Sistemasi: Jurnal Sistem Informasi ISSN:2302-8149
Volume 15, Nomor 11, 2026: 4742-4757 e-ISSN:2540-9719

Perbandingan *XGBoost* dan *LightGBM* untuk Prediksi Tinggi Gelombang Signifikan 1 Jam

Comparison of XGBoost and LightGBM for 1-Hour-Ahead Significant Wave Height Prediction

[1]Aulia Resty Nur Aini*, [2]Rumini
[1,2]Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta
[1,2]Jl. Ring Road Utara, Condong Catur, Depok, Sleman, Yogyakarta, Indonesia 55283
*e-mail: auliaresty@students.amikom.ac.id, rumini@amikom.ac.id

(***received*:** 23 October 2026, ***revised*:** 9 November 2026, ***accepted*:** 11 November 2026)

Abstrak

Prediksi tinggi gelombang signifikan (Hs) jangka sangat pendek berguna bagi peringatan dini keselamatan pantai dan pelayaran. Penelitian ini membandingkan XGBoost dan LightGBM untuk memprediksi Hs 1 jam ke depan pada data *buoy* Gold Coast periode 2014–2024 dengan interval 30 menit. Evaluasi memakai skema *rolling-origin* dengan tiga tahun uji (2022, 2023, dan 2024); hiperparameter disetel ulang di setiap lipatan dengan Optuna (50 percobaan per model) pada ruang pencarian yang disetarakan antarmodel. Model memprediksi perubahan Hs, lalu hasilnya dikembalikan ke *level* Hs. Pada gabungan tiga tahun uji, LightGBM mencapai RMSE 0,0787 m dan R^2 perubahan Hs 0,131, sedangkan XGBoost 0,0790 m dan 0,123 (acuan *persistence*: 0,0844 m). Uji *block bootstrap* berpasangan menunjukkan selisih RMSE relatif XGBoost terhadap LightGBM sebesar −0,49% (selang kepercayaan 95%: −0,85 hingga −0,12; negatif berarti LightGBM lebih baik), dan hasil ini sama pada panjang blok 48 dan 144 langkah. Selisih tersebut tegas pada 2022, setara pada 2023, dan belum tegas pada 2024; pada ruang pencarian yang lebih sempit kedua model setara, sehingga selisihnya sensitif terhadap penyetelan. Kedua model hanya menjelaskan sekitar 12–13% variasi perubahan Hs satu jam. Dengan demikian, pada horizon 1 jam LightGBM sedikit lebih akurat daripada XGBoost, namun selisihnya kecil (sekitar 0,4 mm).

Kata kunci: LightGBM, prediksi jangka sangat pendek, *rolling-origin*, Tinggi gelombang signifikan, XGBoost

Abstract

Very short-term significant wave height (Hs) forecasting is important for early warnings related to coastal safety and maritime navigation. This study compares XGBoost and LightGBM for one-hour-ahead Hs prediction using Gold Coast buoy data collected from 2014 to 2024 at 30-minute intervals. Model evaluation employed a rolling-origin validation scheme with three test years (2022, 2023, and 2024). Hyperparameters were independently tuned for each fold using Optuna, with 50 trials per model and equivalent search spaces across both algorithms. The models predicted changes in Hs, which were subsequently converted back to Hs levels. Across the three test years combined, LightGBM achieved an RMSE of 0.0787 m and an R^2 of 0.131 for Hs changes, whereas XGBoost achieved an RMSE of 0.0790 m and an R^2 of 0.123. The persistence baseline yielded an RMSE of 0.0844 m. A paired block bootstrap test indicated that the relative RMSE difference of XGBoost compared with LightGBM was −0.49% (95% confidence interval: −0.85% to −0.12%), with negative values indicating better performance by LightGBM. This result remained consistent across block lengths of 48 and 144 time steps. The performance difference was statistically distinguishable in 2022, negligible in 2023, and inconclusive in 2024. Under a narrower hyperparameter search space, the two models performed similarly, suggesting that the observed performance difference was sensitive to hyperparameter tuning. Both models explained only approximately 12–13% of the variance in one-hour Hs changes. Overall, LightGBM was slightly more accurate than XGBoost for one-hour-ahead Hs prediction, although the difference was small, at approximately 0.4 mm.

Keywords: *LightGBM, rolling-origin, Significant wave height, very short-term prediction, XGBoost*

 ISSN:2302-8149
e-ISSN:2540-9719

1 Pendahuluan

Tinggi gelombang signifikan (*significant wave height*/Hs) adalah parameter utama untuk menggambarkan kondisi laut. Nilainya dipakai dalam keselamatan pelayaran, perencanaan bangunan pantai, mitigasi bencana pesisir, dan pemanfaatan energi gelombang [1], [2]. Peringatan dini dan keputusan operasional hanya bisa diambil tepat waktu bila prediksi Hs jangka pendeknya akurat.

Peramalan gelombang masih banyak bertumpu pada model numerik berbasis fisika, padahal keluarannya membawa bias dan harus dikoreksi [3]. Pembelajaran mesin berkembang sebagai alternatif atau pelengkap untuk memperbaiki akurasi estimasi dan *hindcast* gelombang [4]. Untuk deret waktu Hs, *deep learning* seperti GRU, CNN-GRU, dan LSTM banyak dipakai [5], tetapi arsitekturnya umumnya rumit dan menuntut komputasi serta penyetelan hiperparameter yang besar.

Gradient boosting berbasis pohon keputusan menawarkan alternatif dengan arsitektur dan penyetelan yang lebih sederhana. XGBoost [6] dan LightGBM [7] adalah dua algoritma yang banyak dipakai; keduanya berbeda dalam strategi pertumbuhan pohon dan cara menangani data, dan perbedaan itu bisa memengaruhi akurasi maupun efisiensi. XGBoost sudah diterapkan pada peramalan tinggi gelombang, misalnya dibandingkan dengan AdaBoost di Pangandaran [8] dan dengan LSTM di Danau Erie [9].

Rentang horizon yang dikaji dalam literatur sangat lebar, dari 3 hingga 24 jam [10] sampai 7 hingga 60 hari [8]. Pada horizon sangat pendek, nilai Hs saat ini sudah sangat informatif, sehingga perbaikan akibat model yang lebih canggih sulit diukur dan mudah terlihat lebih besar daripada kenyataannya. Karena itu, penelitian pada horizon 1 jam perlu menguji selisih kecil antarmodel secara cermat, dan pengujiannya perlu mencakup beberapa tahun yang belum dilihat model.

Masih ada dua masalah yang belum terjawab. Pertama, studi-studi di atas berbeda dalam lokasi, resolusi waktu, fitur masukan, dan skema evaluasi [8], [9], sehingga sulit memastikan algoritma mana yang lebih unggul. Perbandingan XGBoost dan LightGBM pada horizon 1 jam dengan fitur yang seluruhnya berasal dari pengamatan *buoy* sendiri, diuji pada beberapa tahun uji, belum ditemukan dalam referensi yang dikaji. Kedua, banyak studi mengujinya pada satu pemisahan latih-uji saja, padahal pada deret waktu satu tahun uji belum tentu mewakili tahun lain [8], [9].

Kontribusi penelitian ini terletak pada rancangan evaluasi yang terkontrol, bukan pada klaim akurasi baru: kedua algoritma diuji dengan data, fitur, jumlah percobaan penyetelan, dan ruang pencarian yang disetarakan, dalam skema *rolling-origin* dengan tiga tahun uji, dengan uji *block bootstrap* berpasangan, dan dilengkapi analisis sensitivitas terhadap ruang pencarian. Penelitian ini bertujuan membandingkan kinerja LightGBM dan XGBoost dalam memprediksi Hs 1 jam ke depan, memakai Gold Coast City Wave Dataset periode 2014–2024 dari Coastal Data System Pemerintah Queensland. Kinerja dinyatakan sebagai RMSE, selisih RMSE relatif antarmodel, R^2 dari perubahan Hs, serta MAE dan MAPE. Secara praktis, hasilnya diharapkan menjadi acuan memilih algoritma untuk prediksi gelombang jangka sangat pendek; secara akademis, penelitian ini menambah bukti empiris perbandingan XGBoost dan LightGBM dengan evaluasi pada beberapa tahun uji.

2 Tinjauan Literatur

Riset peramalan Hs cenderung mengejar akurasi dengan menambahkan komponen di atas model dasar. Titik awalnya model rekuren: GRU, CNN-GRU, dan LSTM [5]. Kemudian muncul dekomposisi sinyal seperti EEMD yang digabungkan dengan *deep learning* [11], sistem hibrida dua tahap [12], dan *autoencoder*–BiGRU yang dikombinasikan dengan *empirical wavelet transform* serta seleksi fitur Boruta-CatBoost [13]. Mekanisme atensi [14] dan hibrida CNN-BiLSTM untuk perairan Australia [2] mengikuti jalur yang sama. Peningkatan akurasi umumnya dicapai dengan menambah tahap praproses dan kompleksitas arsitektur, dengan konsekuensi kebutuhan komputasi serta penyetelan yang ikut meningkat; bagi sistem peringatan dini yang ringan, beban itu dapat menjadi kendala.

Model berbasis pohon keputusan dan *ensemble* memiliki peran yang beragam. Sebagai peramal utama, XGBoost dipakai di Pangandaran [8] dan Danau Erie [9], sementara *stacking ensemble* diusulkan untuk prediksi tinggi gelombang laut [15]. Pembelajaran mesin juga dipakai untuk memperbaiki estimasi tinggi gelombang [16], untuk prediksi Hs sekaligus analisis gelombang ekstrem [17], dan melalui *ensemble learning* dari pengamatan *buoy real-time* untuk memprediksi potensi energi gelombang di Laut Hitam [1]. Pada studi lain, *gradient boosting* dipakai untuk mengoreksi

prakiraan numerik [3] atau dibandingkan dengan jaringan saraf untuk memperbaiki prakiraan Hs [18]. Hal ini menunjukkan bahwa model berbasis pohon fleksibel dan kompetitif. Namun, setiap studi membawa data, lokasi, horizon, dan skema evaluasinya sendiri, sehingga hasilnya tidak dapat langsung dipakai untuk memilih algoritma pada prediksi jangka sangat pendek.

XGBoost [6] dan LightGBM [7] sama-sama membangun pohon keputusan secara berurutan, dengan tiap pohon memperbaiki galat pohon sebelumnya, sebagaimana pendekatan *gradient boosting* pada studi gelombang [3], [18]. Keduanya memakai metode histogram untuk mempercepat pelatihan. Bedanya, XGBoost secara bawaan menumbuhkan pohon per tingkat kedalaman, sedangkan LightGBM memilih daun dengan penurunan galat terbesar, sehingga pohonnya bisa lebih dalam dan tidak simetris. Pertumbuhan *leaf-wise* perlu dibatasi lewat parameter seperti jumlah daun. XGBoost sudah dipakai untuk prediksi tinggi gelombang [8], [9].

Masukan model juga beragam. Sebagian studi memasukkan variabel angin dan atmosfer: fitur angin yang berkorelasi spasial di Laut Jawa [19], parameter angin dan atmosfer yang digeneralisasi [20], yang juga membandingkan XGBoost dan LightGBM, dan prakiraan kecepatan angin bersama Hs [18]. Informasinya memang kaya, tetapi dalam pemakaian operasional variabel tersebut harus sudah tersedia, atau diprakirakan terlebih dahulu, pada saat prediksi dibuat. Untuk jangka sangat pendek, riwayat Hs dari pengamatan sendiri selalu tersedia, sehingga layak diuji sebagai masukan utama.

Studi-studi tersebut juga berbeda jauh dalam horizon dan skala data. Horizon prediksi berkisar dari 30 menit hingga 6 jam pada model hibrida CLSTM-BiGRU di perairan Queensland [21], 3 hingga 24 jam pada RNN-LSTM [10], dan 7 hingga 60 hari pada XGBoost di Pangandaran [8]. Perbedaan horizon dan skala data ini membuat angka kesalahan antarstudi tidak dapat dibandingkan langsung, sehingga tidak dapat dijadikan dasar untuk memilih algoritma pada horizon 1 jam.

Kinerja pada gelombang tinggi dan cara membagi data juga layak dicermati. Pada [9], XGBoost dan LSTM memperbaiki prediksi gelombang saat angin kencang dibandingkan model fisika WAVEWATCH III. Temuan ini menunjukkan bahwa model berbasis pohon layak dipertimbangkan untuk sistem peringatan dini yang ringan, sekaligus bahwa kinerja pada gelombang tinggi perlu dilaporkan secara terpisah, sebagaimana kebutuhan analisis gelombang ekstrem [17]. Pada deret waktu, pengacakan membuat model melihat data yang lebih baru daripada data uji sehingga hasilnya terlalu optimis; sebagian studi menghindarinya, misalnya [8] memakai enam bulan data sebagai data uji dan [9] menetapkan periode berurutan masing-masing untuk pelatihan, validasi, dan pengujian. Namun uji pada satu atau dua tahun dari satu stasiun belum tentu mewakili kondisi tahun lain.

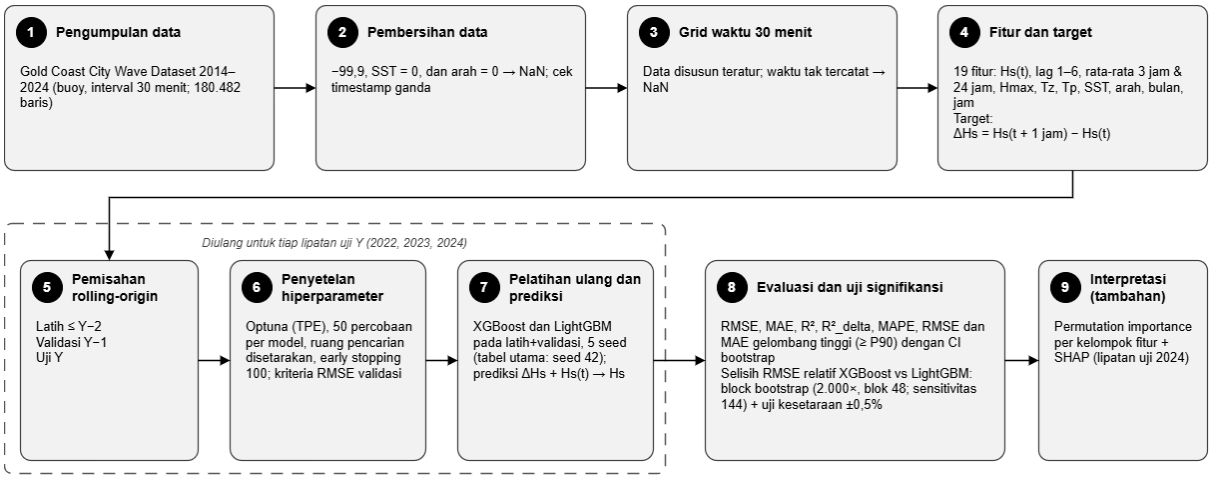
Dari sisi evaluasi, studi yang dikaji umumnya membandingkan model dengan model lain yang setara atau lebih kompleks: GRU, CNN-GRU, dan LSTM saling dibandingkan [5], XGBoost dibandingkan dengan AdaBoost [8] dan LSTM [9]. Perbedaan data, fitur, dan skema evaluasi antarstudi tersebut membuat hasilnya tidak dapat dipakai untuk membandingkan algoritma *boosting* secara terkontrol.

Dari tinjauan tersebut terdapat dua celah penelitian: (1) perbandingan XGBoost dan LightGBM pada horizon 1 jam dengan fitur yang seluruhnya berasal dari pengamatan *buoy* sendiri belum ditemukan dalam referensi yang dikaji; dan (2) prediksi jangka pendek pada data pengamatan berfrekuensi tinggi yang mencakup beberapa tahun uji, dengan fitur yang hanya berisi informasi yang tersedia saat prediksi, masih jarang dikaji. Penelitian ini mengisi keduanya dengan membandingkan LightGBM dan XGBoost untuk memprediksi Hs 1 jam ke depan pada Gold Coast City Wave Dataset periode 2014–2024. Perbandingan biaya komputasi langsung dengan model *deep learning* tidak dilakukan.

3 Metode Penelitian

Alur penelitian terdiri atas sembilan langkah (Gambar 1). Data *buoy* Gold Coast dikumpulkan (1) dan dibersihkan (2), lalu disusun pada *grid* waktu 30 menit (3). Dari *grid* itu dibentuk 19 fitur dan target perubahan Hs satu jam (4). Data dipisah dengan skema *rolling-origin* (5); untuk setiap tahun uji dilakukan penyetelan hiperparameter dengan Optuna (6) serta pelatihan ulang dan prediksi dengan lima *seed* (7). Hasil ketiga lipatan digabung untuk evaluasi dan uji signifikansi selisih antarmodel (8), dan langkah terakhir adalah interpretasi kepentingan fitur (9).

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 1 Alur penelitian

3.1 Pengumpulan Data

Penelitian memakai Gold Coast City Wave Dataset periode 2014–2024 yang dikumpulkan Pemerintah Queensland melalui Coastal Data System dan diakses melalui repositori Kaggle [22] (diakses 23 Juni 2026). Berkas gabungan *combined_all_years.csv* memuat 180.482 baris pengamatan *buoy* dari 1 Januari 2014 pukul 00.30 hingga 31 Desember 2024 pukul 23.30, dengan interval 30 menit dan enam variabel: Hs, Hmax, Tz, Tp, arah puncak gelombang, dan suhu permukaan laut (SST). Deskripsi variabel disajikan pada Tabel 1 dan cuplikan data mentah pada Gambar 2. Lokasinya di Queensland, Australia, wilayah yang juga menjadi lokasi studi [2], [21]. Seluruh proses dijalankan dengan Python di Google Colab (CPU) menggunakan pandas, NumPy, scikit-learn, XGBoost 3.4.1, LightGBM 4.6.0, Optuna 5.0.0, dan SHAP.

Tabel 1 Deskripsi variabel pada dataset

Variabel	Satuan	Keterangan
Hs	m	Tinggi gelombang signifikan; variabel yang diprediksi
Hmax	m	Tinggi gelombang individu maksimum pada periode pengamatan
Tz	s	Periode rata-rata gelombang (*zero-crossing*)
Tp	s	Periode gelombang pada puncak spektrum energi
Arah puncak	°	Arah gelombang pada periode puncak (0–360°)
SST	°C	Suhu permukaan laut

	Timestamp	Hs_m	Hmax_m	Tz_s	Tp_s	Peak_Direction_deg	SST_degC
0	2014-01-01 00:30:00	0.6211	1.01	4.6196	5.0390	0.000	0.00
1	2014-01-01 01:00:00	0.6136	1.01	4.1473	15.9741	89.161	24.00
2	2014-01-01 01:30:00	0.6627	1.27	4.1951	7.0494	101.161	24.00
3	2014-01-01 02:00:00	0.7042	1.40	4.1625	4.2759	77.161	23.95
4	2014-01-01 02:30:00	0.7122	1.09	4.3272	15.9105	77.161	23.90

Gambar 2 Cuplikan data mentah

3.2 Praproses Data dan *Grid* Waktu

Pembersihan data dilakukan dalam beberapa tahap. Nilai −99,9 diubah menjadi NaN, begitu pula SST = 0 dan arah = 0 yang dianggap penanda isian, bukan hasil pengukuran. *Timestamp* ganda diperiksa

 ISSN:2302-8149
e-ISSN:2540-9719

dan dibuang bila ada, sedangkan nilai ekstrem tidak dibuang karena gelombang tinggi justru kejadian yang paling penting bagi peringatan dini. Data kemudian disusun pada *grid* 30 menit yang teratur sehingga fungsi *shift* untuk *lag* dan target selalu tepat berselisih kelipatan 30 menit; waktu yang tidak tercatat diisi NaN. Baris yang Hs saat ini atau targetnya (1 jam ke depan) hilang dibuang. Nilai fitur lain yang hilang tidak diisi karena XGBoost dan LightGBM dapat memprosesnya langsung, dan tidak dilakukan imputasi agar tidak ada informasi dari masa depan yang masuk.

3.3 Fitur dan Target

Fitur dibentuk hanya dari informasi sampai waktu t agar tidak ada kebocoran data dari masa depan [23]. Kesembilan belas fitur dikelompokkan pada Tabel 2. Arah gelombang dan waktu (bulan dan jam) dikodekan dengan sin dan cos karena sifatnya melingkar. Tahun tidak dipakai sebagai fitur karena model diuji pada tahun yang belum dilihat. Variabel angin dan atmosfer tidak dipakai karena harus tersedia atau diprakirakan lebih dulu saat prediksi dibuat [18],[19],[20].

Tabel 2 Variabel masukan model

Kelompok	Fitur	Jumlah
Hs saat ini dan riwayat	Hs(t), Hs *lag* 1–6 (30 menit hingga 3 jam sebelumnya)	7
Rata-rata bergerak Hs	Rata-rata 3 jam (6 langkah) dan 24 jam (48 langkah) terakhir	2
Parameter gelombang dan laut saat ini	Hmax, Tz, Tp, SST	4
Arah gelombang	Sin dan cos arah puncak	2
Waktu siklik	Sin dan cos bulan, sin dan cos jam	4
Total	–	19

Lag ke-k dihitung sebagai $\mathrm{Hs}(t - k\Delta t)$ dengan $k = 1$ hingga 6 dan $\Delta t = 30$ menit, sedangkan rata-rata bergerak adalah rata-rata w nilai Hs terakhir dengan w = 6 dan 48 langkah, dihitung bila tersedia minimal setengah jendela (3 dan 24 nilai). Arah gelombang, bulan, dan jam dikodekan dengan $\sin(2\pi x/P)$ dan $\cos(2\pi x/P)$, dengan $P = 360$ untuk arah (derajat), 12 untuk bulan, dan 24 untuk jam.

Target adalah perubahan Hs satu jam ke depan, yang dihitung dengan Persamaan (1):

$$\Delta\mathrm{Hs}(t) = \mathrm{Hs}(t + 1\,\mathrm{jam}) - \mathrm{Hs}(t) \quad (1)$$

Prediksi model ditambahkan kembali ke $\mathrm{Hs}(t)$, yaitu $\hat{\mathrm{H}}\mathrm{s}(t + 1\ \mathrm{jam}) = \mathrm{Hs}(t) + \Delta\hat{\mathrm{H}}\mathrm{s}(t)$, sehingga seluruh metrik dihitung pada *level* Hs. Baris yang Hs saat ini atau targetnya hilang dibuang, begitu pula baris yang waktu targetnya jatuh di luar periode latih, validasi, atau uji yang sama, sehingga tiga periode itu tidak saling tumpang tindih. Horizon prediksi ditetapkan satu jam (dua langkah 30 menit).

3.4 Pemisahan Data *Rolling-Origin*

Data dipisah menurut urutan waktu dengan skema *rolling-origin* [23]. Untuk tiap tahun uji Y (2022, 2023, 2024), data latih adalah semua tahun sampai $Y-2$, data validasi adalah tahun $Y-1$, dan data uji adalah tahun Y. Penyetelan hiperparameter diulang di setiap lipatan, dan data uji hanya dipakai pada evaluasi akhir. Gelombang tinggi didefinisikan sebagai Hs aktual pada atau di atas persentil ke-90 Hs aktual pada data latih lipatan itu [17].

3.5 Penyetelan Hiperparameter

Hiperparameter dicari dengan Optuna [24] berbasis Tree-structured Parzen Estimator (TPE), 15 percobaan awal acak, dan total 50 percobaan per model per lipatan; jumlah percobaan sama untuk kedua model. Kriterianya adalah RMSE pada data validasi lipatan itu, dengan *early stopping* 100

 ISSN:2302-8149
e-ISSN:2540-9719

ronde dan maksimum 3.000 pohon. Keduanya memakai metode histogram untuk mempercepat pelatihan dan meminimalkan fungsi objektif pada Persamaan (2). Ruang pencarian disetarakan menurut efek regularisasinya, bukan hanya menurut nilai numeriknya (Tabel 3). Pada fungsi kerugian galat kuadrat, turunan kedua (Hessian) tiap sampel bernilai 1, sehingga batas minimum bobot Hessian di daun pada XGBoost *(min_child_weight)* setara dengan batas minimum jumlah sampel di daun pada LightGBM *(min_child_samples)*; keduanya dicari pada rentang yang sama, 5–1000. Kedalaman maksimum (3–16) dan ambang minimum penurunan galat untuk membelah simpul (*gamma* pada XGBoost, *min_split_gain* pada LightGBM; 0–2) juga disamakan. Jumlah daun hanya dicari pada LightGBM karena pohonnya tumbuh per daun, sedangkan pohon XGBoost tumbuh per tingkat kedalaman; perbedaan ini bersifat bawaan algoritma dan tidak dapat disamakan sepenuhnya. Rentang ini diperluas setelah penyetelan awal (*min_child* 5–500, kedalaman 3–12) menunjukkan sebagian parameter mendekati batas atas; hasil penyetelan awal dibandingkan pada Subbab 4.8 sebagai analisis sensitivitas. Setelah itu tiap model dilatih ulang pada gabungan data latih dan validasi dengan jumlah pohon 1,1 kali iterasi terbaik, diulang dengan lima *seed* (42 hingga 46); *seed* 42 dipakai untuk penyetelan.

$$\text{Obj} = \sum_{i=1}^{n} l\left(\Delta\text{Hs}_i, \Delta\widehat{\text{Hs}}_i\right) + \sum_{k=1}^{K} \Omega(f_k) \quad (2)$$

dengan n jumlah sampel latih, ΔHs_i target aktual, $\Delta\widehat{\text{Hs}}_i = \sum_{k=1}^{K} f_k(x_i)$ prediksi dari K pohon, x_i vektor 19 fitur, l fungsi kerugian (galat kuadrat), f_k pohon ke-k, dan Ω suku regularisasi yang membatasi kompleksitas pohon [6], [7].

Tabel 3 Ruang pencarian hiperparameter

Hiperparameter	Rentang pencarian	XGBoost	LightGBM
Learning rate	0,02–0,2 (skala log)	✓	✓
Kedalaman maksimum	3–16	✓	✓
Jumlah daun	16–256 (skala log)	–	✓
Batas minimum di daun *(min_child_weight / min_child_samples)*	5–1000 (skala log)	✓	✓
Subsample	0,6–1,0	✓	✓
Colsample by tree	0,6–1,0	✓	✓
Reg lambda	0,001–30 (skala log)	✓	✓
Ambang minimum penurunan galat *(gamma / min_split_gain)*	0–2	✓	✓

Catatan: pada setiap baris yang memuat dua nama parameter, nama pertama untuk XGBoost dan nama kedua untuk LightGBM.

3.6 Evaluasi dan Uji Signifikansi

Metrik utama adalah RMSE pada *level* Hs (dalam meter) dan R^2 dari perubahan Hs (R^2_delta). RMSE dihitung dengan Persamaan (3):

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\text{Hs}_i - \widehat{\text{Hs}}_i\right)^2} \quad (3)$$

dengan n jumlah sampel uji, Hs_i nilai aktual, dan $\widehat{\text{Hs}}_i$ nilai prediksi pada *level* Hs. R^2_delta mengukur seberapa baik model menebak perubahan itu sendiri dan dihitung sebagai R^2 antara perubahan Hs aktual dan perubahan yang diprediksi model, dengan Persamaan (4):

$$R^2{}_{\text{delta}} = 1 - \frac{\sum_{i=1}^{n}\left(\Delta\text{Hs}_i - \Delta\widehat{\text{Hs}}_i\right)^2}{\sum_{i=1}^{n}\left(\Delta\text{Hs}_i - \overline{\Delta\text{Hs}}\right)^2} \quad (4)$$

dengan ΔHs_i perubahan aktual, $\Delta\widehat{\text{Hs}}_i$ perubahan yang diprediksi model, dan $\overline{\Delta\text{Hs}}$ rata-rata perubahan aktual. Dengan definisi ini *persistence* bernilai sekitar 0 dan nilai positif berarti model lebih baik daripada menebak Hs tidak berubah. Selisih RMSE antara dua model A dan B dinyatakan secara relatif terhadap RMSE model B dengan Persamaan (5):

 ISSN:2302-8149
e-ISSN:2540-9719

$$\mathrm{RMSE}_{\mathrm{rel}} = 100 \times \frac{\mathrm{RMSE}_B - \mathrm{RMSE}_A}{\mathrm{RMSE}_B} \quad (5)$$

Metrik pendukung adalah MAE, R^2 *level* (R^2 antara Hs aktual dan prediksi), MAPE (hanya untuk Hs aktual di atas 0,05 m), serta RMSE dan MAE pada gelombang tinggi. Sebagai acuan, *persistence* menganggap Hs(t + 1 jam) = Hs(t) tanpa pelatihan, dan dilaporkan pada Tabel 7, 8, dan 10. *Persistence* hanya acuan: tidak ikut uji signifikansi, dan kontras utama tetap XGBoost terhadap LightGBM.

Seluruh tabel utama (per lipatan, gabungan tiga tahun, dan gelombang tinggi) serta uji *bootstrap* memakai satu sumber prediksi, yaitu model dengan *seed* 42. Rata-rata dan simpangan baku lima *seed* (42–46) hanya dilaporkan sebagai uji ketahanan terhadap *seed*, dan *seed* yang dipakai dicantumkan pada keterangan setiap tabel. Karena galat prediksi antarwaktu saling berkaitan, selisih antarmodel diuji dengan *block bootstrap* [25]. Galat kedua model dihitung pada blok yang sama (berpasangan): 2.000 pengulangan, blok 48 langkah (satu hari). Untuk XGBoost (A) terhadap LightGBM (B), pada setiap pengulangan dihitung selisih RMSE relatif dalam persen dengan Persamaan (5) (positif berarti XGBoost lebih baik), dan selang kepercayaan 95% diambil dari persentil ke-2,5 dan ke-97,5 hasil pengulangan itu. Selang yang seluruhnya di atas nol berarti XGBoost lebih baik, seluruhnya di bawah nol berarti LightGBM lebih baik. Selang yang memotong nol tetapi seluruhnya berada dalam ±0,5% dinyatakan setara, dan selain itu dinyatakan belum tegas. Margin 0,5% adalah pilihan penulis. Sensitivitas terhadap panjang blok diuji dengan blok 144 langkah (tiga hari). RMSE dan MAE gabungan pada gelombang tinggi dilengkapi selang kepercayaan 95% dari prosedur *block bootstrap* yang sama pada deret penuh, dengan hanya menghitung waktu yang tergolong gelombang tinggi pada tiap pengulangan; selang ini dilaporkan per model dan tidak dipakai untuk menyatakan selisih antarmodel pada gelombang tinggi. Horizon evaluasi (1 jam) dipilih setelah uji pendahuluan pada horizon 1, 3, dan 6 jam yang hasilnya tidak dilaporkan dalam naskah ini (lihat Subbab 4.9). Setelah horizon dipilih dan sebelum analisis signifikansi statistik dilakukan, ditetapkan kontras utama, yaitu selisih RMSE relatif XGBoost terhadap LightGBM pada gabungan tiga tahun uji. Hasil per tahun uji bersifat sekunder dan dibaca sebagai indikatif, karena tidak ada koreksi perbandingan ganda.

Kepentingan fitur dihitung pada lipatan uji 2024 dengan *permutation importance* per kelompok fitur (fitur dalam satu kelompok diacak bersama, lima pengulangan, diukur sebagai kenaikan RMSE pada target perubahan Hs) dan dengan SHAP sebagai pelengkap (rata-rata nilai SHAP mutlak per kelompok pada 3.000 sampel uji acak), memakai model *seed* 42. Waktu pelatihan dan waktu prediksi diukur pada CPU Google Colab tanpa kontrol beban dan dirata-ratakan pada tiga lipatan dan lima seed.

4 Hasil dan Pembahasan

Hasil disajikan berurutan: karakteristik data dan hasil praproses (4.1), penyetelan hiperparameter (4.2), kinerja gabungan tiga tahun uji (4.3), kinerja per tahun uji (4.4), uji signifikansi (4.5), gelombang tinggi dan waktu komputasi (4.6), kepentingan fitur (4.7), pembahasan (4.8), dan keterbatasan (4.9).

4.1 Karakteristik Data dan Hasil Praproses

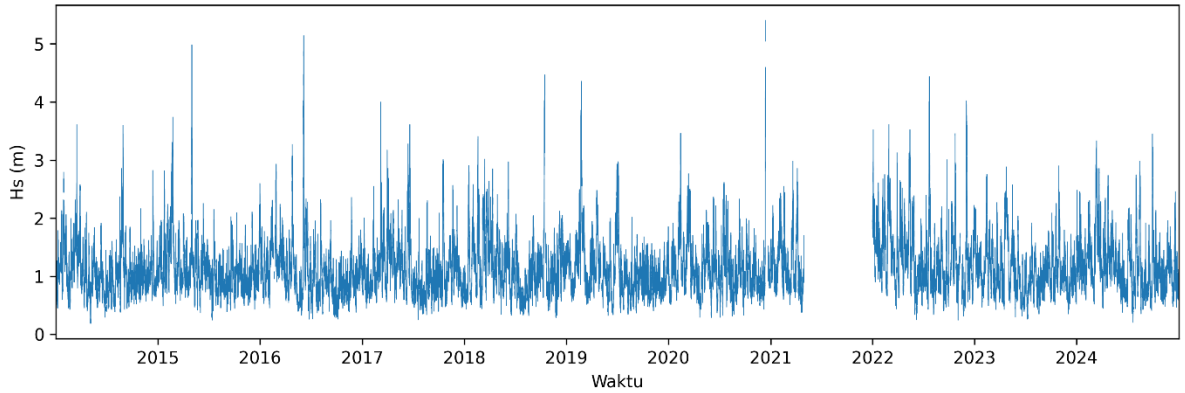
Pembersihan dan penyusunan *grid* waktu menghasilkan jumlah data pada Tabel 4. Setelah pembersihan, data hilang 1,10% pada Hs, Hmax, Tz, dan Tp, 1,11% pada arah puncak, dan 10,43% pada SST. Hs berkisar 0,186–5,405 m, rentang yang wajar, sehingga tidak ada nilai ekstrem yang dibuang. Deret waktu Hs ditunjukkan pada Gambar 3. Setelah baris dengan Hs saat ini atau target yang hilang dibuang, tersisa 178.170 baris layak pakai.

Tabel 4 Jumlah data pada tahap pembersihan dan penyusunan *grid*

Tahap	Keterangan	Jumlah (baris)

 ISSN:2302-8149
e-ISSN:2540-9719

Data mentah	Pengamatan *buoy*, 1 Jan 2014 00.30–31 Des 2024 23.30	180.482
Kode −99,9 menjadi NaN	Hs, Hmax, Tz, dan Tp (masing-masing)	1.984
Kode −99,9 menjadi NaN	Arah puncak gelombang	1.622
Kode −99,9 menjadi NaN	SST	18.447
Penanda isian menjadi NaN	SST = 0	378
Penanda isian menjadi NaN	Arah = 0	382
Timestamp ganda	Jumlah yang ditemukan	0
Grid 30 menit	180.482 asli + 12.381 waktu tidak tercatat	192.863
Hs hilang pada *grid*	Termasuk 11.886 baris pada 2021	14.365
Target 1 jam hilang	Hs(t) ada, Hs(t + 1 jam) hilang	328
Baris layak pakai	Hs saat ini dan target 1 jam tersedia	178.170



Gambar 3 Deret waktu Hs

Ukuran data latih, validasi, dan uji tiap lipatan serta ambang gelombang tinggi dirinci pada Tabel 5. Validasi untuk lipatan 2022 hanya 5.543 baris karena 2021 kehilangan sekitar dua pertiga data Hs.

Tabel 5 Ukuran data dan ambang gelombang tinggi tiap lipatan

Tahun uji	Latih	Validasi	Uji	Ambang gelombang tinggi (m)	Sampel uji gelombang tinggi
2022	121.106	5.543	16.819	1,659	4.076
2023	126.651	16.819	17.457	1,673	1.505
2024	143.470	17.457	17.239	1,739	1.877

4.2 Hasil Penyetelan Hiperparameter

Pencarian Optuna 50 percobaan per model per lipatan menghasilkan RMSE validasi yang hampir sama untuk kedua model, dengan LightGBM sedikit lebih rendah pada ketiga lipatan. RMSE validasi terbaik tiap model pada setiap lipatan disajikan pada Tabel 6.

Tabel 6 RMSE validasi terbaik hasil penyetelan (m)

Tahun uji (lipatan)	XGBoost	LightGBM
2022	0,08147	0,08109
2023	0,09287	0,09251
2024	0,06729	0,06724

Setelah ruang pencarian disetarakan dan diperluas, hiperparameter terbaik LightGBM tidak lagi berada di tepi rentang: *min child samples* bernilai 446, 446, dan 660 dari rentang 5–1000, dan kedalaman maksimum 13, 13, dan 12 dari batas atas 16. Pada XGBoost, *min child weight* bernilai 12,3, 148,5, dan 9,6, dengan kedalaman maksimum 3, 15, dan 8. Beberapa parameter XGBoost dan satu parameter LightGBM masih dekat batas: kedalaman maksimum XGBoost 3 (batas bawah) pada

 ISSN:2302-8149
e-ISSN:2540-9719

2022, *learning rate* XGBoost 0,0201 (batas bawah 0,02) pada 2024, *reg lambda* XGBoost 0,00135 (batas bawah 0,001) pada 2023, dan *colsample by tree* LightGBM 0,618 (batas bawah 0,6) pada 2024. Nilai *gamma* dan *min_split_gain* pada kedua model mendekati nol, yaitu tanpa pemangkasan tambahan. Konfigurasi XGBoost yang terpilih berbeda jauh antarlipatan (kedalaman 3 hingga 15) dengan RMSE validasi yang hampir sama, yang mengisyaratkan permukaan optimasi yang relatif datar. Implikasinya dibahas pada Subbab 4.8.

4.3 Kinerja pada Gabungan Tiga Tahun Uji

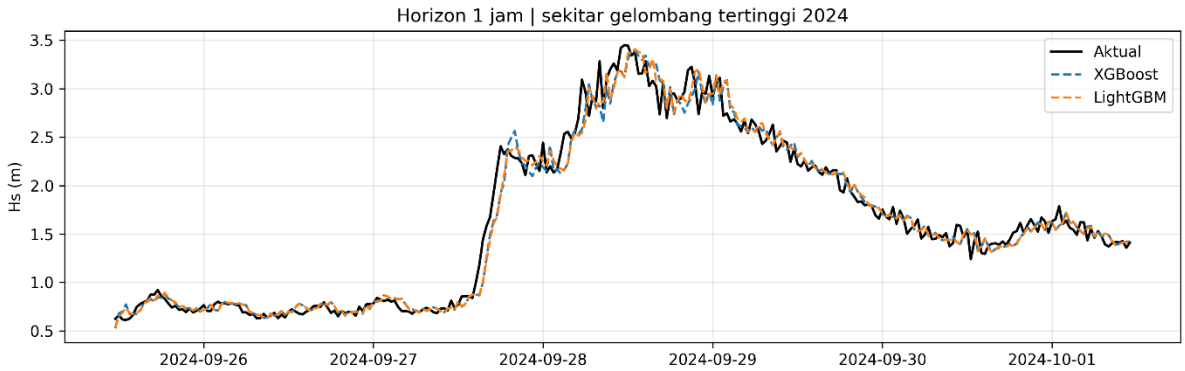
Prediksi tiga tahun uji (2022, 2023, 2024) digabung karena seluruhnya bersifat *out-of-sample*. Kinerja gabungan tiga tahun uji disajikan pada Tabel 7. LightGBM memiliki RMSE terendah (0,0787 m) dan R^2 perubahan Hs tertinggi (0,131), diikuti XGBoost (0,0790 m; 0,123). Selisih RMSE kedua model kecil, sekitar 0,4 mm atau 0,49%, dan arahnya sama pada kelima *seed* (selisih relatif −0,49% hingga −0,54%). Sebagai acuan, RMSE *persistence* adalah 0,0844 m, sehingga RMSE LightGBM sekitar 6,8% dan XGBoost sekitar 6,4% lebih rendah, atau sekitar 5,4–5,7 mm. R^2 *level* hampir tidak membedakan model (0,9758 dan 0,9756) dan sudah tinggi pada *persistence* (0,9722), sehingga tidak dijadikan ukuran utama; R^2 perubahan Hs menunjukkan bahwa model hanya menjelaskan sekitar 12–13% variasi perubahan Hs satu jam.

Tabel 7 Kinerja pada gabungan tiga tahun uji, horizon 1 Jam

Model	R^2_delta	RMSE (m)	MAE (m)	R^2 *level*	MAPE (%)
LightGBM	0,131	0,0787	0,0563	0,9758	4,84
XGBoost	0,123	0,0790	0,0566	0,9756	4,86
Persistence (acuan)	0	0,0844	0,0602	0,9722	5,11

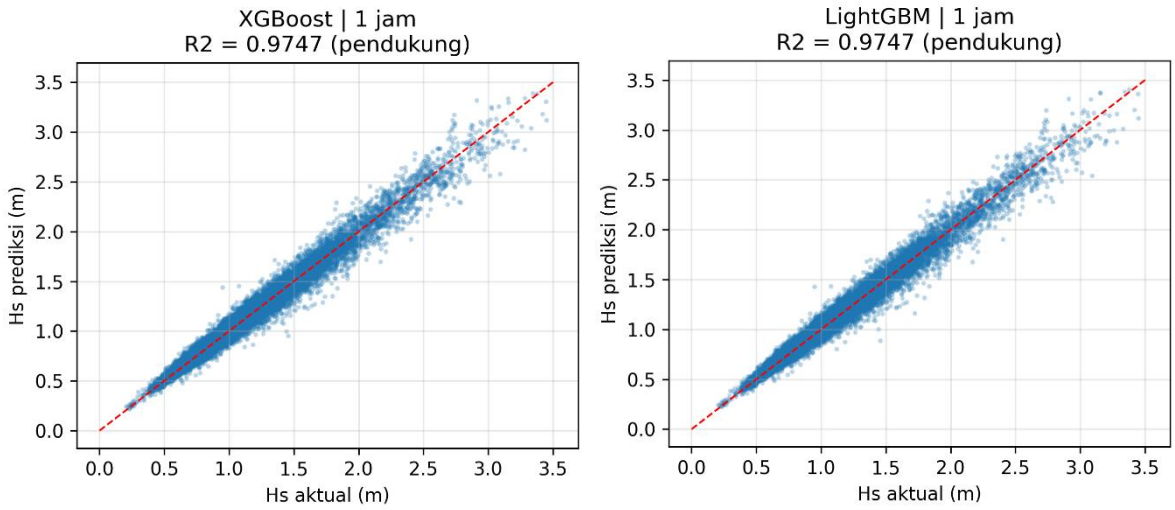
Catatan: XGBoost dan LightGBM memakai prediksi *seed* 42 dengan hiperparameter hasil 50 percobaan per lipatan.

Contoh prediksi Hs 1 jam ke depan dibandingkan dengan nilai aktual pada tahun uji 2024 ditunjukkan pada Gambar 4 dan Gambar 5.



Gambar 4 Prediksi Hs 1 jam ke depan vs aktual pada periode gelombang tertinggi 2024 (seed 42)

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 5 Sebaran Hs prediksi vs aktual pada tahun uji 2024 (seed 42)

Kedua model mengikuti pola Hs aktual, tetapi pada kenaikan gelombang yang tiba-tiba prediksi cenderung terlambat. Hal ini sejalan dengan R^2 perubahan Hs yang hanya 0,12 sampai 0,13.

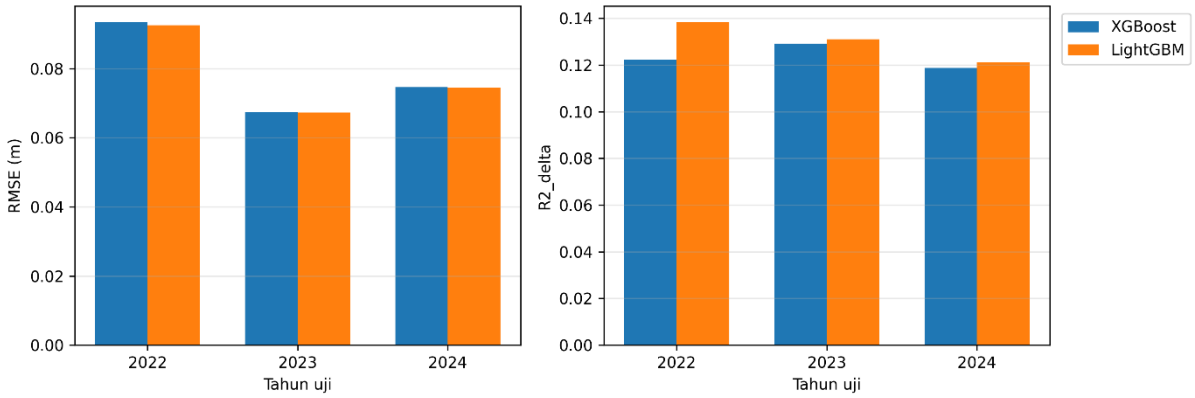
4.4 Kinerja per Tahun Uji

Kinerja per tahun uji disajikan pada Tabel 8 dan divisualisasikan pada Gambar 6. LightGBM memiliki RMSE lebih rendah pada ketiga tahun uji, dengan selisih terhadap XGBoost paling besar pada 2022 (0,0925 dan 0,0934 m) dan nyaris hilang pada 2023 (0,0673 m pada keduanya) dan 2024 (0,0745 dan 0,0746 m). RMSE kedua model paling besar pada 2022 dan paling kecil pada 2023, jadi ketiga tahun uji memiliki tingkat kesulitan yang berbeda. Simpangan baku RMSE antar lima *seed* sangat kecil (0,00002 hingga 0,0001 m); angka ini jauh lebih kecil daripada selisih antarmodel pada 2022, tetapi sebanding dengan selisihnya pada 2023 dan 2024, sehingga perbedaan pada dua tahun terakhir tidak dapat dibedakan dari variasi *seed*.

Tabel 8 Kinerja per tahun uji (*seed* 42)

Tahun Uji	Model	RMSE (m)	MAE (m)	R^2 delta
2022	LightGBM	0,0925	0,0657	0,138
2022	XGBoost	0,0934	0,0662	0,122
2022	*Persistence* (acuan)	0,0997	0,0709	0
2023	LightGBM	0,0673	0,0492	0,131
2023	XGBoost	0,0673	0,0494	0,129
2023	*Persistence* (acuan)	0,0721	0,0524	0
2024	LightGBM	0,0745	0,0543	0,121
2024	XGBoost	0,0746	0,0545	0,119
2024	*Persistence* (acuan)	0,0795	0,0576	0

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 6 RMSE dan R^2 Perubahan Hs per Tahun Uji

4.5 Uji Signifikansi

Kontras utama (XGBoost terhadap LightGBM, gabungan tiga tahun) menunjukkan LightGBM lebih baik: selisih RMSE relatif XGBoost terhadap LightGBM adalah −0,49% (Tabel 9), dan selangnya (−0,85% hingga −0,12%) seluruhnya di bawah nol. Hasil yang sama diperoleh pada blok 48 dan 144 langkah (selang pada blok 144: −0,91% hingga −0,12%). Selang itu tidak seluruhnya berada dalam margin kesetaraan ±0,5% (batas bawahnya −0,85%), sehingga kedua model tidak dapat dinyatakan setara; namun batas atasnya hanya −0,12%, sehingga keunggulan LightGBM, walaupun terdeteksi secara statistik, bernilai kecil. Per tahun uji, LightGBM lebih baik pada 2022, kedua model setara pada 2023, dan pada 2024 selisihnya belum tegas. Perbandingan per tahun bersifat indikatif karena tidak ada koreksi perbandingan ganda.

Tabel 9 Selisih RMSE relatif XGBoost terhadap LightGBM (%), *seed* 42 (negatif berarti LightGBM lebih baik)

Tahun Uji	Selisih RMSE relatif (%)	Selang 95%	Keputusan
Gabungan (kontras utama)	−0,49	−0,85 s.d. −0,12	LightGBM lebih baik
2022	−0,93	−1,58 s.d. −0,25	LightGBM lebih baik
2023	−0,11	−0,43 s.d. 0,25	Setara
2024	−0,14	−0,64 s.d. 0,38	Belum tegas

4.6 Gelombang Tinggi dan Waktu Komputasi

Pada gelombang tinggi (Hs aktual ≥ persentil ke-90 data latih), galat kedua model sekitar 1,6 hingga 1,8 kali galat keseluruhan, sehingga kesalahan terbesar terjadi pada kondisi yang paling penting bagi keselamatan. Pola ini sejalan dengan perlunya analisis gelombang ekstrem secara terpisah [17]. RMSE dan MAE pada gelombang tinggi per tahun uji, serta nilai gabungan tiga tahun beserta selang kepercayaan 95%, disajikan pada Tabel 10. RMSE LightGBM lebih rendah daripada XGBoost pada 2022 (0,1449 dan 0,1464 m) dan 2024 (0,1295 dan 0,1301 m), sedangkan pada 2023 keduanya hampir identik (0,1193 m); MAE menunjukkan pola serupa, dengan XGBoost sedikit lebih rendah pada 2023. Keduanya lebih rendah daripada acuan *persistence* pada ketiga tahun uji. Pada gabungan tiga tahun (7.458 sampel gelombang tinggi), RMSE LightGBM adalah 0,1363 m dan XGBoost 0,1373 m, sedangkan MAE masing-masing 0,1044 m dan 0,1053 m. Selang kepercayaan kedua model saling menimpa hampir seluruhnya, dan selisih antarmodel pada gelombang tinggi tidak diuji secara statistik, sehingga selisih ini dibaca sebagai deskriptif.

Tabel 10 RMSE dan MAE pada gelombang tinggi (m), *seed* 42. angka dalam kurung pada baris gabungan adalah selang kepercayaan 95% *block bootstrap*.

Tahun Uji	RMSE	RMSE	RMSE	MAE	MAE

 ISSN:2302-8149
e-ISSN:2540-9719

(jumlah sampel)	LightGBM	XGBoost	*Persistence* (acuan)	LightGBM	XGBoost
2022 (4.076)	0,1449	0,1464	0,1542	0,1109	0,1121
2023 (1.505)	0,1193	0,1193	0,1310	0,0926	0,0925
2024 (1.877)	0,1295	0,1301	0,1374	0,0998	0,1006
Gabungan (7.458)	0,1363 (0,1298–0,1426)	0,1373 (0,1303–0,1439)	0,1456	0,1044 (0,0998–0,1091)	0,1053 (0,1004–0,1101)

Waktu pelatihan rata-rata per model (tiga lipatan, lima seed) adalah 11,1 detik untuk LightGBM dan 14,5 detik untuk XGBoost, dengan waktu prediksi 0,99 detik untuk LightGBM dan 0,26 detik untuk XGBoost. Urutan kecepatan pelatihan berubah antar lipatan (LightGBM lebih cepat pada 2022 dan 2024, XGBoost lebih cepat pada 2023), sedangkan XGBoost lebih cepat memprediksi pada ketiga lipatan. Waktu ini dipengaruhi hiperparameter hasil penyetelan yang berbeda antarmodel dan antarlipatan (misalnya kedalaman dan jumlah pohon), dan pengukuran dilakukan pada CPU Colab tanpa kontrol beban, sehingga waktu komputasi tidak dipakai sebagai dasar memilih model.

4.7 Kepentingan Fitur

Kepentingan fitur dihitung pada lipatan uji 2024 dengan *permutation importance* per kelompok fitur dan SHAP [26], memakai model *seed* 42. Hasil *permutation importance* dan SHAP per kelompok fitur disajikan pada Tabel 11. Fitur diacak per kelompok karena Hs saat ini, *lag*, dan rata-ratanya saling berkorelasi kuat. Pada kedua model, riwayat Hs paling berpengaruh: mengacaknya menaikkan RMSE pada target perubahan Hs sebesar 11,4% (XGBoost) dan 10,6% (LightGBM). Berikutnya adalah periode gelombang (Tz dan Tp), sebesar 4,4% dan 2,8%. SST praktis tidak berpengaruh (−0,05% dan 0,01%). Urutan kelompok sedang pada SHAP sedikit berbeda dari *permutation importance* (misalnya musim dan jam serta arah puncak di atas Hmax), tetapi kelompok teratasnya sama.

Tabel 11 Kepentingan kelompok fitur pada lipatan uji 2024

Kelompok Fitur	Kenaikan RMSE bila diacak, XGBoost (%)	Kenaikan RMSE bila diacak, LightGBM (%)	Rata-rata nilai SHAP mutlak, XGBoost (m)	Rata-rata nilai SHAP mutlak, LightGBM (m)
Riwayat Hs (Hs(t), *lag*, rata-rata)	11,42	10,63	0,0143	0,0159
Periode (Tz, Tp)	4,43	2,81	0,0087	0,0089
Hmax	2,02	1,11	0,0011	0,0017
Musim dan jam	1,31	1,23	0,0045	0,0046
Arah puncak	0,83	0,82	0,0030	0,0032
SST	−0,05	0,01	0,0010	0,0012

Kepentingan ini mencerminkan seberapa jauh model bergantung pada suatu kelompok fitur, bukan hubungan sebab-akibat.

4.8 Pembahasan

Pada horizon 1 jam, LightGBM sedikit lebih akurat daripada XGBoost pada penyetelan akhir: RMSE XGBoost 0,49% lebih tinggi (sekitar 0,4 mm pada gabungan) dengan selang kepercayaan −0,85% hingga −0,12%, dan arah selisih yang sama pada kelima *seed*. Selisih ini tegas pada gabungan dan pada 2022, setara pada 2023, dan belum tegas pada 2024. Karena itu temuan ini sebaiknya dibaca sebagai “LightGBM cenderung sedikit lebih baik”, bukan sebagai perbedaan besar antara kedua algoritma.

Selisih tersebut sensitif terhadap ruang pencarian hiperparameter (Tabel 12). Pada penyetelan awal dengan ruang yang lebih sempit tetapi sama untuk kedua model (*min child* 5–500, kedalaman maksimum 3–12, 50 percobaan), RMSE gabungan XGBoost dan LightGBM praktis sama (0,0787 m), dengan selisih relatif 0,01% yang dinyatakan setara. Ketika ruang diperluas, hasil LightGBM nyaris tidak berubah, tetapi RMSE gabungan XGBoost naik dari 0,0787 menjadi 0,0790 m. Karena ruang yang lebih lebar memuat ruang yang lebih sempit, RMSE validasi XGBoost yang lebih tinggi pada

 ISSN:2302-8149
e-ISSN:2540-9719

ketiga lipatan (misalnya 0,0812 menjadi 0,0815 m pada 2022) menunjukkan bahwa 50 percobaan belum tentu cukup untuk menemukan konfigurasi terbaik XGBoost pada ruang yang lebih lebar; konfigurasi yang terpilih pun berbeda jauh antarlipatan (Subbab 4.2). Dengan demikian, selisih 0,4 mm pada Tabel 7 tidak dapat sepenuhnya dikaitkan dengan perbedaan algoritma: perbedaan ruang pencarian, tingkat regularisasi, dan kecukupan anggaran penyetelan merupakan penjelasan alternatif yang tidak dapat dikesampingkan, dan pengaruhnya sebanding dengan selisih antarmodel itu sendiri.

Tabel 12 Sensitivitas hasil gabungan tiga tahun uji terhadap ruang pencarian hiperparameter (*seed* 42)

Ruang pencarian (sama untuk kedua model)	RMSE XGBoost (m)	RMSE LightGBM (m)	Selisih RMSE relatif (%)	Selang 95%	Keputusan
Min child 5–500; kedalaman 3–12	0,0787	0,0787	0,01	−0,24 s.d. 0,27	Setara
Min child 5–1000; kedalaman 3–16 (hasil utama)	0,0790	0,0787	−0,49	−0,85 s.d. −0,12	LightGBM lebih baik

Dari sisi algoritma, salah satu kemungkinan penyebab keunggulan tipis LightGBM adalah struktur pohon *leaf-wise* yang cocok dengan data sekitar 120–145 ribu baris latih, tetapi ini dugaan yang tidak diuji. Studi [20] juga membandingkan XGBoost dan LightGBM, tetapi dengan masukan angin dan atmosfer serta skema evaluasi yang berbeda, sehingga hasilnya tidak dapat dibandingkan langsung dengan temuan ini.

Kedua model hanya menjelaskan sebagian kecil variasi perubahan Hs: R^2_delta hanya 0,12–0,13. Artinya sebagian besar perubahan Hs dalam satu jam tidak dapat dijelaskan oleh fitur yang hanya berisi pengamatan sampai waktu t. Hal ini sejalan dengan analisis kepentingan fitur: riwayat Hs mendominasi, periode gelombang menyumbang sedikit, dan SST nyaris tidak berpengaruh. Informasi tambahan seperti angin atau prakiraan numerik [18], [20] mungkin diperlukan untuk menangkap kenaikan gelombang yang tiba-tiba, tetapi tidak diuji di sini. R^2 *level* (0,976 untuk LightGBM) tampak tinggi, namun nilainya hampir sama pada kedua model dan sebagian besar mencerminkan kedekatan Hs(t + 1 jam) dengan Hs(t), sehingga tidak dijadikan ukuran utama.

Pada gelombang tinggi, galat naik sekitar 1,6–1,8 kali pada kedua model. Pada gabungan tiga tahun, RMSE LightGBM (0,1363 m) dan XGBoost (0,1373 m) memiliki selang kepercayaan yang saling menimpa hampir seluruhnya, dan selisih per tahun hilang pada 2023. Jadi keunggulan LightGBM pada kondisi yang paling penting bagi keselamatan tidak dapat dinyatakan, dan perlu diuji tersendiri dengan uji berpasangan pada gabungan tiga tahun.

Hasil ini tidak dapat dibandingkan langsung dengan angka kesalahan dari studi lain, misalnya [8] dan [9], karena lokasi, horizon, fitur masukan, dan skema pembagian datanya berbeda. Studi [21] memang memakai wilayah perairan Queensland, tetapi dengan horizon, fitur, dan skema evaluasi yang berbeda, sehingga angkanya pun tidak dapat dibandingkan langsung. Kelebihan penelitian ini terletak pada rancangan evaluasinya: kedua algoritma dibandingkan dengan data, fitur, jumlah percobaan penyetelan, dan ruang pencarian yang disetarakan, diuji pada tiga tahun uji yang tidak dilihat saat pelatihan, dengan selisih yang diuji memakai *block bootstrap* berpasangan dan dilengkapi analisis sensitivitas terhadap ruang pencarian.

4.9 Keterbatasan

Penelitian ini memiliki beberapa keterbatasan. Data berasal dari satu stasiun sehingga generalisasi ke lokasi lain belum diuji. Hanya satu horizon (1 jam) yang dilaporkan, dan pemilihannya didahului uji pendahuluan pada horizon 1, 3, dan 6 jam yang hasilnya tidak dilaporkan dalam naskah ini (lihat Subbab 3.6); karena horizon dipilih setelah penulis melihat hasil uji pendahuluan, pemilihan horizon berpotensi memengaruhi interpretasi hasil. Tahun 2021 kehilangan sekitar 11.886 baris Hs sehingga validasi lipatan 2022 hanya 5.543 baris dan penyetelan pada lipatan itu kurang andal. Pergantian jenis *buoy* ke Mk4 pada Oktober 2022 jatuh di dalam tahun uji 2022, sedangkan 2023–2024 seluruhnya Mk4 (menurut metadata resmi Pemerintah Queensland), dan pengaruhnya tidak dianalisis. Meskipun ruang pencarian telah disetarakan dan diperluas, beberapa hiperparameter XGBoost masih dekat batas

 ISSN:2302-8149
e-ISSN:2540-9719

rentang (Subbab 4.2), anggaran penyetelan dibatasi 50 percobaan per model per lipatan, dan hasil gabungan, terutama XGBoost, berubah ketika ruang pencarian diubah (Subbab 4.8). Penyetelan berulang dengan beberapa *seed sampler* tidak dilakukan, sehingga selisih antarmodel tidak dapat sepenuhnya dipisahkan dari variasi penyetelan. Hanya kontras utama yang ditetapkan sebelum analisis signifikansi dilakukan; hasil per tahun uji bersifat indikatif tanpa koreksi perbandingan ganda, dan margin kesetaraan 0,5% adalah pilihan penulis. Seluruh tabel utama dan uji signifikansi memakai prediksi satu *seed* (42), sedangkan hasil lima *seed* hanya dipakai sebagai uji ketahanan. Selisih antarmodel pada gelombang tinggi tidak diuji secara statistik (hanya RMSE dan MAE gabungan dengan selang kepercayaan per model), dan kepentingan fitur hanya dihitung pada lipatan 2024 serta tidak menunjukkan hubungan sebab-akibat. Terakhir, waktu komputasi diukur pada CPU Colab tanpa kontrol beban sehingga tidak dapat dijadikan kesimpulan kecepatan.

5 Kesimpulan

Penelitian ini membandingkan XGBoost dan LightGBM untuk memprediksi Hs 1 jam ke depan pada data *buoy* Gold Coast periode 2014–2024, dengan evaluasi *rolling-origin* pada tiga tahun uji (2022, 2023, 2024), ruang pencarian hiperparameter yang disetarakan, dan fitur yang hanya memakai informasi sampai waktu prediksi. Pada gabungan tiga tahun uji, LightGBM mencapai RMSE 0,0787 m dan XGBoost 0,0790 m, sehingga RMSE XGBoost 0,49% lebih tinggi (selang kepercayaan 95%: −0,85 hingga −0,12); selisih ini tegas pada gabungan dan pada 2022, setara pada 2023, dan belum tegas pada 2024. Selisih tersebut sensitif terhadap ruang pencarian hiperparameter: pada ruang yang lebih sempit kedua model setara, sehingga keunggulan tipis LightGBM tidak dapat sepenuhnya dikaitkan dengan algoritmanya. Kedua model hanya menjelaskan sekitar 12–13% variasi perubahan Hs satu jam, dan pada gelombang tinggi selang kepercayaan RMSE dan MAE gabungan kedua model saling menimpa. Kesimpulannya, pada horizon 1 jam LightGBM sedikit lebih akurat daripada XGBoost, tetapi selisihnya kecil (sekitar 0,4 mm) dan sensitif terhadap penyetelan, sehingga kedua model dapat dipandang sebanding. Penelitian selanjutnya dapat menambah horizon dan stasiun, memperbanyak percobaan penyetelan dengan beberapa *seed sampler*, memasukkan data angin atau prakiraan numerik, dan menganalisis kinerja pada gelombang tinggi secara lebih rinci.

Referensi

[1] D. Saydam Irim, M. Sarıkaya, and S. Dağli, "*Data-Driven Prediction of Wave Energy Potential in the Black Sea using Ensemble Learning and Real-Time Buoy Observations*," *Arab. J. Sci. Eng.*, Vol. 51, No. 14, pp. 17069–17089, Mar. 2026, DOI: 10.1007/S13369-026-11224-Y.

[2] N. Raj and R. Prakash, "*Assessment and Prediction of Significant Wave Height using Hybrid CNN-BiLSTM Deep Learning Model for Sustainable Wave Energy in Australia*," *Sustain. Horizons*, Vol. 11, p. 100098, Sep. 2024, DOI: 10.1016/J.HORIZ.2024.100098.

[3] I. Yanchin and C. Guedes Soares, "*Gradient Boosted Trees and Denoising Autoencoder to Correct Numerical Wave Forecasts*," *J. Mar. SCI. Eng.*, Vol. 12, No. 9, p. 1573, Sep. 2024, DOI: 10.3390/JMSE12091573.

[4] M. O. Costa, R. M. Campos, and C. Guedes Soares, "*Enhancing the Accuracy of Metocean Hindcasts with Machine Learning Models*," *Ocean Eng.*, Vol. 287, p. 115724, Nov. 2023, DOI: 10.1016/J.OCEANENG.2023.115724.

[5] C. S. Alfredo and D. Adytia, "*Time Series Forecasting of Significant Wave Height using GRU, CNN-GRU, and LSTM*," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, Vol. 6, No. 5, pp. 776–781, Oct. 2022, DOI: 10.29207/RESTI.V6I5.4160.

[6] T. Chen and C. Guestrin, "*XGBoost: A Scalable Tree Boosting System*," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 785–794, Aug. 2016, DOI: 10.1145/2939672.2939785.

[7] G. Ke *et al.*, "*LightGBM: A Highly Efficient Gradient Boosting Decision Tree*," *Adv. Neural Inf. Process. Syst.*, Vol. 30, pp. 3146–3154, 2017, DOI: 10.5555/3294996.3295074.

[8] F. Anggraeni, D. Adytia, and A. W. Ramadhan, "*Forecasting of Wave Height Time Series using AdaBoost and XGBoost, Case Study in Pangandaran, Indonesia*," *2021 Int. Conf. Data*

 ISSN:2302-8149
e-ISSN:2540-9719

SCI. Its Appl. ICoDSA 2021, pp. 97–101, 2021, DOI: 10.1109/ICODSA53588.2021.9617524.

[9] H. Hu, A. J. van der Westhuysen, P. Chu, and A. Fujisaki-Manome, "*Predicting Lake Erie Wave Heights and Periods using XGBoost and LSTM*," *Ocean Model.*, Vol. 164, p. 101832, Aug. 2021, DOI: 10.1016/J.OCEMOD.2021.101832.

[10] P. Martina Maria Pushpam and V. S. Felix Enigo, "*Forecasting Significant Wave Height using RNN-LSTM Models*," *Proc. Int. Conf. Intell. Comput. Control Syst. ICICCS 2020*, pp. 1141–1146, May 2020, DOI: 10.1109/ICICCS48265.2020.9121040.

[11] T. Song *et al.*, "*Prediction of Significant Wave Height based on EEMD and Deep Learning*," *Front. Mar. SCI.*, Vol. 10, p. 1089357, Feb. 2023, DOI: 10.3389/FMARS.2023.1089357.

[12] E. Androulakis and G. Galanis, "*A Two-Step Hybrid System Towards Optimized Wave Height Forecasts*," *Stoch. Environ. Res. Risk Assess.*, Vol. 36, No. 3, pp. 753–766, Aug. 2021, DOI: 10.1007/S00477-021-02075-0.

[13] M. Karbasi, M. Jamei, M. Ali, S. Abdulla, X. Chu, and Z. M. Yaseen, "*Developing a Novel Hybrid Auto Encoder Decoder Bidirectional Gated Recurrent Unit model Enhanced with Empirical Wavelet Transform and Boruta-Catboost to Forecast Significant Wave Height*," *J. Clean. Prod.*, Vol. 379, p. 134820, Dec. 2022, DOI: 10.1016/J.JCLEPRO.2022.134820.

[14] J. Shi *et al.*, "*A Machine-Learning Approach based on Attention Mechanism for Significant Wave Height Forecasting*," *J. Mar. SCI. Eng.*, Vol. 11, No. 9, p. 1821, Sep. 2023, DOI: 10.3390/JMSE11091821.

[15] Y. Zhan, H. Zhang, J. Li, and G. Li, "*Prediction Method for Ocean Wave Height based on Stacking Ensemble Learning Model*," *J. Mar. SCI. Eng.*, Vol. 10, No. 8, p. 1150, Aug. 2022, DOI: 10.3390/JMSE10081150.

[16] S. Gracia, J. Olivito, J. Resano, B. Martin-del-Brio, M. de Alfonso, and E. Álvarez, "*Improving Accuracy on Wave Height Estimation through Machine Learning Techniques*," *Ocean Eng.*, Vol. 236, p. 108699, Sep. 2021, DOI: 10.1016/J.OCEANENG.2021.108699.

[17] M. S. Afzal, L. Kumar, V. Chugh, Y. Kumar, and M. Zuhair, "*Prediction of Significant Wave Height using Machine Learning and Its Application to Extreme Wave Analysis*," *J. Earth Syst. SCI.*, Vol. 132, No. 2, p. 51, Mar. 2023, DOI: 10.1007/S12040-023-02058-5.

[18] M. Ré Henriques, D. Silva, I. Yanchin, M. Latas, and C. Guedes Soares, "*Improving the Forecast of Wind Speed and Significant Wave Height using Neural Networks and Gradient Boosting Trees*," *Ocean Eng.*, Vol. 327, p. 120925, May 2025, DOI: 10.1016/J.OCEANENG.2025.120925.

[19] D. Adytia, D. Saepudin, S. R. Pudjaprasetya, S. Husrin, and A. Sopaheluwakan, "*A Deep Learning Approach for Wave Forecasting based on a Spatially Correlated Wind Feature, with a Case Study in the Java Sea, Indonesia*," *Fluids*, Vol. 7, No. 1, p. 39, Jan. 2022, DOI: 10.3390/FLUIDS7010039.

[20] A. Hasan, I. Kayes, M. Alam, T. Shahriar, and M. Ahsan Habib, "*Generalized Machine Learning Models to Predict Significant Wave Height Utilizing Wind and Atmospheric Parameters*," *Energy Convers. Manag. X*, Vol. 23, p. 100623, Jul. 2024, DOI: 10.1016/J.ECMX.2024.100623.

[21] A. A. M. Ahmed *et al.*, "*Hybrid Deep Learning Model for Wave Height Prediction in Australia's Wave Energy Region*," *Appl. Soft Comput.*, Vol. 150, Jan. 2024, DOI: 10.1016/J.ASOC.2023.111003.

[22] Rohitgds, "*Gold Coast City Wave Dataset from 2014-2024*," Kaggle, 2025. Accessed: Jun. 23, 2026. [Online]. Available: https://www.kaggle.com/datasets/rohitgds/gold-coast-city-wave-dataset-from-2014-2024

[23] H. Hewamalage, K. Ackermann, and C. Bergmeir, "*Forecast Evaluation for Data Scientists: Common Pitfalls and Best Practices*," *Data Min. Knowl. Discov.*, Vol. 37, No. 2, pp. 788–832, 2023, DOI: 10.1007/S10618-022-00894-5.

[24] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "*Optuna: A Next-Generation*

ISSN:2302-8149
e-ISSN:2540-9719

Hyperparameter Optimization Framework," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 2623–2631, Jul. 2019, DOI: 10.1145/3292500.3330701.

[25] R. Lunde and C. Shalizi, "*Bootstrapping Generalization Error Bounds for Time Series*," *Sankhya A*, pp. 1–34, Aug. 2026, DOI: 10.1007/S13171-026-00452-X.

[26] S. M. Lundberg and S.-I. Lee, "*A Unified Approach to Interpreting Model Predictions*," *Adv. Neural Inf. Process. Syst.*, Vol. 30, pp. 4765–4774, 2017, DOI: 10.48550/arXiv.1705.07874.