

Perbandingan LSTM dan XGBoost untuk Prediksi Multi-Horizon Suhu dan Curah Hujan Harian pada Grid ERA5 Jakarta Pusat

Ahmad Julaibib Syahr¹, Raden Teddy Iswahyudi², Nizirwan Anwar³, Imam Sutanto⁴

^{1,2,3,4} Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Esa Unggul

Email: ahmadjulaibib@student.esaunggul.ac.id, raden.teddy@esaunggul.ac.id, nizirwan.anwar@esaunggul.ac.id, imam.sutanto@esaunggul.ac.id

Abstrak

Prediksi deret meteorologis perkotaan tropis menghadapi dua karakter target yang berbeda: suhu relatif kontinu dan berautokorelasi, sedangkan curah hujan harian bersifat zero-inflated, menceng, dan mengandung kejadian ekstrem. Penelitian ini membandingkan Long Short-Term Memory (LSTM) dan Extreme Gradient Boosting (XGBoost) untuk direct multi-horizon forecasting $t+1$, $t+7$, dan $t+14$ pada satu grid reanalysis ERA5 yang dipilih menggunakan koordinat acuan Jakarta Pusat. Data harian 1 Januari 2005-21 Juli 2026 diakses melalui Open-Meteo, kemudian dibagi secara kronologis menjadi train, validation, main test 2024-2025, dan out-of-time (OOT) test Januari-Juli 2026. Model dibangun terpisah untuk setiap target-horizon, dioptimalkan dengan random search dan early stopping, serta dibandingkan dengan baseline kausal menggunakan RMSE, MAE, R^2 , dan bias. Pada main test, XGBoost memperoleh RMSE terendah pada enam tugas; selisih terhadap LSTM adalah 0,34%-5,48% untuk suhu dan 1,72%-3,18% untuk hujan. Semua model suhu melampaui baseline, tetapi model hujan hanya melampauinya pada $t+1$; klimatologi bulanan tetap lebih baik pada $t+7$ dan $t+14$. Peringkat XGBoost konsisten pada OOT, tetapi periode OOT yang lebih pendek dan berbeda komposisi musim tidak dapat dibandingkan langsung dengan main test. Hasil mendukung XGBoost sebagai model yang lebih stabil dan efisien dalam ruang eksperimen ini, bukan sebagai bukti superioritas universal. Prediksi intensitas hujan ekstrem tetap menjadi keterbatasan utama.

Kata kunci: ERA5; LSTM; multi-horizon forecasting; suhu rata-rata; curah hujan; XGBoost

Abstract

Forecasting tropical urban meteorological series involves targets with different statistical properties: temperature is relatively continuous and autocorrelated, whereas daily rainfall is zero-inflated, skewed, and contains extremes. This study compares Long Short-Term Memory (LSTM) and Extreme Gradient Boosting (XGBoost) for direct multi-horizon forecasts at $t+1$, $t+7$, and $t+14$ on a single ERA5 reanalysis grid selected using a Central Jakarta reference coordinate. Daily data from 1 January 2005 to 21 July 2026 were accessed through Open-Meteo and chronologically divided into training, validation, a 2024-2025 main test, and a January-July 2026 out-of-time (OOT) test. Separate target-horizon models were optimized using random search and early stopping, then evaluated against causal baselines with RMSE, MAE, R^2 , and bias. On the main test, XGBoost achieved the lowest RMSE in all six tasks; its advantage over LSTM ranged from 0.34%-5.48% for temperature and 1.72%-3.18% for rainfall. All temperature models outperformed their baselines, whereas rainfall models did so only at $t+1$; monthly climatology remained stronger at $t+7$ and $t+14$. XGBoost retained its ranking on the OOT period, although the shorter and seasonally different OOT window is not directly comparable with the main test. The findings support XGBoost as the more stable and computationally efficient option within this experimental space, not as evidence of universal superiority. Accurate estimation of extreme rainfall intensity remains the principal limitation.

Keywords: daily rainfall; ERA5; LSTM; mean temperature; multi-horizon forecasting; XGBoost.

1. PENDAHULUAN

Informasi suhu dan curah hujan harian dibutuhkan untuk pengelolaan air, transportasi, energi, kesehatan, dan mitigasi risiko hidrometeorologi. Kebutuhan tersebut menonjol di Jakarta karena kepadatan penduduk, perubahan tutupan lahan, serta interaksi iklim regional dan proses perkotaan. Analisis historis menunjukkan adanya perubahan pola suhu dan hujan Jakarta, tetapi interpretasi lokal tetap memerlukan kehati-hatian karena heterogenitas spasial wilayah metropolitan [16].

Deret yang panjang dan konsisten diperlukan untuk eksperimen peramalan berbasis data. ERA5 menggabungkan observasi dan model atmosfer melalui asimilasi data untuk menghasilkan reanalysis global per jam [11]. Namun, reanalysis merupakan estimasi berbasis grid, bukan pengukuran stasiun; presipitasi ERA5 juga memiliki ketidakpastian yang lebih besar pada proses konvektif dan wilayah tropis [14]. Studi domestik di DAS Anai menunjukkan bahwa evaluasi dan koreksi

terhadap stasiun dapat menurunkan MAE curah hujan ERA5 [1]. Karena itu, unit analisis penelitian ini adalah nilai harian pada satu grid ERA5 yang dipilih dari koordinat acuan Jakarta Pusat. Framing ini membatasi inferensi: hasil tidak secara otomatis mewakili seluruh Jakarta atau akurasi terhadap observasi lapangan.

Machine learning telah digunakan untuk post-processing, downscaling, dan peramalan cuaca berbasis data, tetapi performanya dipengaruhi representasi fitur, volume data, horizon, dan rancangan evaluasi [7], [27]. LSTM mempelajari dependensi urutan melalui mekanisme memori [13], sedangkan XGBoost membangun ensemble pohon terregularisasi pada representasi tabular [4]. Keduanya relevan untuk dibandingkan karena LSTM mengonstruksi representasi temporal secara internal, sedangkan XGBoost bergantung pada fitur lag, rolling, dan kalender yang eksplisit.

Literatur domestik telah membandingkan FFT, ARIMA, dan ANN untuk prediksi hujan [21], menguji neural-fuzzy pada data hujan harian [20], serta

menerapkan LSTM pada prediksi hujan harian dan mingguan [17]. Studi yang lebih baru di Tangerang Selatan membandingkan LSTM dan CNN-LSTM pada deret hujan 2005-2025 [23]. Rangkaian studi tersebut menunjukkan perkembangan model berbasis data di Indonesia, tetapi belum menjawab perbandingan LSTM-XGBoost pada dua target meteorologis, tiga horizon langsung, baseline kausal, dan OOT yang sama.

Celah penelitian yang dituju adalah perbandingan terkendali kedua algoritma pada target dengan distribusi berbeda, beberapa horizon langsung, baseline kausal, serta evaluasi temporal terpisah. Kontribusi penelitian meliputi: (1) pipeline anti-kebocoran berbasis forecast origin dan target time; (2) perbandingan LSTM-XGBoost pada enam tugas target-horizon; (3) evaluasi terhadap baseline sederhana, main test, dan OOT; serta (4) analisis khusus kejadian hujan untuk memisahkan deteksi kejadian dari estimasi intensitas. Kajian multi-horizon penting karena model terbaik dan sumber

kesalahan dapat berubah menurut jarak prediksi [5], [15].

2. METODE PENELITIAN

2.1 Data, Unit Analisis, dan Pembagian Waktu

Data diakses dari Historical Weather API Open-Meteo dengan parameter model ERA5 untuk koordinat permintaan -6,1818; 106,8223 [18]. Respons API menunjuk pusat grid -6,25; 106,75 pada elevasi 15 m. Periode 1 Januari 2005-21 Juli 2026 menghasilkan 7.872 observasi harian. Target adalah temperature_2m_mean (°C) dan rain_sum (mm); prediktor tambahan terdiri atas temperature_2m_max, temperature_2m_min, shortwave_radiation_sum, dan wind_speed_10m_max. Nilai historis kedua target juga digunakan sebagai prediktor kausal.

Tabel 1. Rancangan Data dan Pembagian Kronologis

Komponen	Periode/variabel	Fungsi
Train	1 Jan 2005-31 Des 2021	Parameter, model, dan scaler
Validation	1 Jan 2022-31 Des 2023	Tuning dan early stopping
Main test	1 Jan 2024-31 Des 2025	Evaluasi utama
OOT test	1 Jan-21 Jul 2026	Uji konsistensi temporal
Target	Suhu rata-rata; rain_sum	Regresi terpisah
Prediktor	Suhu min/maks, radiasi, angin, histori target	Informasi tersedia hingga waktu t

Sumber: rancangan penelitian berdasarkan deret ERA5 yang diakses melalui Open-Meteo.

Split ditentukan menggunakan target_time, bukan hanya tanggal baris fitur. Konteks historis dari periode sebelumnya boleh digunakan jika telah tersedia pada forecast origin t, tetapi nilai setelah t tidak boleh masuk ke fitur. Pemeriksaan integritas meliputi frekuensi harian, urutan waktu, duplikasi, missing value, nilai tak hingga, dan rentang fisik. Tidak ditemukan tanggal hilang, duplikasi, atau nilai kosong, sehingga imputasi tidak dilakukan dan nilai ekstrem tidak dihapus otomatis.

2.2 Pembentukan Fitur dan Strategi Prediksi

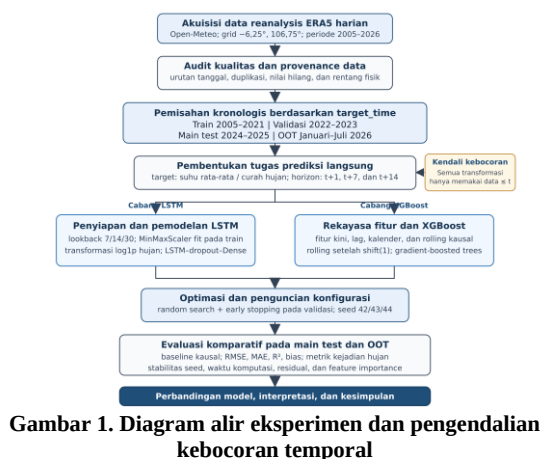
LSTM menerima tensor [sampel, lookback, fitur] dengan kandidat lookback 7, 14, dan 30 hari. Variabel input dan target dinormalisasi menggunakan MinMaxScaler yang dipelajari hanya dari train saat tuning. Setelah konfigurasi dikunci, scaler final dipelajari ulang pada train+validation. Untuk hujan, target ditransformasi log1p; prediksi dikembalikan dengan expm1 dan dibatasi minimum nol. XGBoost menerima fitur current yang tersedia pada t, lag 1, 2, 3, 7, 14, dan 29 hari, rolling 3, 7, 14, dan 30 hari, kalender siklik, rentang suhu, indikator hari hujan, serta musim basah. Semua rolling dihitung setelah shift(1), sehingga tidak menggunakan observasi target pada atau setelah forecast origin.

Strategi direct forecasting membentuk model terpisah untuk setiap target y dan horizon $h \in \{1, 7, 14\}$:

$\hat{y}(t+h) = f_h(X \leq t)$. Dengan demikian, t+7 dan t+14 tidak menggunakan prediksi t+1 dan terhindar dari propagasi galat rekursif. Target hujan pada setiap horizon adalah intensitas tepat pada hari t+h, bukan akumulasi selama tujuh atau empat belas hari.

2.3 Model, Optimasi, Baseline, dan Evaluasi

LSTM menggunakan satu lapisan LSTM, dropout, dan Dense(1), dengan loss MSE, optimizer Adam, dan shuffle=False. Tidak ada satu konfigurasi LSTM yang selalu optimal di seluruh masalah, sehingga unit, dropout, learning rate, batch size, lookback, dan kelompok fitur dituning [10]. XGBoost menggunakan objective reg:squarederror, tree_method hist, dan evaluasi RMSE; fondasinya adalah gradient boosting bertahap [9].



Tabel 2. Ruang Optimasi dan Anggaran Eksperimen

Komponen	LSTM	XGBoost
Kelompok fitur	basic, calendar, derived	basic, lag, lag_rolling, full
Lookback	7, 14, 30 hari	7, 14, 30 hari
Kapasitas	unit 32/64/128; dropout 0,1/0,2/0,3	depth 2-6; learning rate 0,01-0,10
Batas	150 epoch; patience 15	2.000 round; patience 50
Trial	10 per tugas (60 total)	20 per tugas (120 total)
Seed final	42, 43, 44	42, 43, 44

Sumber: ruang random search dan early stopping penelitian.

Random search digunakan karena hanya sebagian hiperparameter biasanya berpengaruh dominan pada ruang pencarian berdimensi tinggi [2], [25]. Konfigurasi dipilih hanya dari validation. Model final kemudian dilatih pada train+validation dengan best epoch atau best iteration yang telah dikunci. Seed 42, 43, dan 44 digunakan untuk menilai stabilitas, bukan untuk memilih ulang model dari main test.

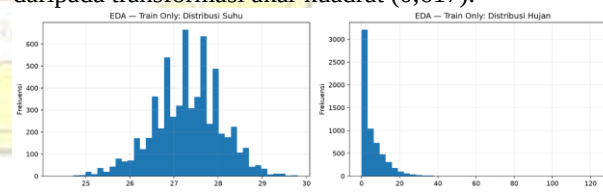
Baseline suhu mencakup persistence, moving average, klimatologi bulanan, dan seasonal naive; baseline hujan mencakup persistence, moving average, klimatologi/median bulanan, dan nol. Moving average juga telah digunakan sebagai representasi hujan antecedent pada studi DAS Ciliwung Hulu, sehingga relevan sebagai benchmark lokal yang transparan [28]. Baseline terbaik dipilih berdasarkan RMSE validation dan dibekukan untuk kedua periode uji. Metrik utama adalah RMSE, dilengkapi MAE, R^2 , dan bias. Penggunaan RMSE dan MAE secara bersama membantu membedakan penalti terhadap galat besar dan besar galat absolut tipikal [3], [24]. Evaluasi luar sampel diperlukan untuk menghindari penilaian pada data pemilihan model [22].

Untuk hujan, observasi ≥ 1 mm dikategorikan hari hujan dan ≥ 20 mm sebagai hujan berintensitas tinggi. Precision, recall, F1, false-wet rate, dan MAE pada kejadian ≥ 20 mm digunakan sebagai diagnostik tambahan. Karena naskah tidak menyertakan deret galat berpasangan dan uji Diebold-Mariano, selisih RMSE dilaporkan sebagai keunggulan numerik, bukan perbedaan yang signifikan secara statistik [8].

3. HASIL

3.1 Karakteristik Data dan Hasil Optimasi

Jumlah sampel train untuk t+1, t+7, dan t+14 masing-masing 6.208, 6.202, dan 6.195; validation dan main test masing-masing memiliki 730 dan 731 target, sedangkan OOT memiliki 202 target. Pasangan forecast_origin_time, target_time, dan nilai aktual telah diperiksa identik antaralgoritma. Distribusi suhu train relatif simetris (skewness -0,161), sedangkan hujan sangat menceng ke kanan (3,355). Transformasi log1p menurunkan skewness hujan menjadi 0,163, lebih kecil daripada transformasi akar kuadrat (0,617).



Gambar 2. Distribusi suhu rata-rata dan curah hujan pada data train

Sumber: pengolahan data train ERA5 periode 2005-2021.

Persistence menjadi baseline suhu t+1 terbaik, sedangkan moving average 14 hari terpilih pada t+7 dan t+14. Untuk hujan, moving average 14 hari terpilih pada t+1 dan klimatologi bulanan pada t+7 serta t+14. Hasil ini menunjukkan bahwa autokorelasi jangka pendek kuat pada suhu, sedangkan informasi musiman menjadi benchmark kompetitif bagi hujan pada horizon lebih panjang.

3.2 Perbandingan Main Test dan Out-of-Time

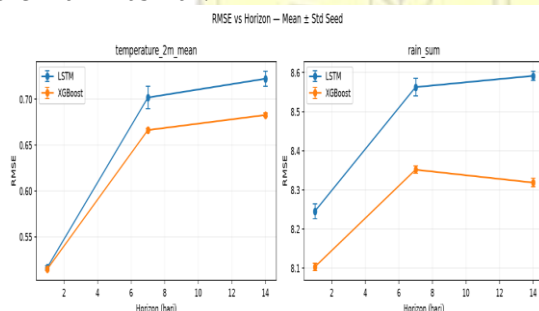
Tabel 3. RMSE Main Test 2024-2025 dan OOT 2026

Target	H	Baseline main	LSTM main	XGB main	XGB vs LSTM	XGB OOT
Suhu	t+1	0,5671	0,5163	0,5145	0,34%	0,4276
Suhu	t+7	0,7135	0,7018	0,6663	5,06%	0,6497
Suhu	t+14	0,7723	0,7223	0,6828	5,48%	0,6615
Hujan	t+1	8,4152	8,2454	8,1034	1,72%	5,9229
Hujan	t+7	8,1054	8,5628	8,3516	2,47%	6,2588
Hujan	t+14	8,1050	8,5916	8,3186	3,18%	6,3776

Sumber: evaluasi tiga seed model final; satuan RMSE suhu = °C dan hujan = mm.

XGBoost memperoleh rata-rata RMSE terendah pada enam tugas main test. Pada suhu t+1, selisih 0,0018 °C (0,34%) sebanding dengan variasi seed dan tidak mendukung klaim keunggulan substantif tanpa uji berpasangan. Perbedaan lebih besar muncul pada t+7 dan t+14, masing-masing 5,06% dan 5,48%. R² XGBoost pada dua horizon tersebut adalah 0,2458 dan 0,2081, dibandingkan 0,1631 dan 0,1136 untuk LSTM.

Pada hujan, XGBoost mengungguli LSTM secara numerik 1,72%-3,18%, tetapi R² tetap rendah: 0,0705 (t+1), 0,0127 (t+7), dan 0,0204 (t+14); LSTM bahkan menghasilkan R² negatif pada t+7 dan t+14. Semua model suhu melampaui baseline, sedangkan kedua model hujan hanya melampaui baseline pada t+1. Dengan demikian, kemenangan relatif terhadap LSTM tidak sama dengan skill yang memadai terhadap benchmark musiman.



Gambar 3. RMSE suhu dan hujan menurut model dan horizon

Sumber: evaluasi tiga seed pada main test 2024-2025.

H	Model	F1	MAE ≥20 mm	Underprediction ≥20 mm	False-wet rate
t+1	LSTM	0,8762	25,52	100%	56,4%
t+1	XGBoost	0,8823	24,93	100%	50,7%
t+7	LSTM	0,8485	26,35	100%	75,6%
t+7	XGBoost	0,8470	25,37	100%	82,0%
t+14	LSTM	0,8505	26,40	100%	79,4%
t+14	XGBoost	0,8465	24,87	100%	81,0%

Sumber: analisis kejadian main test; MAE dinyatakan dalam mm.

F1 0,85-0,88 menunjukkan bahwa kedua model relatif mampu mengidentifikasi hari hujan pada ambang 1 mm. Namun, seluruh 37 observasi dengan hujan aktual ≥20 mm diprediksi lebih rendah. Residual hujan (prediksi-aktual) memiliki median dekat nol, tetapi mean negatif dan minimum sekitar -103 mm. False-wet rate juga meningkat pada horizon lebih panjang. Jadi, kemampuan mendeteksi

Optimasi memperbaiki 10 dari 12 kombinasi. Peningkatan terbesar terjadi pada LSTM suhu t+7 (10,38%), LSTM suhu t+14 (9,13%), XGBoost suhu t+14 (6,56%), dan XGBoost suhu t+7 (5,40%). LSTM suhu t+1 dan XGBoost hujan t+1 memburuk masing-masing 0,50% dan 0,45%, yang menegaskan bahwa performa validation tidak selalu berpindah ke periode uji. Pada OOT Januari-21 Juli 2026, XGBoost mempertahankan peringkat terendah pada enam tugas. Namun, OOT hanya memuat 202 hari dan tidak mencakup satu tahun penuh; RMSE OOT yang lebih rendah tidak dapat dianggap sebagai peningkatan antartahun. Relatif terhadap baseline OOT, XGBoost hujan hanya unggul 0,18% pada t+7 dan kalah 1,74% pada t+14.

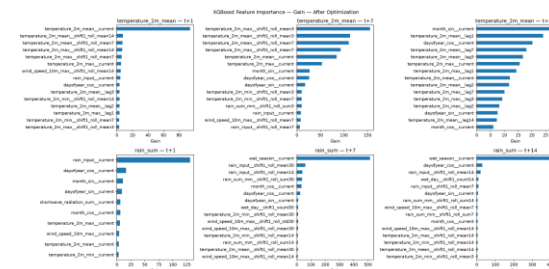
3.3 Deteksi Hujan, Residual, dan Interpretasi Fitur

Tabel 4. Diagnostik Kejadian Hujan pada Main Test

kejadian tidak boleh disamakan dengan kemampuan mengestimasi intensitas, terutama pada ekor distribusi yang jarang. Pola ini konsisten dengan sifat zero-inflated dan menceng pada hujan harian [19].

Pada suhu t+1, bias LSTM 0,0923 °C dan XGBoost -0,0794 °C. Pada t+7 dan t+14, keduanya cenderung underpredict; bias LSTM mencapai -0,2285 dan -0,2583 °C, sedangkan XGBoost -0,1441 dan -0,1759 °C. Analisis kuartil menunjukkan regression toward the

mean: suhu rendah diprediksi terlalu tinggi dan suhu tinggi terlalu rendah.



Gambar 4. Gain feature importance XGBoost menurut target dan horizon

Sumber: model XGBoost final setelah optimasi.

Pada suhu $t+1$, `temperature_2m_mean_current` dominan; pada horizon lebih panjang, rolling mean, lag, dan fitur kalender memperoleh kontribusi lebih besar. Untuk hujan, `rain_input_current` dominan pada $t+1$, sedangkan `wet_season_current` lebih dominan pada $t+7$ dan $t+14$. Gain importance menjelaskan penggunaan fitur oleh model, bukan hubungan sebab-akibat; korelasi antarfitur juga dapat membagi atau mengalihkan importance.

4. PEMBAHASAN

Keunggulan XGBoost pada penelitian ini masuk akal secara metodologis. Sekitar 6.200 sampel train per tugas relatif terbatas bagi deep learning, sedangkan XGBoost memperoleh representasi waktu yang telah disusun sebagai current, lag, rolling, kalender, dan musim. LSTM harus mempelajari representasi tersebut dari urutan dengan ruang konfigurasi yang lebih mahal. Enam puluh trial LSTM memerlukan sekitar 5,28 jam, sedangkan 120 trial XGBoost sekitar 6,16 menit pada lingkungan komputasi yang digunakan. Pelatihan final per seed berkisar 58-697 detik untuk LSTM dan 0,36-4,03 detik untuk XGBoost. Angka ini bukan benchmark perangkat keras universal, tetapi relevan untuk efisiensi pipeline penelitian.

Hasil hujan memperlihatkan batas regresi titik pada data zero-inflated. Loss MSE pada target $\log_{1p}$ menyeimbangkan rentang nilai dan memperbaiki stabilitas, tetapi kompresi skala dapat mendorong prediksi ke intensitas yang sering terjadi dan mengurangi sensitivitas terhadap ekstrem. Model dua tahap—klasifikasi kejadian diikuti regresi intensitas kondisional—atau objective yang memberi bobot lebih besar pada ekor dapat diuji pada penelitian lanjutan. Validasi terhadap stasiun BMKG juga diperlukan sebelum hasil digunakan untuk keputusan lokal.

Tiga pembatas utama memengaruhi generalisasi. Pertama, satu grid tidak menangkap heterogenitas spasial Jakarta; analisis domestik di Jawa Barat juga menunjukkan distribusi hujan yang berbeda antarlokasi dan keterkaitannya dengan angin serta topografi [26]. Kedua, anggaran random search LSTM dan XGBoost tidak setara dan konfigurasi terbaik LSTM suhu $t+7/t+14$ muncul pada trial akhir, sehingga pencarian LSTM mungkin belum jenuh. Ketiga, OOT 2026 hanya mencakup Januari-Juli dan belum ada uji signifikansi galat berpasangan. Oleh

sebab itu, rumusan yang tepat adalah bahwa XGBoost unggul secara numerik dan efisien dalam desain ini; superioritas algoritma secara universal belum dapat disimpulkan.

Kebutuhan terhadap baseline kuat juga didukung studi domestik pada deret muka laut Jakarta: ARIMA mengungguli LSTM berdasarkan MAE, MAPE, dan RMSE [12]. Temuan lintas-domain tersebut tidak dapat dipindahkan langsung ke suhu atau hujan, tetapi memperkuat prinsip bahwa kompleksitas model tidak menjamin performa dan setiap klaim harus dibandingkan dengan benchmark yang sesuai.

Untuk meningkatkan daya terbit dan replikasi, artikel perlu disertai repositori kode, parameter API, daftar fitur per kelompok, seed, versi pustaka, spesifikasi perangkat keras, serta file prediksi per target-horizon. Jika deret galat tersedia, uji Diebold-Mariano atau bootstrap blok berpasangan sebaiknya ditambahkan dengan koreksi multipel untuk enam tugas. Penilaian eksternal akan lebih kuat apabila mencakup beberapa grid atau stasiun, rolling-origin evaluation, dan perbandingan dengan model statistik yang dituning pada anggaran serupa.

5. KESIMPULAN

Penelitian membangun direct multi-horizon forecasting suhu rata-rata dan curah hujan harian pada satu grid ERA5 beracuan Jakarta Pusat untuk $t+1$, $t+7$, dan $t+14$ dengan split kronologis berbasis target_time. XGBoost menghasilkan RMSE main test terendah pada seluruh enam tugas dan mempertahankan peringkat pada OOT. Keunggulan numeriknya terhadap LSTM adalah 0,34%-5,48% untuk suhu dan 1,72%-3,18% untuk hujan. Semua model suhu melampaui baseline, tetapi model hujan hanya melampauinya pada $t+1$; klimatologi bulanan tetap lebih baik pada $t+7$ dan $t+14$. Kedua model mampu mendeteksi hari hujan secara relatif baik, tetapi meng-underpredict seluruh 37 kejadian ≥ 20 mm. Dengan mempertimbangkan stabilitas seed dan biaya komputasi, XGBoost merupakan pilihan yang lebih sesuai untuk pipeline ini. Kesimpulan dibatasi pada target grid ERA5, periode, fitur, dan anggaran optimasi yang digunakan.

DAFTAR PUSTAKA

- [1] Abdillah, M. R., Harisuseno, D., & Sajali, M. A. (2025). Pemanfaatan data ERA5 untuk estimasi curah hujan di DAS Anai Provinsi Sumatera Barat. *Jurnal Teknologi dan Rekayasa Sumber Daya Air*, 5(1), 628-638. <https://doi.org/10.21776/ub.jtresda.2025.005.01.059>
- [2] Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13, 281-305. <https://www.jmlr.org/papers/v13/bergstra12.a.html>
- [3] Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the

literature. *Geoscientific Model Development*, 7, 1247-1250. <https://doi.org/10.5194/gmd-7-1247-2014>

- [4] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). Association for Computing Machinery. <https://doi.org/10.1145/2939672.2939785>
- [5] Chevillon, G. (2007). Direct multi-step estimation and forecasting. *Journal of Economic Surveys*, 21(4), 746-785. <https://doi.org/10.1111/j.1467-6419.2007.00518.x>
- [6] Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, e623. <https://doi.org/10.7717/peerj-cs.623>
- [7] de Burgh-Day, C. O., & Leeuwenburg, T. (2023). Machine learning for numerical weather and climate modelling: A review. *Geoscientific Model Development*, 16, 6433-6477. <https://doi.org/10.5194/gmd-16-6433-2023>
- [8] Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253-263. <https://doi.org/10.1080/07350015.1995.10524599>
- [9] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189-1232. <https://doi.org/10.1214/aos/1013203451>
- [10] Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10), 2222-2232. <https://doi.org/10.1109/TNNLS.2016.2582924>
- [11] Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., ... Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999-2049. <https://doi.org/10.1002/qj.3803>
- [12] Hilal, Y. N., Nainggolan, G. D. A., Syahputri, S. H., & Kartiasih, F. (2024). Comparison of ARIMA and LSTM methods in predicting Jakarta sea level. *Jurnal Ilmu dan Teknologi Kelautan Tropis*, 16(2), 163-178. <https://doi.org/10.29244/jitkt.v16i2.52818>
- [13] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [14] Lavers, D. A., Simmons, A., Vamborg, F., & Rodwell, M. J. (2022). An evaluation of ERA5 precipitation for climate monitoring. *Quarterly Journal of the Royal Meteorological Society*, 148(748), 3152-3165. <https://doi.org/10.1002/qj.4351>
- [15] Lim, B., & Zohren, S. (2021). Time-series forecasting with deep learning: A survey. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194), 20200209. <https://doi.org/10.1098/rsta.2020.0209>
- [16] Maheng, D., Bhattacharya, B., Zevenbergen, C., & Pathirana, A. (2024). Changing urban temperature and rainfall patterns in Jakarta: A comprehensive historical analysis. *Sustainability*, 16(1), 350. <https://doi.org/10.3390/su16010350>
- [17] Monita, V., Raniprima, S., & Cahyadi, N. (2024). Comparative analysis of daily and weekly heavy rain prediction using LSTM and cloud data. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, 10(4), 833-842. <https://doi.org/10.26555/jiteki.v10i4.30374>
- [18] Open-Meteo. (n.d.). Historical Weather API. Retrieved July 26, 2026, from <https://open-meteo.com/en/docs/historical-weather-api>
- [19] Pakoksung, K., & Takagi, M. (2017). Mixed of zero-inflation method and probability distribution in fitting daily rainfall data. *Engineering Journal*, 21(2), 63-80. <https://doi.org/10.4186/ej.2017.21.2.63>
- [20] Ritha, N., & Wardoyo, R. (2016). Implementasi neural fuzzy inference system dan algoritma pelatihan Levenberg-Marquardt untuk prediksi curah hujan. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 10(2), 125-136. <https://doi.org/10.22146/ijccs.15532>
- [21] Susilokarti, D., Arif, S. S., Susanto, S., & Sutiarto, L. (2015). Studi komparasi prediksi curah hujan metode fast Fourier transformation (FFT), autoregressive integrated moving average (ARIMA), dan artificial neural network (ANN). *agriTECH*, 35(2), 241-247. <https://doi.org/10.22146/agritech.9412>
- [22] Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting*, 16(4), 437-450. [https://doi.org/10.1016/S0169-2070\(00\)00065-0](https://doi.org/10.1016/S0169-2070(00)00065-0)
- [23] Uliyatunisa, & Supriatna, D. (2025). Komparasi model LSTM dan CNN-LSTM untuk peramalan curah hujan di Kota Tangerang Selatan. *Bulletin of Information Technology*, 6(3), 294-301. <https://doi.org/10.47065/bit.v6i3.2235>

- [24] Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79-82. <https://doi.org/10.3354/cr030079>
- [25] Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415, 295-316. <https://doi.org/10.1016/j.neucom.2020.07.061>
- [26] Zahroh, N. F., Syahdiza, R., Prayoga, M. B. R., Mulyana, E., Wirahma, S., Purwadi, & Djamil, Y. S. (2024). Spatial rainfall distribution in the eastern part of West Java Province using principal component analysis and cluster analysis. *Jurnal Sains & Teknologi Modifikasi Cuaca*, 25(2), 51-62. <https://ejournal.brin.go.id/JSTMC/article/view/5738>
- [27] Zhang, H., Liu, Y., Zhang, C., & Li, N. (2025). Machine learning methods for weather forecasting: A survey. *Atmosphere*, 16(1), 82. <https://doi.org/10.3390/atmos16010082>
- [28] Zuma, D. S., Murtiaksono, K., & Suharnoto, Y. (2017). Analisis curah hujan dan debit model SWAT dengan metode moving average di DAS Ciliwung Hulu. *Jurnal Pengelolaan Sumberdaya Alam dan Lingkungan*, 7(2), 98-106. <https://doi.org/10.29244/jpsl.7.2.98-106>
- Zuma, D. S., Murtiaksono, K., & Suharnoto, Y. (2017). Analisis curah hujan dan debit model SWAT dengan metode moving average di DAS Ciliwung Hulu. *Jurnal Pengelolaan Sumberdaya Alam dan Lingkungan*, 7(2), 98-106. <https://doi.org/10.29244/jpsl.7.2.98-106>

