ 2.610,09	19,08 $\pm$ 6,90
S4	30,67 $\pm$ 0,72	7,8282 $\pm$ 1,0388	6.397,49 $\pm$ 2.094,07	8.056,76 $\pm$ 2.410,08	17,84 $\pm$ 5,86
S5	38,14 $\pm$ 0,93	7,4855 $\pm$ 0,6056	6.930,46 $\pm$ 1.310,96	8.629,44 $\pm$ 1.361,46	19,31 $\pm$ 3,83

Berbeda dari Holt-Winters yang *fit time*-nya tidak meningkat tegas seiring penambahan data (sebagaimana dibahas pada evaluasi Holt-Winters), LSTM menunjukkan pola monoton yang jelas: *fit time* naik dari 11,27 detik (S1) menjadi 38,14 detik (S5), mengonfirmasi sensitivitas khas *deep learning* terhadap ukuran data [5], [7]. Sementara itu, *inference time* relatif stabil (7,49–8,19 detik) karena proses *recursive* tetap membutuhkan 90 iterasi *forward pass* yang identik terlepas dari panjang data latih. Standar deviasi (SD) *fit time* pada S1 dan S2 (masing-masing $\pm$ 2,59 dan $\pm$ 4,15 detik, atau sekitar 23% dan 22% dari rata-rata) jauh lebih besar secara proporsional dibandingkan S3–S5 ($<$ 3,5%), kemungkinan akibat kondisi *system warm-up* di awal siklus pengujian, bukan karena karakteristik data S1 maupun S2.



Gambar 4 Kurva Loss S2 dan S5

Karena *epoch* tetap (50) digunakan sebagai pengganti *early stopping*, kurva *training/validation loss* dianalisis untuk mendeteksi *under/overfitting*. (Gambar 4, panel loss S2 dan S5). Pada S2 (akurasi terbaik), *training loss* menurun tajam dari $\sim$ 0,043 ke $<$ 0,010 pada *epoch* ke-20, konvergen $\sim$ 0,0085; *validation loss* mengikuti pola serupa tanpa divergensi *epoch* tetap tidak menyebabkan *overfitting* pada S2. Pada S5 (akurasi terburuk), *training* dan *validation loss* justru berimpitan mulus sepanjang pelatihan, konvergen lebih rendah dari S2 ($\sim$ 0,004) tampak sebagai konvergensi ideal tanpa *gap*. Namun *loss* rendah ini tidak berbanding lurus dengan performa akhir: S5 justru mencatat MAPE data uji tertinggi (19,31%), karena *loss* pelatihan *single-step-ahead* dan MAPE evaluasi *recursive multi-step* mengukur horizon berbeda *loss* mencerminkan galat satu langkah dengan *input* aktual, MAPE mencerminkan galat kumulatif 90 langkah yang saling bergantung pada prediksi sebelumnya [6]. Pola serupa teramati kualitatif pada S1, S3, S4.

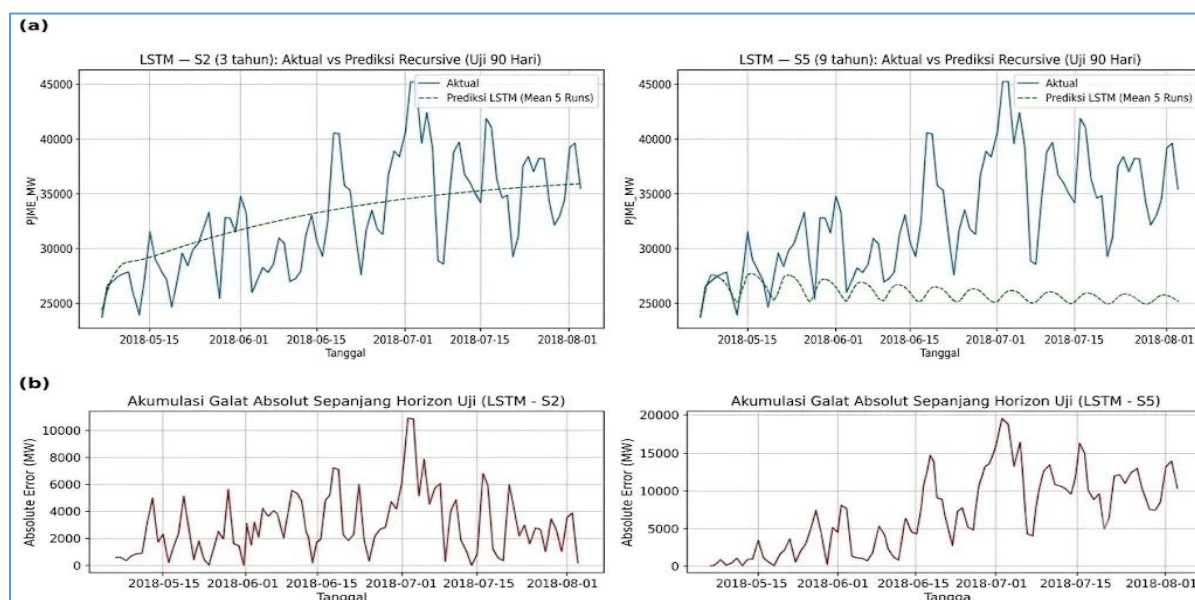
Urutan performa: S2 (9,69%) terbaik, diikuti S1 (13,02%), S4 (17,84%), S3 (19,08%), dan S5 (19,31%) terburuk. Pola tidak monoton namun dengan karakter berbeda dari Holt-Winters: performa LSTM cenderung memburuk dan berfluktuasi pada tingkat galat yang tinggi setelah S2, sedangkan Holt-Winters lebih bergejolak (S2/S4 terburuk, S5 terbaik) perbedaan arah inilah yang mendasari temuan *crossover* pada subbagian Perbandingan Akurasi Holt-Winters vs LSTM. Hanya S2 yang MAPE-nya di bawah 10% [4]; SD MAPE terkecil pada S2 ($\pm$ 0,31), memuncak pada S3 ($\pm$ 6,90),

dengan S1, S4, S5 masing-masing $\pm 1,96$; $\pm 5,86$; $\pm 3,83$. Variabilitas antar-run tidak meningkat seiring panjang data latih, melainkan memuncak di S3 sebelum menurun kembali. Dibandingkan MAPE LSTM murni 4,39% [2] atau 1,5% pada peramalan *hourly* [18], seluruh skenario di sini jauh lebih tinggi sebagian besar dijelaskan oleh horizon evaluasi kedua studi pembandingan yang jauh lebih pendek tanpa skema *recursive* 90 hari. Horizon uji 90 hari dipecah menjadi tiga segmen 30 hari untuk mengamati akumulasi galat secara empiris.

Tabel 6 Breakdown MAPE per segmen horizon uji

Skenario	Hari 1–30 (%)	Hari 31–60 (%)	Hari 61–90 (%)
S1	7,521	13,684	17,855
S2	7,739	11,867	9,449
S3	8,763	20,827	27,660
S4	7,562	19,094	26,849
S5	7,976	21,161	28,805

Empat skenario (S1, S3, S4, S5) pada Tabel 6 menunjukkan pola MAPE meningkat konsisten seiring horizon, mengindikasikan akumulasi galat khas skema *recursive* [6]. S2 menjadi pengecualian: MAPE hari 61–90 (9,45%) justru lebih rendah dari hari 31–60 (11,87%), menegaskan posisinya sebagai skenario paling resilien. Gambar 4 memperkuat temuan inikurva prediksi S2 mengikuti tren naik yang melandai mendekati pita fluktuasi aktual meski tak menangkap fluktuasi harian tajam, sementara kurva S5 nyaris datar dan gagal merespons lonjakan tajam data aktual (puncak >45.000 MW awal Juli 2018), manifestasi regresi-ke-rata-rata akibat setiap langkah prediksi *recursive* menggunakan hasil prediksi sebelumnya sebagai input.



Gambar 5 Aktual vs prediksi dan akumulasi galat absolut LSTM (S2 dan S5)

Kurva akumulasi galat absolut (Gambar 5, panel S2 dan S5) menunjukkan kontras serupa: pada S2, galat berfluktuasi terkendali (mayoritas <6.000 MW) dengan satu lonjakan ~10.500 MW bertepatan puncak aktual awal Juli, lalu kembali turun. Pada S5, galat absolut naik jelas seiring waktu, dari <2.000 MW di awal horizon hingga konsisten >10.000 MW dan puncak ~19.500 MW awal Juli sejalan dengan MAPE segmen akhir (28,81%) lebih dari tiga kali MAPE segmen awal (7,98%). Secara keseluruhan, karakteristik akumulasi galat LSTM *recursive* tidak seragam di seluruh skenario: skenario berperforma baik (S2) juga lebih resilien terhadap akumulasi galat, sedangkan skenario berperforma buruk (S5) menunjukkan baik MAPE tinggi maupun akumulasi galat tajam mengindikasikan akurasi rendah dan akumulasi galat tinggi cenderung muncul bersamaan, bukan dua persoalan independen.

4.4 Perbandingan Akurasi Holt Winters vs LSTM

Tabel 7 Perbandingan akurasi dan waktu komputasi holt-winters vs LSTM

Skenario	HW MAPE(%)	LSTM MAPE (%)	Selisih (poin)	Unggul	HW Waktu (s)	LSTM Waktu (s)	Tambahan Beban LSTM (s)
S1	14,09	13,02 ± 1,96	1,07	LSTM	0,1402	18,9537	+18,82
S2	22,58	9,69 ± 0,31	12,89	LSTM	0,2240	26,8638	+26,64
S3	12,25	19,08 ± 6,90	6,83	HW	0,3302	32,0865	+31,76
S4	25,47	17,84 ± 5,86	7,63	LSTM	0,5360	38,4982	+37,96
S5	9,37	19,31 ± 3,83	9,94	HW	0,5213	45,6255	+45,10

Berdasarkan Tabel 7 terlihat bahwa tidak ada satu metode yang konsisten unggul: LSTM lebih rendah MAPE-nya pada tiga skenario (S1, S2, S4), Holt-Winters pada dua (S3, S5). Rata-rata MAPE keseluruhan LSTM (15,79%) sedikit lebih rendah dari Holt-Winters (16,75%), namun rentang performa Holt-Winters jauh lebih lebar (9,37–25,47%) dibanding LSTM (9,69–19,31%) Holt-Winters lebih ekstrem antar-skenario, LSTM relatif lebih stabil meski tidak selalu unggul. Menggunakan SD LSTM sebagai proksi variabilitas *run-to-run*, selisih S1 dan S3 masih dalam rentang SD LSTM (tidak dapat dibedakan meyakinkan dari variabilitas inisialisasi bobot acak), sedangkan selisih S2, S4, dan S5 melampaui SD LSTM yang bersangkutan (perbedaan bermakna).

Temuan paling menonjol: pola *crossover* pada S5 skenario data latih terpanjang (9 tahun) adalah skenario terbaik Holt-Winters (MAPE 9,37%) namun sekaligus terburuk LSTM (MAPE 19,31%). Konsisten dengan mekanisme masing-masing metode (lihat evaluasi Holt-Winters dan LSTM sebelumnya): performa Holt-Winters ditentukan representativitas estimasi awal komponen musiman (*boundary solution*), sehingga data lebih panjang berarti lebih banyak siklus musiman untuk estimasi stabil; performa LSTM pada skema *recursive multi-step* justru memburuk setelah S2, kemungkinan akibat akumulasi galat yang lebih parah dan variabilitas antar-*run* yang meningkat tajam. Pola *alternating superiority* ini berbeda dari Butt *et al.* [19], yang menemukan LSTM konsisten unggul dibanding ARIMA maupun *attention-based transformer* tanpa pembalikan keunggulan seperti di sini, mengindikasikan kesimpulan "metode X lebih unggul" dari satu konfigurasi data tunggal berisiko tidak tergeneralisasi. Temuan Holt-Winters mengungguli LSTM justru pada data latih terpanjang juga sejalan dengan Royal *et al.* [10], yang menemukan LSTM menghasilkan RMSE jauh lebih tinggi dibanding model statistik klasik pada peramalan jangka panjang. Dari sisi waktu komputasi, LSTM konsisten membutuhkan waktu total jauh lebih besar, dengan tambahan beban meningkat teratur dari +18,82 detik (S1) hingga +45,10 detik (S5) sejalan hubungan linear antara durasi pelatihan dan ukuran data [6], sekaligus bukti empiris bahwa metode *deep learning* menanggung kerumitan komputasi yang terus bertambah seiring volume data, terlepas manfaat akurasi yang sepadan [8], [12].

Tabel 8 Hasil uji signifikansi statistik (one-sample T-Test)

Skenario	t	p-value	Signifikan ($\alpha=0,05$)
S1	-1,221	0,289	Tidak
S2	-92,977	<0,001	Ya
S3	2,213	0,091	Tidak
S4	-2,911	0,044	Ya
S5	5,803	0,004	Ya

Untuk memformalkan penalaran berbasis SD LSTM di atas, dilakukan uji *one-sample t-test* yang menguji apakah rata-rata selisih MAPE Holt-Winters dan LSTM lintas lima pengulangan *seed* berbeda signifikan dari nol pada tiap skenario (Tabel 8). Hasil uji menunjukkan bahwa selisih pada S2 ($t=-92,977$, $p<0,001$), S4 ($t=-2,911$, $p=0,044$), dan S5 ($t=5,803$, $p=0,004$) signifikan secara statistik, sedangkan selisih pada S1 ($t=-1,221$, $p=0,289$) dan S3 ($t=2,213$, $p=0,091$) tidak signifikan pada $\alpha=0,05$. Temuan ini memperkuat secara formal pengamatan sebelumnya bahwa selisih S1 dan S3 masih berada dalam rentang variabilitas *run-to-run* LSTM sehingga tidak dapat dibedakan meyakinkan dari kebetulan inisialisasi bobot acak, sementara selisih pada S2, S4, dan S5 cukup besar dan konsisten untuk disimpulkan sebagai perbedaan performa yang bermakna antara kedua metode.

<http://sistemasi.ftik.unisi.ac.id>

Tabel 9 Hasil uji diebold-mariano (HW vs LSTM per skenario)

Skenario	Varian HW	DM Statistic	p-value	Model Unggul	Signifikan ($\alpha=0,05$)
S1	Multiplicative	0,795	0,427	LSTM (tipis)	Tidak
S2	Multiplicative	6,480	<0,001	LSTM	Ya
S3	Additive	-5,864	<0,001	Holt-Winters	Ya
S4	Multiplicative	2,118	0,034	LSTM	Ya
S5	Multiplicative	-6,895	<0,001	Holt-Winters	Ya

Untuk memperkuat validitas temuan *crossover performance* dengan mempertimbangkan struktur temporal galat harian secara eksplisit, dilakukan pula uji Diebold-Mariano (DM) yang membandingkan *loss* kuadrat harian (90 hari) antara Holt-Winters (varian terbaik) dan LSTM (rata-rata lima *run*) pada tiap skenario (Tabel 9). Hasil uji menunjukkan pola yang konsisten dengan Tabel 7: LSTM unggul signifikan pada S2 (DM=6,48, $p<0,001$) dan S4 (DM=2,12, $p=0,034$), sedangkan Holt-Winters unggul signifikan pada S3 (DM=-5,86, $p<0,001$) dan S5 (DM=-6,90, $p<0,001$), sementara selisih pada S1 tidak signifikan (DM=0,80, $p=0,427$), sejalan dengan hasil *one-sample t-test* pada Tabel 8. Menariknya, hasil uji DM pada S3 berbeda dari *one-sample t-test* yang sebelumnya tidak signifikan ($p=0,091$); perbedaan ini beralasan karena kedua uji menguji hipotesis berbeda — *one-sample t-test* menguji konsistensi selisih rata-rata MAPE lintas lima pengulangan *seed* (variabilitas antar-*run*), sedangkan uji DM menguji signifikansi selisih *loss* harian dalam satu realisasi peramalan (variabilitas intra-horizon) yang cenderung lebih kecil, sehingga uji DM memiliki daya statistik lebih tinggi untuk mendeteksi perbedaan performa yang konsisten sepanjang horizon uji.

4.5 Titik Keseimbangan Optimal dan Posisi Temuan terhadap Penelitian Terdahulu

Tabel 10 Titik keseimbangan optimal holt-winters dan LSTM

Metode	Skenario Optimal	Panjang Data Latih	MAPE (%)	Total Waktu Komputasi (s)
Holt-Winters	S5	9 tahun	9,37	0,5213
LSTM	S2	3 tahun	9,69	26,8638

Penentuan titik optimal bersandar pada durasi data historis yang menghasilkan MAPE data uji paling rendah [7]. Ditemukan pola unik: Holt-Winters optimal pada data latih terpanjang (S5), LSTM optimal pada data latih tersingkat kedua (S2) kendati bertolak belakang dari sisi volume data, keduanya sama-sama menekan MAPE di bawah 10%, ambang klasifikasi sangat akurat [4]. Hasil ini mengurai hipotesis awal penelitian: kebutuhan panjang data historis bersifat kasuistik, bergantung pada cara kerja intrinsik masing-masing algoritma. Titik optimal Holt-Winters konsisten dengan mekanisme *boundary solution* (lihat evaluasi Holt-Winters): semakin panjang data latih, semakin banyak siklus musiman penuh untuk estimasi awal komponen musiman yang representatif. Titik optimal LSTM di S2 tidak dapat dijelaskan mekanisme serupa karena LSTM tidak bergantung estimasi musiman eksplisit; berdasarkan evaluasi LSTM di atas, performa LSTM lebih dipengaruhi resiliensi terhadap akumulasi galat *recursive* S2 paling resilien, sementara S3/S4 menunjukkan variabilitas antar-*run* tajam sekaligus akumulasi galat lebih parah.

Pola ini berbeda karakter dari Simani *et al.* [6], yang melaporkan peningkatan akurasi menjadi marjinal (bukan memburuk nyata) setelah sekitar 90 hari data latih kemungkinan karena studi tersebut mengevaluasi horizon pendek dengan koreksi berkala nilai aktual, sedangkan penelitian ini menerapkan skema *recursive* 90 hari tanpa koreksi, sehingga penambahan data pelatihan di sini aktif merugikan akurasi, bukan sekadar *diminishing returns*. Dari sisi komputasi, total waktu Holt-Winters pada titik optimalnya (S5) hanya 0,5213 detik, sedangkan LSTM pada titik optimalnya (S2) mencapai 26,8638 detik lebih dari 51 kali lipat, memperkuat argumen Zhang *et al.* [8] bahwa kerumitan komputasi *deep learning* bersifat inheren terhadap arsitekturnya. Pola *crossover* serupa juga dilaporkan Royal *et al.* [10] dan Lazcano *et al.* [20], menegaskan titik optimal yang berbeda tajam antar-metode bukan anomali data PJME, melainkan cerminan sensitivitas spesifik-model dan spesifik-domain.

Implikasi praktis: jika prioritas akurasi maksimal dengan beban komputasi minimal, Holt-Winters dengan data latih sembilan tahun paling efisien (MAPE 9,37%, waktu <1 detik); jika data

<http://sistemasi.ftik.unisi.ac.id>

historis terbatas, LSTM dengan tiga tahun data tetap kompetitif (MAPE 9,69%) meski beban komputasinya jauh lebih besar. Penyedia layanan energi perlu menghindari asumsi generik bahwa penambahan data historis akan selalu memperbaiki akurasi tanpa mempertimbangkan sensitivitas masing-masing metode.

Secara keseluruhan, temuan pada Tabel 10 memperkuat posisi penelitian ini dalam mengisi keempat kesenjangan yang telah diuraikan pada Tinjauan Literatur: sensitivitas Holt-Winters terhadap panjang data yang belum diuji Royal *et al.* [10]; skala regional/multi-wilayah (PJME, $\pm 16,7$ tahun) yang belum diuji dibanding skala bangunan tunggal [6], pasar harga tahunan [9], atau kampus [10]; desain pemisahan data latih-uji yang mengisolasi murni efek panjang data latih; serta arsitektur LSTM tunggal yang sengaja dipertahankan sederhana. Dengan demikian, titik keseimbangan optimal pada Tabel 10 bukan sekadar konfirmasi ulang temuan terdahulu, melainkan perluasan pemahaman pada domain, metode, dan desain eksperimen yang belum pernah diuji secara bersamaan.

4.6 Interpretasi kedalam Knowledge Discovery

Di luar posisinya terhadap penelitian terdahulu, temuan-temuan pada bagian ini juga dapat dimaknai sebagai manifestasi empiris dari prinsip-prinsip statistik yang berlaku lintas-bidang. *Boundary solution* pada Holt-Winters (lihat evaluasi Holt-Winters) sejalan dengan konsep *effective sample size*: pada data berautokorelasi musiman tinggi, performa lebih ditentukan oleh kecukupan siklus musiman penuh dibandingkan volume data semata. Memburuknya akurasi LSTM seiring data latih (lihat evaluasi LSTM dan perbandingan kedua metode) selaras dengan *trade-off* bias-varians dan *error propagation* pada sistem dinamis, di mana penambahan data pada skema *recursive multi-step* tidak menekan varians akibat akumulasi galat yang membesar secara sistematis [7]. Selain menguatkan prinsip statistik lintas-bidang tersebut, temuan ini juga relevan dimaknai sebagai penerapan *knowledge discovery in databases* (KDD) dalam disiplin sistem informasi Xie *et al.* [1] memosisikan peramalan energi listrik sebagai salah satu penerapan utama KDD, dan akurasi peramalan pada penelitian ini merepresentasikan kualitas pengetahuan dari keseluruhan alur KDD, mulai dari seleksi data, pra-pemrosesan, transformasi skenario panjang data, penambangan pola melalui Holt-Winters dan LSTM, hingga interpretasi hasil. Manfaat konkret dari kualitas pengetahuan tersebut dapat dijabarkan pada tiga tataran berikut.

Pada tataran kualitas data input bagi *decision support system* (DSS), akurasi peramalan yang rendah akan menghasilkan estimasi kebutuhan pembangkitan dan distribusi yang menyimpang dari kondisi aktual, sehingga keputusan alokasi sumber daya yang diambil berbasis output tersebut ikut bias sebaliknya, akurasi tinggi (MAPE <10% pada titik optimal kedua metode dalam penelitian ini) menjamin DSS bekerja di atas basis pengetahuan yang valid. Lebih jauh, pada tataran integrasi sistem operasional, hasil peramalan yang akurat memungkinkan model diintegrasikan langsung sebagai modul prediktif pada sistem informasi kelistrikan yang sudah berjalan, seperti *dashboard* eksekutif untuk pemantauan dan proyeksi beban, tanpa memerlukan lapisan validasi tambahan yang membebani proses bisnis. Tidak kalah penting, pada tataran efisiensi tata kelola data (*data governance*), temuan bahwa Holt-Winters mencapai akurasi optimal dengan beban komputasi minimal (0,5213 detik) menunjukkan bahwa sistem informasi peramalan yang andal tidak selalu menuntut infrastruktur komputasi mahal, sehingga organisasi dengan sumber daya terbatas tetap dapat membangun kapabilitas peramalan berbasis pengetahuan yang kredibel. Dengan demikian, kontribusi penelitian ini terhadap proses *knowledge discovery* beroperasi pada dua lapis sekaligus yaitu memperkuat prinsip statistik dan teori estimasi yang berlaku lintas-bidang, sekaligus menghasilkan pengetahuan yang aplikatif secara langsung bagi pengambilan keputusan dalam ekosistem sistem informasi kelistrikan.

Pendekatan KDD serupa telah diterapkan pada peramalan harga beras [21], prediksi *churn* pelanggan *e-commerce* [22], dan analisis persepsi publik terhadap AI generatif [23], di mana keluaran model diterjemahkan menjadi pengetahuan yang mendukung keputusan. Bahri dan Tania [21] menyimpulkan bahwa pemilihan model prediksi yang tepat bergantung pada karakteristik data historis dan stabilitas harga, sejalan dengan temuan penelitian ini bahwa tidak ada satu konfigurasi data dan metode yang selalu terbaik. Temuan yang bersifat kontekstual juga muncul pada studi perbandingan metode statistik dan *deep learning* di luar kerangka KDD: Fadhilahrizka *et al.* [24] menyimpulkan, dalam kerangka *No Free Lunch Theorem*, bahwa tidak ada satu model peramalan yang optimal untuk seluruh jenis data, sedangkan Miranda *et al.* [25] melaporkan bahwa metode statistik klasik mengungguli LSTM pada peramalan harga bawang merah harian ketika pola data tidak terlalu

<http://sistemasi.ftik.unisi.ac.id>

kompleks. Kesesuaian lintas domain ini memperkuat bahwa pengetahuan hasil KDD tidak berhenti pada model dengan akurasi tertinggi, tetapi mencakup pemahaman tentang kondisi data dan karakteristik metode yang membuat suatu model layak digunakan.

5 Kesimpulan

Penelitian ini menemukan bahwa pengaruh panjang data historis terhadap hasil ramalan konsumsi energi listrik bersifat situasional dan sangat bergantung pada ciri khas masing-masing metode. Holt-Winters mencapai titik terbaik pada skenario S5 dengan sembilan tahun data latih. Ini menghasilkan MAPE data uji sebesar 9,37%. Sementara itu, LSTM mencapai titik terbaiknya pada skenario S2 dengan tiga tahun data latih. Ini menghasilkan MAPE sebesar 9,69%. Keduanya termasuk punya hasil prediksi yang sangat baik. Titik terbaik Holt-Winters sejalan dengan cara estimasi awal komponen musiman yang makin baik ketika jumlah siklus musiman penuh dalam data latih bertambah. Sebaliknya, titik terbaik LSTM cocok dengan tingkat ketahanan terhadap penumpukan galat pada skema peramalan recursive multi-step. Hal ini malah memburuk jika data latihnya lebih panjang. Dari sisi beban komputasi, Holt-Winters pada titik terbaiknya hanya butuh waktu komputasi total 0,5213 detik. LSTM pada titik terbaiknya butuh 26,8638 detik, atau lebih dari 51 kali lebih lama walaupun dilatih dengan data yang jauh lebih singkat. Pola performa crossover ini didukung oleh dua uji signifikansi statistik formal yang saling melengkapi. Uji one-sample t-test menunjukkan bahwa selisih MAPE Holt-Winters dan LSTM signifikan secara statistik pada S2, S4, dan S5. Pada S1 dan S3, selisihnya tidak signifikan sehingga tidak bisa dibedakan dengan jelas dari variasi run-to-run LSTM. Uji Diebold-Mariano, yang menguji signifikansi selisih loss harian sepanjang horizon uji, memberikan pola serupa. LSTM lebih unggul secara signifikan pada S2 dan S4. Holt-Winters lebih unggul secara signifikan pada S3 dan S5. Sementara pada S1 tidak ada perbedaan signifikan. Uji ini juga memastikan bahwa perbedaan performa pada keempat skenario tadi bukan kebetulan, melainkan bermakna secara statistik. Temuan ini juga membuktikan bahwa pola titik optimal yang berbeda tajam antar-metode bukanlah anomali data, tetapi mencerminkan sensitivitas yang khas untuk setiap model dan domain. Pada akhirnya, tidak ada jawaban tunggal tentang berapa lama data historis yang ideal untuk meramal konsumsi energi listrik. Kebutuhan ini sepenuhnya tergantung pada infrastruktur komputasi dan metode peramalan yang dipakai. Kesimpulan ini sesuai dengan hasil penelitian dan memperlihatkan bahwa kebutuhan panjang data historis bergantung pada karakteristik metode. Penelitian ini mempunyai sejumlah keterbatasan. Pertama, studi ini hanya memakai satu kumpulan data tunggal (PJME) yang bersifat univariat, tanpa variabel luar seperti cuaca, penanda ekonomi, atau elemen lingkungan lain, jadi penemuan ini paling sesuai untuk kondisi serupa dan belum tentu berlaku pada area yang sangat dipengaruhi oleh variabel tersebut. Kedua, studi ini hanya menguji satu arsitektur LSTM tunggal yang sengaja dibuat sederhana. Berdasarkan keterbatasan itu, studi lanjutan disarankan untuk: (1) menguji keterulangan penemuan pada kumpulan data konsumsi listrik regional lain di luar PJME, sebaiknya dengan variabel luar seperti cuaca atau data ekonomi; dan (2) menjajaki berbagai arsitektur LSTM (multilayer, attention) sebagai perbandingan terhadap arsitektur tunggal yang dipakai di sini, untuk menguji apakah pola persilangan yang ditemukan tetap sama pada kerumitan model yang berbeda.

Referensi

- [1] B. Xie, C. Zhu, L. Zhao, and J. Zhang, "A gradient boosting machine-based framework for electricity energy knowledge discovery," *Front. Environ. Sci.*, vol. 10, no. November, pp. 1–16, 2022, doi: 10.3389/fenvs.2022.1031095.
- [2] Y. Zhou, Z. Su, K. Gao, Z. Wang, W. Ye, and J. Zeng, "A short-term electricity load forecasting method integrating empirical modal decomposition with SAM-LSTM," *Front. Energy Res.*, vol. 12, no. June, pp. 1–11, 2024, doi: 10.3389/fenrg.2024.1423692.
- [3] J. M. Ahmad Rofii, Muhammad Sobirin, "Prediksi Konsumsi Listrik Bangunan Gedung Menggunakan Model Lstm Untuk Efisiensi Energi Dan Mitigasi Lingkungan," 2024.
- [4] N. M. W. Citrayanti, I. Wiguna, and ..., "Comparison of Holt-Winter Exponential Smoothing and Lstm Methods on Rainfall Forecasting in Subak Abian Wanasari Kenjung," ... , *Des. E ...*, vol. 3, no. December, pp. 230–237, 2024, [Online]. Available: <https://eprosiding.idbbali.ac.id/index.php/imade/article/view/879%0Ahttps://eprosiding.idbbali>
<http://sistemasi.ftik.unisi.ac.id>

- .ac.id/index.php/imade/article/download/879/577
- [5] F. Firdaus, M. Komaro, and V. Dwiyantri, "Perbandingan Metode Holt-Winters dan SARIMA untuk Memprediksi Permintaan Produk Fashion pada PT XYZ," *Fact. J. Ind. Manaj. dan Rekayasa Sist. Ind.*, vol. 4, no. 2, pp. 287–308, 2026, doi: 10.56211/factory.v4i2.1401.
 - [6] K. N. Simani, Y. O. Genga, and Y.-C. J. Yen, "Using LSTM to Perform Load Predictions for Grid-Interactive Buildings," *SAIEE Africa Res. J.*, vol. 115, no. 2, pp. 42–47, 2024, doi: 10.23919/saiee.2024.10520212.
 - [7] R. J. H. and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. Melbourne. 2021. [Online]. Available: <https://otexts.com/fpp3/>
 - [8] A. J. S. Zack C. Lipton, Zhang Aston, Mu Li, "Dive into Deep Learning," 2023, [Online]. Available: <https://d2l.ai/>
 - [9] V. Michalakopoulos, C. Menos-Aikateriniadis, E. Sarvas, A. Zakyntinos, P. S. Georgilakis, and D. Askounis, "Data-driven Day Ahead Market Prices Forecasting: A Focus on Short Training Set Windows," vol. xxx, pp. 1–13, 2025, [Online]. Available: <http://arxiv.org/abs/2506.10536>
 - [10] E. Royal, S. Bandyopadhyay, A. Newman, Q. Huang, and P. C. Tabares-Velasco, "A statistical framework for district energy long-term electric load forecasting," *Appl. Energy*, vol. 384, 2025, doi: 10.1016/j.apenergy.2025.125445.
 - [11] M. I. C. Rachmatullah, "Forecasting Data Time Series Menggunakan MLP dan LSTM untuk Memprediksi Jumlah Produksi Bir," *TIN Terap. Inform. Nusantara*, vol. 6, no. 4, pp. 409–416, 2025, doi: 10.47065/tin.v6i4.7755.
 - [12] L. Moosbrugger *et al.*, "Load Forecasting for Households and Energy Communities: Are Deep Learning Models Worth the Effort?," 2025, [Online]. Available: <http://arxiv.org/abs/2501.05000>
 - [13] F. Bayram, P. Aupke, B. S. Ahmed, A. Kassler, A. Theocharis, and J. Forsman, "DA-LSTM: A dynamic drift-adaptive learning framework for interval load forecasting with LSTM networks," *Eng. Appl. Artif. Intell.*, vol. 123, no. January, p. 106480, 2023, doi: 10.1016/j.engappai.2023.106480.
 - [14] H. Li, S. Li, Y. Wu, Y. Xiao, Z. Pan, and M. Liu, "Short-term power load forecasting for integrated energy system based on a residual and attentive LSTM-TCN hybrid network," *Front. Energy Res.*, vol. 12, no. May, pp. 1–21, 2024, doi: 10.3389/fenrg.2024.1384142.
 - [15] J. Xiao, S. Li, Y. Tian, J. Huang, X. Jiang, and S. Wang, "Example dependent cost sensitive learning based selective deep ensemble model for customer credit scoring," pp. 1–24, 2025.
 - [16] R. Mulla, "Hourly energy consumption [Data set]." [Online]. Available: <https://www.kaggle.com/datasets/robikscube/hourly-energy-consumption>
 - [17] F. Chollet and M. Watson, *Deep Learning With Python*. 2021. [Online]. Available: <https://deeplearningwithpython.io/>
 - [18] W. Waheed, Q. Xu, M. Aurangzeb, S. Iqbal, S. H. Dar, and Z. M. S. Elbarbary, "Empowering data-driven load forecasting by leveraging long short-term memory recurrent neural networks," *Heliyon*, vol. 10, no. 24, p. e40934, 2024, doi: 10.1016/j.heliyon.2024.e40934.
 - [19] T. F. Butt, S. Tameer, M. Saleem, J. S. Ur Rehman, and S. K. Selvaperumal, "A comparative study on electricity load forecasting using statistical and deep learning approaches," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 38, no. 3, p. 1540, 2025, doi: 10.11591/ijeecs.v38.i3.pp1540-1552.
 - [20] A. Lazcano, P. Hidalgo, and J. E. Sandubete, "Walking Back the Data Quantity Assumption to Improve Time Series Prediction in Deep Learning," *Appl. Sci.*, vol. 14, no. 23, 2024, doi: 10.3390/app142311081.
 - [21] C. A. Bahri and K. D. Tania, "Perbandingan Kinerja LSTM , Random Forest , dan SVR Berbasis Knowledge Discovery untuk Prediksi Harga Beras Sumatera Selatan," vol. 12, no. 5, pp. 721–732, 2025, doi: 10.30865/jurikom.v12i5.9140.
 - [22] D. A. Ardhani and K. D. Tania, "Knowledge Discovery on E-Commerce Customer Churn Using Interpretable Machine Learning : A Comparative Study of SHAP-Based Classifiers," vol. 9, no. 5, pp. 2695–2702, 2025.
 - [23] M. R. Maulizidan and K. D. Tania, "(Indonesian Journal of Informatics Education) Knowledge Discovery Based on Sentiment Analysis of Public Perceptions About Generative
<http://sistemasi.ftik.unisi.ac.id>

- AI on X,” vol. 9, no. 2, 2025.
- [24] D. N. Fadhilahrizka, K. D. Tania, and R. D. Kurnia, “Analisis performa holt-winters dan sarima dalam peramalan multivariabel iklim bulanan di wilayah pesisir kota semarang 1),” vol. 11, no. 1, pp. 1878–1893, 2026.
- [25] F. A. Miranda, K. D. Tania, and R. D. Kurnia, “Comparative Performance Evaluation of ARIMA , SARIMA , and LSTM for Daily Shallot Price Forecasting in Palembang City,” vol. 6, no. 2, pp. 1–12, 2025.