

Pengaruh Hiperparameter dan Variabel Eksogen pada Prediksi Multi-Langkah Kecepatan Angin menggunakan LightGBM

Nina Fitriyati^{*1}, Muhaza Liebenlito², Salma Hasna Tsabitah³

^{1,2,3}*Program Studi Matematika, Fakultas Sains dan Teknologi, UIN Syarif Hidayatullah Jakarta
Jl. Ir. H. Juanda No 95 Ciputat, Tangerang Selatan, Banten, Indonesia*

Korespondensi; Nina Fitriyati, Email: nina.fitriyati@uinjkt.ac.id

Abstrak

Penelitian ini menganalisis pengaruh konfigurasi hiperparameter dan variabel eksogen terhadap prediksi kecepatan angin multi-langkah menggunakan algoritma LightGBM dengan strategi Recursive dan Direct. Hasil penelitian menunjukkan bahwa panjang horizon prediksi merupakan salah satu faktor utama yang mempengaruhi tingkat kesalahan prediksi, dimana nilai kesalahan meningkat secara konsisten dari horizon 1 langkah hingga 30 langkah. Penambahan variabel eksogen pada model berupa suhu minimum dan kelembaban relatif terbukti mampu meningkatkan kinerja model pada seluruh horizon, dengan dampak yang lebih besar pada horizon menengah dan panjang. Selain itu, tuning hiperparameter menunjukkan pengaruh yang bergantung pada horizon, dimana tuning memberikan manfaat yang lebih jelas pada prediksi jangka panjang. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa kinerja prediksi kecepatan angin tidak hanya dipengaruhi oleh pemilihan algoritma, tetapi juga dipengaruhi oleh kombinasi strategi prediksi, variabel eksogen, konfigurasi hiperparameter, dan panjang horizon prediksi.

Kata Kunci: horizon, konfigurasi hiperparameter, pendekatan Direct, pendekatan Recursive.

Abstract

This study analyzes the effect of hyperparameter configuration and exogenous variables on multi-step wind speed prediction using the LightGBM algorithm with Recursive and Direct strategies. The results show that the length of the prediction horizon is one of the main factors affecting the level of prediction error, with the error increasing consistently from 1 step to 30 steps. The addition of exogenous variables to the model, such as minimum temperature and relative humidity, has been shown to improve model performance across all horizons, with a greater impact on medium and long horizons. The addition of exogenous variables to the model, such as minimum temperature and relative humidity, has been shown to improve model performance across all horizons, with a greater impact on medium- and long-horizon forecasts. In addition, hyperparameter tuning exhibits a horizon-dependent effect, with greater benefits for long-term predictions. Overall, the results of this study indicate that wind speed prediction performance is influenced not only by the choice of algorithm but also by the combination of prediction strategy, exogenous variables, hyperparameter configuration, and prediction horizon length.

Keywords: horizon, hyperparameter configuration, Direct approach, Recursive approach.

Pendahuluan

Pengembangan Pembangkit Listrik Tenaga Bayu (PLTB) di Indonesia memiliki peran strategis dalam mendukung peralihan energi terbarukan, khususnya di wilayah timur seperti Nusa Tenggara Barat (NTB) dan Nusa Tenggara Timur (NTT), yang memiliki potensi angin yang relatif stabil dan tinggi sepanjang tahun. Oleh karena itu, prediksi kecepatan angin yang akurat sangat penting untuk memastikan produksi

energi yang optimal, perencanaan operasional yang efektif, manajemen risiko, dan keandalan jaringan. Namun, dinamika kecepatan angin sangatlah berfluktuasi dan nonlinier. Hal ini didorong oleh interaksi atmosfer yang kompleks dan saling ketergantungan antarvariabel meteorologi. Karakteristik tersebut menuntut pengembangan model prediksi yang mampu merepresentasikan struktur nonlinier dan ketergantungan temporal secara memadai guna meningkatkan akurasi prediksi kecepatan angin.

Berbagai pendekatan prediksi deret waktu telah dikembangkan untuk memodelkan perilaku kecepatan angin, mulai dari metode statistik klasik hingga teknik pembelajaran mesin (*machine learning*) modern. Pendekatan statistik seperti Autoregressive Integrated Moving Average (ARIMA) tetap banyak digunakan karena kesederhanaan dan interpretasinya, khususnya untuk data linier dan stasioner [1]. Namun, model ARIMA terbatas dalam menangkap pola nonlinier dan interaksi kompleks antarvariabel meteorologi. Dengan kemajuan pada pembelajaran mesin, model nonlinier seperti Random Forests, Long Short-Term Memory (LSTM), dan algoritma Gradient Boosting mampu menunjukkan kinerja prediksi yang unggul dalam aplikasi prediksi meteorologi dan energi [2][3][4][5]. Penelitian terdahulu menunjukkan bahwa LightGBM berkinerja kompetitif dalam prediksi variabel meteorologi dan energi, termasuk kecepatan angin [6], dan cocok untuk skenario prediksi multi-langkah [7].

Light Gradient Boosting Machine (LightGBM) merupakan salah satu algoritma gradient boosting yang dikenal karena efisiensi komputasi dan kemampuan generalisasinya yang kuat [8]. LightGBM menggunakan strategi pertumbuhan pohon leaf-wise bersama dengan Gradient-based One-Side Sampling (GOSS) dan Exclusive Feature Bundling (EFB), yang memungkinkan pembelajaran yang efisien dari data berdimensi tinggi [9]. Meskipun memiliki kemampuan prediksi yang kuat, kinerja LightGBM sangat sensitif terhadap konfigurasi hyperparameter. Parameter seperti jumlah daun, kedalaman maksimum, laju pembelajaran, fraksi fitur, fraksi bagging, dan iterasi boosting secara langsung memengaruhi kompleksitas model, trade-off bias-varians, perilaku konvergensi, dan kinerja generalisasi. Dalam konteks prediksi multi-langkah, dimana prediksi dihasilkan di beberapa horizon, pengaturan hyperparameter dapat mempengaruhi stabilitas dan konsistensi secara berbeda tergantung pada strategi prediksi yang digunakan. Namun, analisis sistematis tentang bagaimana konfigurasi hyperparameter mempengaruhi kinerja prediksi kecepatan angin multi-langkah masih terbatas, khususnya dalam kerangka kerja LightGBM.

Selain konfigurasi model, penyertaan variabel meteorologi eksogen dapat secara signifikan mempengaruhi kinerja prediksi. Kecepatan angin secara fisik terkait dengan proses termodinamika atmosfer dan dinamika kelembaban. Suhu minimum mencerminkan proses pendinginan nokturnal yang memengaruhi gradien tekanan dan stabilitas lapisan batas atmosfer, sementara kelembaban relatif mewakili kandungan kelembaban atmosfer yang memengaruhi kepadatan udara dan pola sirkulasi lokal [10]. Analisis korelasi empiris menunjukkan bahwa suhu minimum dan kelembaban relatif menunjukkan hubungan yang bermakna dengan kecepatan angin, mendukung pemilihan keduanya sebagai prediktor eksogen. Menggabungkan variabel-variabel ini berpotensi meningkatkan kinerja prediksi. Namun, bukti empiris mengenai bagaimana variabel eksogen berinteraksi dengan tuning hiperparameter dan strategi prediksi multi-langkah masih langka.

Strategi prediksi multi-langkah memainkan peran penting dalam membentuk hasil prediksi. Dua pendekatan utama adalah *Recursive* (Recursive) dan *Direct* (Langsung) [11]. Pendekatan Recursive menggunakan satu model dengan satu langkah ke depan secara iteratif untuk menghasilkan prediksi multi-horizon. Hal ini menyebabkan efisiensi komputasi dan implementasi yang lebih sederhana. Sebaliknya, pendekatan Direct membangun model terpisah untuk setiap horizon prediksi, memungkinkan pembelajaran horizon tertentu yang menyebabkan peningkatan biaya komputasi dan kompleksitas model [12]. Pertimbangan teoritis menunjukkan bahwa strategi ini menggunakan informasi temporal yang berbeda yaitu pendekatan Recursive secara eksplisit memanfaatkan ketergantungan

sekuensial dan pendekatan Direct memperlakukan setiap horizon sebagai tugas pembelajaran terawasi (*supervised learning*) yang independen [13]. Akibatnya, konfigurasi model dan penyertaan fitur dapat menghasilkan efek yang berbeda pada setiap strategi.

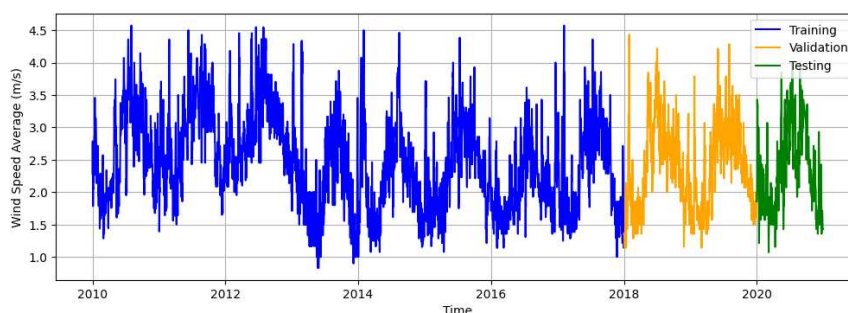
Meskipun banyak penelitian hanya membandingkan model berdasarkan nilai akurasi akhir, beberapa studi mulai melihat bagaimana pemilihan hiperparameter dan penambahan variabel eksogen dapat mempengaruhi kinerja prediksi multi-langkah pada suatu algoritma tertentu. Dalam praktiknya, peningkatan akurasi yang tampak kecil sekalipun dapat berdampak pada keputusan yang lebih baik, seperti penjadwalan operasi yang lebih tepat, perencanaan distribusi energi yang lebih efisien, dan peningkatan keandalan sistem. Selain itu, memahami seberapa sensitif kinerja model terhadap tuning hiperparameter dan penggunaan variabel atmosfer memberikan wawasan penting untuk pengembangan dan penerapan model prediksi yang lebih efektif. Oleh karena itu, artikel ini secara sistematis membahas efek konfigurasi hiperparameter dan variabel eksogen pada prediksi kecepatan angin multi-langkah menggunakan LightGBM untuk strategi Recursive dan Direct. Penelitian ini menekankan evaluasi eksperimental terstruktur untuk menilai bagaimana keputusan tuning hyperparameter dan penyertaan fitur eksogen mempengaruhi kinerja prediksi di berbagai horizon. Selain itu, penelitian ini secara eksplisit menganalisis mekanisme propagasi error pada prediksi multi-langkah, khususnya pada pendekatan Recursive dan Direct. Analisis ini memberikan pemahaman yang lebih mendalam mengenai bagaimana ketidakpastian prediksi berkembang dari horizon jangka pendek ke jangka panjang, yang masih relatif jarang dibahas dalam literatur prediksi kecepatan angin berbasis LightGBM.

Walaupun tuning hiperparameter dan penambahan variabel eksogen tidak selalu menghasilkan peningkatan akurasi yang besar, penelitian ini tetap menunjukkan bagaimana model LightGBM bereaksi terhadap perubahan konfigurasi tersebut. Hasil penelitian ini membantu memahami kapan dan dalam kondisi apa pengaturan model memberikan manfaat. Dengan demikian, penelitian ini dapat menjadi panduan praktis dalam memilih konfigurasi model yang sesuai untuk prediksi kecepatan angin sebagai energi terbarukan.

Bahan dan Metode

Data and Pra-pemrosesan

Studi ini menggunakan data kecepatan angin rata-rata harian yang dikumpulkan dari beberapa stasiun pengamatan di NTB dan NTT, Indonesia, tanggal 1 Januari 2010 hingga 31 Desember 2020. Data dari semua stasiun dikumpulkan secara temporal untuk mewakili kondisi angin regional. Penggunaan data agregat bertujuan untuk memfokuskan analisis pada dinamika regional, bukan pada karakteristik spesifik per stasiun. Gambar 1 menggambarkan rata-rata kecepatan angin harian (m/s) di NTB dan NTT, dengan pembagian kronologis menjadi periode pelatihan (*training*), validasi (*validation*), dan pengujian (*testing*) untuk menjaga ketergantungan temporal dan mencegah kebocoran informasi pada eksperimen.



Gambar 1. Deret Waktu Kecepatan Angin Rata-Rata Harian (m/s) dengan Pembagian Kronologis menjadi Dataset Pelatihan, Validasi, dan Pengujian

Selain variabel target kecepatan angin rata-rata (ff_avg (m/s)), beberapa variabel meteorologi lainnya digunakan sebagai variabel eksogen, termasuk suhu minimum (Tn (°C)), suhu maksimum (Tx (°C)), suhu rata-rata (Tavg (°C)); kelembaban relatif (RH_avg (%)); curah hujan (RR (mm)); arah angin saat kecepatan angin maksimum (ddd_x (°)); dan kecepatan angin maksimum (ff_x (m/s)). Proses pra-pemrosesan meliputi penanganan data yang hilang, pengecekan konsistensi data, dan normalisasi variabel numerik untuk memastikan stailitas pada pelatihan model. Analisis awal struktur data dilakukan untuk mengidentifikasi tren, musiman, dan ketergantungan temporal, yang berfungsi sebagai dasar untuk memilih strategi prediksi multi-langkah.

Pemilihan Fitur Berdasarkan Analisis Korelasi

Pemilihan variabel eksogen dalam penelitian ini dilakukan melalui prosedur seleksi yang berbasis korelasi. Ini dilakukan untuk memastikan relevansi statistik dan interpretasi fisik. Oleh karena itu, korelasi Pearson dilakukan sebagai tahap awal seleksi fitur karena interpretabilitasnya yang tinggi dan kesesuaiannya dalam menangkap hubungan fisik antar variable meteorologi.

Misalkan y_t adalah kecepatan angin rata-rata (ff_avg) pada waktu t , dan $x_{j,t}$ adalah kandidat variabel meteorologi. Asosiasi linier dihitung menggunakan koefisien korelasi Pearson berikut:

$$\rho_{y,x_j} = \frac{\sum_{t=1}^T (y_t - \bar{y})(x_{j,t} - \bar{x}_j)}{\sqrt{\sum_{t=1}^T (y_t - \bar{y})^2} \sqrt{\sum_{t=1}^T (x_{j,t} - \bar{x}_j)^2}},$$

dengan T adalah jumlah total pengamatan, $\bar{y}$ dan $\bar{x}_j$ adalah rata-rata sampel. Variabel dengan korelasi sedang dan bermakna secara statistik dengan kecepatan angin dianggap sebagai kandidat untuk dimasukkan sebagai input eksogen. Matriks korelasi pada Gambar 2 menunjukkan bahwa suhu minimum (Tn) dan kelembaban relatif (RH_avg) memiliki korelasi sedang dengan kecepatan angin rata-rata. Meskipun kecepatan angin maksimum (ff_x) menunjukkan korelasi kuat dengan kecepatan angin rata-rata, namun variabel ini dikecualikan untuk menghindari redundansi informasi dan potensi kebocoran data, karena ff_x mewakili statistik ekstrem yang berasal dari dinamika angin harian yang sama.



Gambar 2. Matriks Korelasi Antarvariabel

Dengan demikian, kumpulan fitur akhir didefinisikan sebagai:

Model Dasar (Univariat): $Z_t = \{y_{t-1}, y_{t-2}, \dots, y_{t-p}\}$.

Model multivariate dengan variable eksogen: $Z_t = \{y_{t-1}, \dots, y_{t-p}, Tn_t, RH_t\}$.

Desain ini memungkinkan evaluasi langsung terhadap kontribusi tambahan dari variabel eksogen di berbagai horizon waktu prediksi dan strategi bertahap.

Perumusan Masalah Prediksi Multi-langkah

Masalah prediksi angin dirumuskan sebagai masalah prediksi deret waktu multi-langkah, dengan tujuan utama memprediksi kecepatan angin pada waktu $t + h$ berdasarkan informasi historis hingga waktu t . Untuk setiap horizon prediksi, kualitas prediksi dievaluasi menggunakan metrik kesalahan yang mengukur perbedaan antara nilai aktual dan nilai prediksi. Studi ini tidak hanya berfokus pada kesalahan absolut tetapi juga pada evolusinya sebagai fungsi dari horizon prediksi.

Pendekatan Prediksi Multi-langkah

Studi ini menggunakan dua pendekatan prediksi multi-langkah yaitu Recursive dan Direct. Kedua pendekatan ini berbeda secara prosedural dan memiliki implikasi matematis yang berbeda untuk perilaku pertumbuhan kesalahan seiring bertambahnya horizon prediksi. Secara umum, kesalahan prediksi selama horizon h didefinisikan sebagai $e_{t+h} = y_{t+h} - \hat{y}_{t+h}$, dengan kualitas prediksi dievaluasi melalui ekspektasi kesalahan kuadrat $E(h) = \mathbb{E}[(y_{t+h} - \hat{y}_{t+h})^2]$ yang pada praktiknya dihitung menggunakan Mean Squared Error (MSE).

Pendekatan Recursive

Pada pendekatan ini, model dilatih untuk memprediksi satu langkah ke depan; prediksi yang dihasilkan kemudian digunakan secara iteratif sebagai input untuk memprediksi langkah selanjutnya hingga horizon yang diinginkan tercapai [14]. Secara matematis, prediksi pada horizon h dapat ditulis secara recursive sebagai $\hat{y}_{t+h} = f(\hat{y}_{t+h-1}, \hat{y}_{t+h-2}, \dots, \hat{y}_{t+1})$, di mana fungsi $f(\cdot)$ menyatakan model prediktor satu langkah ke depan yang memprediksi kecepatan angin pada langkah $t + h$ berdasarkan prediksi sebelumnya.

Pada kerangka analitis sederhana, prediksi pada setiap langkah dihitung secara iteratif, dengan model satu Langkah yang digunakan untuk memprediksi langkah selanjutnya. Error pada setiap prediksi $\hat{y}_{t+k}$ pada waktu $t + k$ akan mempengaruhi prediksi pada langkah-langkah berikutnya, dimana error pada $t + 1$ akan dipropagasikan ke langkah-langkah selanjutnya $t + 2, t + 3, \dots, t + h$. Secara matematis, prediksi pada horizon $t + h$ dapat ditulis sebagai $\hat{y}_{t+h} = f(\hat{y}_{t+h-1}, \hat{y}_{t+h-2}, \dots, \hat{y}_{t+1}) + \sum_{k=1}^h \alpha_k \epsilon_{t+k}$, dimana ϵ_{t+k} adalah error pada langkah $t + k$ dan α_k menggambarkan sensitivitas model terhadap error pada langkah tersebut.

Pendekatan Direct

Sebaliknya, pada pendekatan Direct, model prediksi dibangun secara terpisah untuk setiap horizon prediksi h , sehingga nilai pada waktu $t + h$ diprediksi langsung dari data historis hingga waktu t tanpa menggunakan hasil prediksi sebelumnya [15]. Secara matematis, pendekatan ini dapat dinyatakan sebagai $\hat{y}_{t+h} = f_h(y_t, y_{t-1}, \dots)$ dengan $f_h(\cdot)$ adalah fungsi prediksi spesifik untuk horizon h . Pendekatan Direct menghindari penyebaran kesalahan antarlangkah namun meningkatkan kompleksitas pemodelan dan mengurangi jumlah data efektif untuk melatih setiap model horizon. Dalam konteks ini, ekspektasi kesalahan kuadrat ditulis melalui dekomposisi bias-varians sebagai $E(h) = \text{Bias}^2(h) + \text{Var}(h)$.

Seiring dengan semakin panjangnya horizon prediksi, kompleksitas pemetaan antara data historis dan target masa depan cenderung meningkat, sehingga komponen varians $\text{Var}(h)$ menjadi semakin dominan [16]. Kondisi ini menyebabkan peningkatan kesalahan yang lebih tajam pada horizon prediksi yang panjang.

Dengan demikian, perbedaan mendasar antara pendekatan Recursive dan Direct tidak hanya terletak pada mekanisme pembangkitan prediksi namun juga pada bagaimana masing-masing membentuk dinamika kesalahan selama horizon prediksi.

Model LightGBM

Algoritma LightGBM digunakan sebagai model dasar dalam penelitian ini. LightGBM adalah algoritma ensemble berbasis pohon keputusan yang mengoptimalkan fungsi tujuan melalui penguatan gradien [9]. Salah satu keunggulan LightGBM adalah efisiensi komputasi dan kemampuannya untuk menangkap hubungan nonlinier antarvariabel. Model ini dilatih menggunakan data kecepatan angin historis dan variabel meteorologi tambahan sebagai variabel eksogen dengan skenario tertentu untuk mengevaluasi pengaruh informasi tambahan ini terhadap akurasi dan stabilitas prediksi.

Skema Tuning Hiperparameter

Mengingat bahwa kinerja LightGBM sensitif terhadap konfigurasi, tuning hiperparameter dilakukan untuk menilai pengaruhnya pada strategi multi-langkah yang berbeda. Dua konfigurasi yang dibandingkan adalah (1) hiperparameter default, dan (2) hiperparameter yang dioptimalkan melalui optimasi Bayesian. Parameter yang disetel adalah sebagai berikut:

Tabel 1. Rentang Pencarian Hiperparameter pada Proses Tuning Model LightGBM

Hiperparameter	Rentang Pencarian
num_leaves	[20, 150]
max_depth	[3, 15]
learning_rate	[0.01, 0.3]
feature_fraction	[0.6, 1.0]
bagging_fraction	[0.6, 1.0]
n_estimators	[100, 1000]

Misalkan θ adalah vektor hiperparameter dengan tujuan berikut:

$$\theta^* = \arg \min_{\theta \in \Theta} \mathcal{L}(\theta),$$

dimana $\mathcal{L}(\theta)$ adalah kesalahan kuadrat rata-rata validasi. Validasi silang deret waktu (*Time Series Cross Validation*, TSCV) diaplikasikan menggunakan skema *rolling-origin* untuk mempertahankan urutan temporal. Untuk setiap konfigurasi kandidat berlaku (1) latih pada jendela pelatihan, (2) validasi pada jendela validasi maju (*forward validation window*), dan (3) hitung rata-rata kerugian validasi. Prosedur optimasi secara iteratif memperbarui model pengganti (Proses Gaussian) untuk secara efisien mengeksplorasi ruang parameter. Pendekatan ini memungkinkan analisis lebih lanjut tentang hubungan antara kompleksitas model, stabilitas prediksi, dan generalisasi [17].

Desain simulasi lengkap terdiri dari:

1. Dua strategi multi-langkah (Recursive, Direct)
2. Tiga horizon (1, 7, 30)
3. Dua set fitur (dengan dan tanpa variabel eksogen)
4. Dua pengaturan hiperparameter (default vs dioptimalkan)

Total konfigurasi adalah $2 \times 3 \times 2 \times 2 = 24$ skenario. Desain terstruktur ini memungkinkan evaluasi sistematis terhadap pengaruh horizon, pengaruh variabel eksogen, pengaruh hiperparameter, dan interaksi strategi-konfigurasi.

Strategi Pelatihan dan Evaluasi

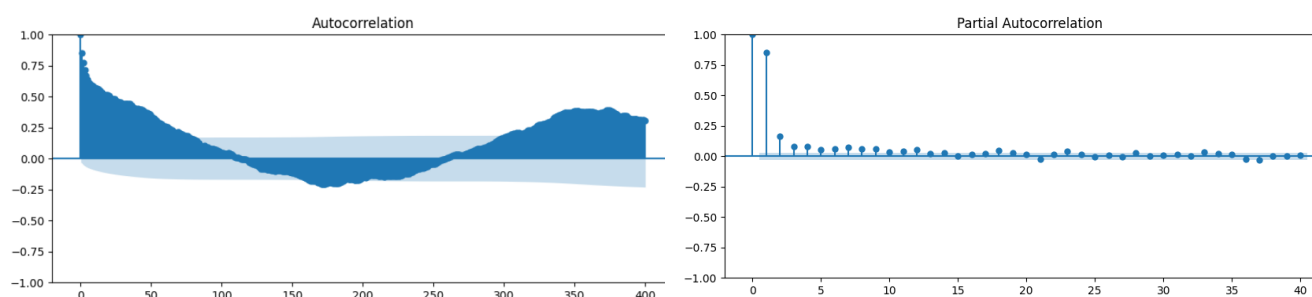
Data dibagi menjadi set pelatihan, validasi, dan pengujian, dengan mempertahankan urutan waktu untuk menghindari kebocoran data. Pelatihan model dilakukan pada beberapa horizon prediksi yang berbeda untuk kedua pendekatan multi-langkah. Evaluasi kinerja dilakukan menggunakan Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), dan Mean Squared Error (MSE).

Hasil dan Pembahasan

Karakteristik Data dan Analisis Eksploratori

Tahap awal analisis data deret waktu melibatkan pembersihan dan eksplorasi data untuk meningkatkan akurasi model, mencegah bias seperti overfitting, dan meminimalkan dampak noise. Karena data berasal dari 14 stasiun di 13 kota/kabupaten di NTB dan NTT, terdapat lebih dari satu baris data per hari, sehingga agregasi dilakukan menggunakan rata-rata harian. Hal ini mengurangi data dari 51.468 menjadi 4.005 baris. Setelah disesuaikan dengan frekuensi harian, 13 hari yang hilang diidentifikasi dan diinterpolasi, menghasilkan dataset bersih sebanyak 4.018 baris dan 9 fitur, siap untuk pemodelan.

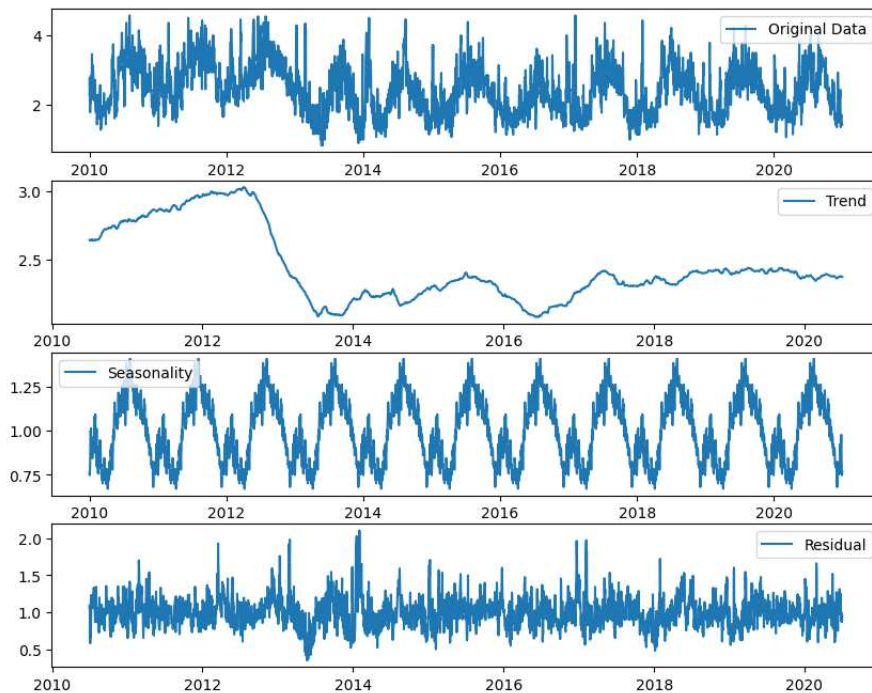
Gambar 3 menyajikan plot ACF dan PACF. ACF menunjukkan autokorelasi tinggi pada lag awal, menurun secara bertahap, berubah tanda sekitar lag 100, dan meningkat lagi di dekat lag 300. Hal ini menunjukkan kemungkinan perilaku siklik dan ketergantungan jangka panjang. PACF hanya menunjukkan lonjakan signifikan pada lag 1, menunjukkan bahwa nilai saat ini terutama dipengaruhi oleh nilai pada hari sebelumnya, dengan efek langsung yang lebih lemah dari lag yang lebih panjang.



Gambar 3. Plot ACF dan PACF Kecepatan Angin Rata-rata

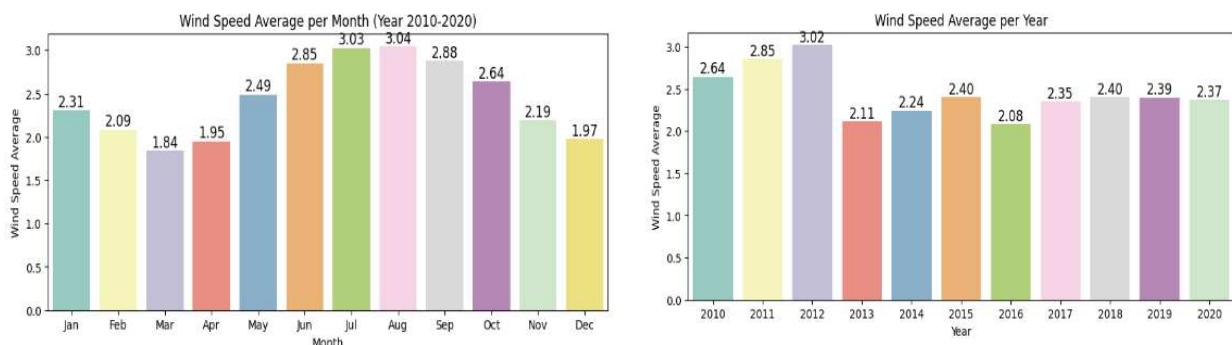
Gambar 4 menunjukkan hasil dari penggunaan dekomposisi musiman pada *library statsmodels.tsa.seasonal* di Python dengan periode 365 hari untuk mengidentifikasi pola musiman.

Gambar ini menunjukkan pola yang berulang setiap tahun. Ini menunjukkan secara alami data memiliki pola musiman.

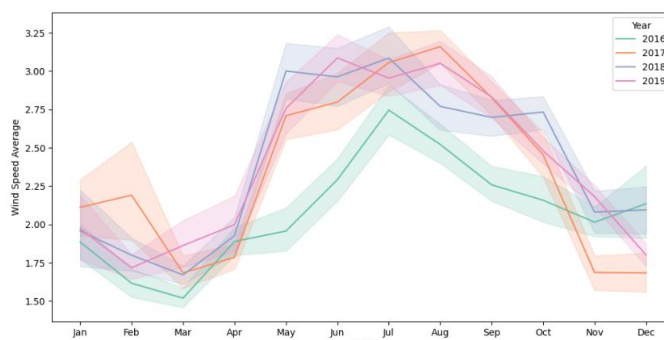


Gambar 4. Plot Dekomposisi Musiman Kecepatan Angin Rata-rata

Setelah mengekstrak kolom tanggal menjadi bulan dan tahun, plot distribusi pada Gambar 5 menunjukkan bahwa selama tahun 2010–2020 kecepatan angin rata-rata tertinggi di NTB dan NTT terjadi pada bulan Agustus (3,04 m/s) dan Juli (3,03 m/s), sedangkan yang terendah terjadi pada bulan Maret (1,84 m/s). Hal ini mencerminkan musim kering (April–September) dan musim hujan (Oktober–Maret), meskipun pergeseran musiman mungkin dipengaruhi oleh perubahan iklim. Secara tahunan, kecepatan angin tetap relatif stabil sekitar 2,4 m/s, dengan nilai tertinggi pada tahun 2012. Pola musiman konsisten, menunjukkan kecepatan yang lebih tinggi pada bulan Juni–September dan lebih rendah pada bulan Desember–Mei, seperti yang diilustrasikan pada Gambar 6.



Gambar 5. Plot Distribusi Kecepatan Angin Rata-rata berdasarkan Bulan dan Tahun



Gambar 6. Kecepatan Angin Rata-Rata Bulanan (m/s) untuk Periode 2016–2019

Gambar 6 menunjukkan pola bulanan yang konsisten, dengan kecepatan angin yang lebih rendah di awal tahun, meningkat menjelang pertengahan tahun, mencapai puncaknya pada bulan Juni–Agustus, dan menurun setelahnya. Struktur yang berulang ini menunjukkan musim tahunan yang jelas dan stabil. Secara keseluruhan, rangkaian data tersebut menunjukkan karakteristik nonlinier, musiman, dan bergantung pada waktu, yang mendukung penggunaan algoritma LightGBM dengan pendekatan prediksi bertahap untuk prediksi kecepatan angin yang akurat dengan berbagai strategi dan konfigurasi.

Hasil Prediksi Multi-langkah Menggunakan LightGBM

Penelitian ini merancang kerangka kerja simulasi yang komprehensif untuk mengevaluasi kinerja model LightGBM dengan berbagai konfigurasi. Dua strategi prediksi multi-langkah (recursive dan direct). Setiap pendekatan diuji pada tiga horizon prediksi yaitu 1 langkah (harian), 7 langkah (mingguan), dan 30 langkah (bulanan) untuk mengevaluasi efek horizon, efek variabel eksogen, efek hiperparameter dan efek interaksi ketiganya pada akurasi prediksi.

Untuk setiap horizon, delapan skenario dievaluasi dengan menggabungkan: (1) strategi recursive vs. direct, (2) model dengan dan tanpa variabel eksogen (suhu dan kelembaban), dan (3) parameter default vs. hiperparameter yang dioptimalkan Bayesian menggunakan TSCV. Desain ini memungkinkan perbandingan sistematis strategi multi-langkah, kompleksitas model, dan input meteorologi tambahan. Hasil kinerja model dinilai menggunakan MAE, MAPE, dan MSE di semua skenario tertera pada Tabel 2.

Tabel 2. Kinerja Prediksi Multi-langkah Kecepatan Angin menggunakan Model LightGBM dengan Pendekatan Recursive dan Direct

No	Model	Horizon prediksi (Langkah)	Variabel Eksogen	Tuning Hyperparameter	MAE	MAPE	MSE
1	Recursif	1	Tidak	Tidak	0.027351	0.011871	0.001197
2	Recursif	1	Ya	Tidak	0.026045	0.011366	0.001086
3	Recursif	1	Tidak	Ya	0.030456	0.013229	0.001448
4	Recursif	1	Ya	Ya	0.027994	0.012184	0.001244
5	Direct	1	Tidak	Tidak	0.027351	0.011871	0.001197
6	Direct	1	Ya	Tidak	0.026045	0.011366	0.001086
7	Direct	1	Tidak	Ya	0.030456	0.013229	0.001448
8	Direct	1	Ya	Ya	0.027994	0.012184	0.001244
9	Recursif	7	Tidak	Tidak	0.091927	0.039286	0.014858
10	Recursif	7	Ya	Tidak	0.082586	0.035968	0.012659
11	Recursif	7	Tidak	Ya	0.096394	0.041634	0.018311
12	Recursif	7	Ya	Ya	0.085775	0.036257	0.014078
13	Direct	7	Tidak	Tidak	0.107164	0.046191	0.021613
14	Direct	7	Ya	Tidak	0.095305	0.040781	0.016663
15	Direct	7	Tidak	Ya	0.117715	0.050383	0.026125

16	Direct	7	Ya	Ya	0.104663	0.044696	0.020853
17	Rekursif	30	Tidak	Tidak	0.334765	0.141861	0.195245
18	Rekursif	30	Ya	Tidak	0.279966	0.117741	0.138232
19	Rekursif	30	Tidak	Ya	0.295598	0.120772	0.154577
20	Rekursif	30	Ya	Ya	0.232515	0.100612	0.099987
21	Direct	30	Tidak	Tidak	0.301913	0.128476	0.149424
22	Direct	30	Ya	Tidak	0.273689	0.116388	0.125068
23	Direct	30	Tidak	Ya	0.234409	0.097962	0.091234
24	Direct	30	Ya	Ya	0.233403	0.098137	0.092288

Pengaruh Horizon

Dimensi analisis pertama membahas pengaruh panjang horizon prediksi terhadap kinerja model. Pada semua konfigurasi yang diuji — baik dari segi strategi, penggunaan variabel eksogen, maupun pengaturan hyperparameter — terlihat pola yang konsisten bahwa nilai kesalahan (MAE, MAPE, dan MSE) meningkat seiring dengan bertambahnya horizon prediksi, dari 1 langkah menjadi 7 langkah, hingga 30 langkah ke depan.

Sebagai contoh, pada pendekatan recursive tanpa variabel eksogen dan tanpa tuning MAE 1 langkah adalah 0,027351, MAE 7 langkah adalah 0,091927, dan MAE 30 langkah adalah 0,334765. Pola ini menegaskan bahwa ketidakpastian prediktif meningkat seiring dengan jarak temporal dari data yang diamati. Besarnya pertumbuhan kesalahan cukup besar antara horizon jangka menengah (7 langkah) dan jangka panjang (30 langkah). Hal ini menunjukkan bahwa prediksi kecepatan angin jangka panjang tetap menjadi tugas yang menantang bahkan untuk model ensemble nonlinier seperti LightGBM. Pola peningkatan ini berlaku secara konsisten di seluruh strategi Recursive dan Direct, menunjukkan bahwa horizon prediksi adalah penentu utama penurunan akurasi. Oleh karena itu, pilihan konfigurasi model harus diinterpretasikan relatif terhadap panjang horizon.

Pengaruh Variabel Eksogen

Untuk mengukur kontribusi variabel eksogen yang terdiri dari suhu minimum dan kelembaban relatif, kami menghitung persentase peningkatan MAE sebagai berikut:

$$\text{Peningkatan (\%)} = \frac{MAE_{\text{without exog}} - MAE_{\text{with exog}}}{MAE_{\text{without exog}}} \times 100.$$

Tabel 3. Pengaruh Variabel Eksogen terhadap MAE di Seluruh Horizon Prediksi (Tanpa Tuning Hiperparameter)

Horizon (Langkah)	Strategi	MAE tanpa Variabel Eksogen	MAE dengan Variabel Eksogen	Peningkatan (%)
1	Recursive	0.027351	0.026045	4.78%
7	Recursive	0.091927	0.082586	10.16%
7	Direct	0.107164	0.095305	11.07%
30	Recursive	0.334765	0.279966	16.36%
30	Direct	0.301913	0.273689	9.35%

Tabel 3 menunjukkan bahwa penambahan variabel eksogen secara konsisten menurunkan nilai MAE pada seluruh horizon peramalan, baik pada strategi Rekursif maupun Direct (tanpa tuning hyperparameter). Pada horizon 1-langkah, penurunan error relatif kecil (4,78%), yang menunjukkan bahwa prediksi jangka pendek lebih banyak dipengaruhi oleh ketergantungan temporal jangka pendek dibandingkan faktor eksternal. Namun, pada horizon 7-langkah, dampak variabel eksogen menjadi lebih signifikan, dengan peningkatan kinerja sekitar 10–11% untuk kedua strategi. Pengaruh terbesar terlihat

pada horizon 30-langkah, khususnya pada strategi Recursif yang mengalami penurunan MAE sebesar 16,36%, sementara strategi Direct sebesar 9,35%. Pola ini menunjukkan bahwa kontribusi variabel atmosfer semakin penting seiring bertambahnya panjang horizon prediksi, karena informasi historis saja menjadi kurang memadai untuk menjelaskan dinamika jangka panjang kecepatan angin.

Analisis Konfigurasi Hiperparameter di Seluruh Horizon Prediksi

Untuk lebih memahami peran kompleksitas model dalam peramalan multi-langkah, pada bagian ini akan dikaji perilaku struktural hiperparameter LightGBM yang dioptimalkan untuk seluruh horizon dan strategi prediksi. Dengan menganalisis perubahan *n_estimators*, *max_depth*, *learning_rate*, *reg_alpha*, dan *reg_lambda*, investigasi ini bertujuan untuk mengungkapkan apakah kesulitan prediksi tercermin pada penyesuaian konfigurasi sistematis. Analisis tersebut memberikan wawasan tentang sensitivitas desain model terhadap panjang horizon dan berkontribusi pada pemahaman yang lebih berprinsip tentang pengaruh hiperparameter dalam prediksi kecepatan angin multi-langkah. Tabel 4 menunjukkan hiperparameter yang dioptimalkan di seluruh strategi dan horizon.

Hasil tuning hiperparameter menunjukkan bahwa konfigurasi model berubah mengikuti panjang horizon prediksi. Pada prediksi jangka pendek (1-langkah), baik strategi Recursif maupun Direct memilih struktur model yang relatif sederhana, seperti *max_depth* yang tidak terlalu dalam, *n_estimators* sedang, *learning_rate* lebih tinggi, dan penggunaan lag pendek (1–7). Hal ini menunjukkan bahwa prediksi jangka pendek terutama dipengaruhi oleh informasi terbaru, sehingga tidak membutuhkan memori historis yang panjang maupun kompleksitas model yang tinggi.

Ketika horizon diperpanjang menjadi 7 dan 30 langkah, konfigurasi optimal berubah secara sistematis. Model mulai menggunakan lag yang lebih panjang (hingga 180 hari), *learning_rate* yang lebih kecil, serta regularisasi yang lebih kuat. Ini menunjukkan bahwa prediksi jangka panjang memerlukan informasi historis yang lebih luas dan pengendalian model yang lebih stabil untuk menghindari *overfitting*. Dengan kata lain, semakin panjang horizon, semakin tinggi kebutuhan akan kestabilan struktur model.

Perbedaan antara strategi Recursif dan Direct paling terlihat pada horizon 7 langkah. Model Direct cenderung menggunakan lebih banyak pohon dan *learning_rate* yang lebih tinggi, yang menunjukkan kebutuhan fleksibilitas lebih besar untuk memodelkan hubungan jangka menengah. Sementara itu, strategi Recursif tetap menggunakan konfigurasi yang lebih konservatif karena memanfaatkan proses prediksi bertahap. Penambahan variabel eksogen tidak banyak mengubah struktur hiperparameter, namun meningkatkan akurasi dengan menambah informasi tambahan. Secara keseluruhan, hasil ini menegaskan bahwa tuning hiperparameter perlu disesuaikan dengan panjang horizon prediksi agar kinerja model optimal.

Tabel 4. Hiperparameter yang Dioptimalkan untuk Seluruh Strategi dan Horizon

Strategi	Variabel Eksogen	Horizon (Langkah)	Langkah	<i>n_estimators</i>	<i>max_depth</i>	<i>learning_rate</i>	<i>reg_alpha</i>	<i>reg_lambda</i>
Recursive	Tidak	1	1–7	600	6	0.298578	0.530396	0.056119
Recursive	Tidak	7	1–180	700	5	0.184929	0.621441	0.620958
Recursive	Tidak	30	1–180	700	5	0.184929	0.621441	0.620958
Recursive	Ya	1	1–7	600	6	0.298578	0.530396	0.056119
Recursive	Ya	7	1–7	600	6	0.298578	0.530396	0.056119
Recursive	Ya	30	1–180	700	5	0.184929	0.621441	0.620958
Direct	Tidak	1	1–7	600	6	0.298578	0.530396	0.056119
Direct	Tidak	7	1–90	800	6	0.330174	0.378395	0.017203
Direct	Tidak	30	1–180	700	5	0.184929	0.621441	0.620958

Strategi	Variabel Eksogen	Horizon (Langkah)	Langkah	n_estimators	max_depth	learning_rate	reg_alpha	reg_lambda
Direct	Ya	1	1–7	600	6	0.298578	0.530396	0.056119
Direct	Ya	7	1–90	800	6	0.330174	0.378395	0.017203
Direct	Ya	30	1–180	700	5	0.184929	0.621441	0.620958

Pengaruh Hiperparameter

Untuk mengevaluasi pengaruh tuning hiperparameter terhadap kinerja model, dilakukan perbandingan nilai MAE antara konfigurasi default dan konfigurasi hasil tuning pada setiap horizon prediksi. Analisis ini bertujuan untuk mengidentifikasi apakah tuning memberikan peningkatan kinerja yang konsisten atau justru bersifat bergantung pada panjang horizon. Hasil perbandingan tersebut disajikan pada Tabel 5.

Tabel 5. Pengaruh Tuning Hiperparameter pada MAE terhadap Horizon Prediksi

Horizon (Langkah)	Strategi	MAE Default	MAE Tuned	Peningkatan (%)	Pengaruh Tuning
1	Recursive	0,027351	0,030456	–11,35%	Menurun
7	Recursive	0,091927	0,096394	–4,86%	Menurun
30	Recursive	0,334765	0,295598	11,70%	Meningkat
30	Direct	0,301913	0,234409	22,35%	Meningkat

Tabel 5 menunjukkan bahwa efek tuning hiperparameter bersifat bergantung pada panjang horizon prediksi. Pada horizon pendek (1 dan 7 langkah) dengan strategi Recursif, tuning justru meningkatkan nilai MAE masing-masing sebesar –11,35% dan –4,86%, yang berarti kinerja model menurun dibandingkan konfigurasi default. Hal ini mengindikasikan bahwa pada prediksi jangka pendek dan menengah, kompleksitas tambahan dari tuning tidak memberikan manfaat signifikan dan bahkan berpotensi menyebabkan overfitting ringan. Sebaliknya, pada horizon 30 langkah, tuning memberikan dampak positif yang cukup besar. Pada strategi Recursif, MAE menurun sebesar 11,70%, sementara pada strategi Direct penurunannya lebih signifikan yaitu 22,35%. Temuan ini menunjukkan bahwa tuning hiperparameter menjadi lebih efektif pada horizon prediksi yang panjang, ketika kompleksitas masalah meningkat dan model membutuhkan kapasitas serta regularisasi yang lebih teroptimasi.

Pengaruh Interaksi Strategi, Eksogen, dan Tuning Hiperparameter

Selain melihat pengaruh masing-masing faktor secara terpisah, penting juga untuk memahami bagaimana strategi peramalan, penggunaan variabel eksogen, dan tuning hiperparameter saling berinteraksi dalam mempengaruhi kinerja model. Analisis interaksi ini dilakukan dengan membandingkan konfigurasi model terbaik pada setiap horizon prediksi. Tujuannya adalah untuk mengidentifikasi kombinasi konfigurasi yang memberikan nilai MAE paling rendah serta melihat bagaimana peran masing-masing komponen berubah seiring dengan bertambahnya horizon prediksi. Ringkasan hasil analisis interaksi tersebut ditampilkan pada Tabel 6.

Tabel 6. Pengaruh Interaksi Strategy, Exogenous, dan Hyperparameter pada MAE vs Horizon

Horizon (Langkah)	Konfigurasi Baseline (Default, Tanpa Eksogen, Tanpa Tuning)	MAE Baseline	Konfigurasi Terbaik	MAE Terbaik	Peningkatan (%)
1	Recursif	0,027351	Dengan Eksogen, Tanpa Tuning (Recursif/Direct)	0,026045	4,78%
7	Recursif	0,091927	Recursif + Eksogen	0,082586	10,16%

Horizon (Langkah)	Konfigurasi Baseline (Default, Tanpa Eksogen, Tanpa Tuning)	MAE Baseline	Konfigurasi Terbaik	MAE Terbaik	Peningkatan (%)
30	Rekursif	0,334765	+ Tanpa Tuning Rekursif + Eksogen + Tuning	0,232515	30,54%

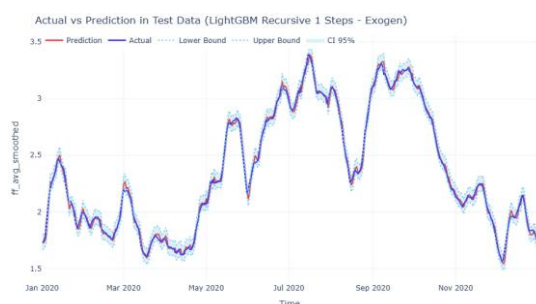
Catatan:

Improvement (%) dihitung relatif terhadap baseline Recursive default tanpa eksogen dan tanpa tuning.

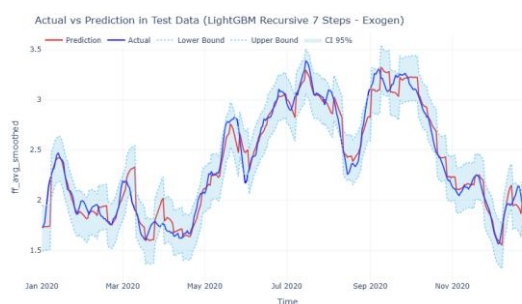
Tabel 6 menunjukkan bahwa kombinasi strategi prediksi, penggunaan variabel eksogen, dan tuning hiperparameter memberikan pengaruh yang berbeda terhadap kinerja model pada setiap horizon prediksi. Pada horizon 1 langkah, konfigurasi terbaik diperoleh dengan menambahkan variabel eksogen tanpa melakukan tuning hiperparameter, baik pada strategi Recursive maupun Direct, yang menghasilkan penurunan MAE sebesar 4,78% dibandingkan baseline. Pada horizon 7 langkah, konfigurasi terbaik diperoleh dengan strategi Recursive yang dikombinasikan dengan variabel eksogen tanpa tuning, dengan peningkatan kinerja sebesar 10,16%. Sementara itu, pada horizon 30 langkah, kombinasi strategi Recursive, variabel eksogen, dan tuning hiperparameter memberikan hasil terbaik dengan penurunan MAE sebesar 30,54%. Hasil ini menunjukkan bahwa konfigurasi model yang optimal bergantung pada panjang horizon prediksi, di mana variabel eksogen berkontribusi pada semua horizon, sedangkan tuning hiperparameter menjadi lebih penting pada horizon yang lebih panjang.

Visualisasi Nilai Aktual dan Prediksi

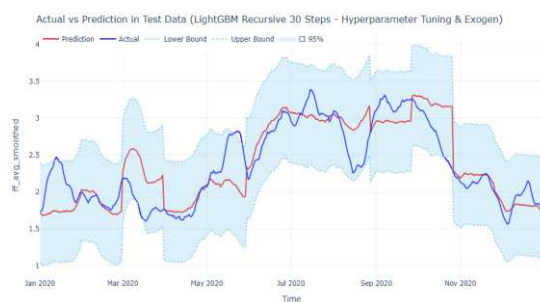
Untuk memberikan gambaran yang lebih jelas mengenai kemampuan model dalam mengikuti pola data sebenarnya, dilakukan perbandingan antara nilai aktual dan nilai prediksi menggunakan tiga konfigurasi model terbaik yang diperoleh dari hasil evaluasi sebelumnya. Perbandingan ini ditampilkan dalam bentuk grafik agar dapat menunjukkan sejauh mana model mampu merepresentasikan dinamika perubahan kecepatan angin pada setiap horizon prediksi. Grafik berikut memperlihatkan kesesuaian antara nilai aktual dan nilai hasil prediksi dari masing-masing model terbaik.



(a) Horizon 1-langkah + Eksogen



(b) Horizon 7-langkah + Eksogen



(c) Horizon 30-langkah + Eksogen + Tuning Hiperparameter

Gambar 7. Perbandingan Nilai Aktual dan Prediksi Kecepatan Angin Menggunakan Model Recursive LightGBM pada Berbagai Horizon Prediksi

Gambar 7 menunjukkan perbandingan antara nilai aktual dan nilai prediksi kecepatan angin menggunakan model Recursive LightGBM pada tiga horizon prediksi yang berbeda, sekaligus menampilkan interval kepercayaan 95%. Pada horizon 1 langkah (a), nilai prediksi sangat mendekati nilai aktual dan mampu mengikuti pola fluktuasi kecepatan angin dengan baik, sementara pita interval kepercayaan relatif sempit. Hal ini menunjukkan tingkat ketidakpastian prediksi yang rendah pada prediksi jangka pendek. Pada horizon 7 langkah (b), model masih mampu mengikuti tren utama data aktual, namun interval kepercayaan mulai melebar dan beberapa perbedaan antara nilai aktual dan prediksi mulai terlihat, terutama pada titik puncak dan lembah. Kondisi ini menunjukkan bahwa ketidakpastian prediksi meningkat seiring bertambahnya horizon. Pada horizon 30 langkah (c), selisih antara nilai aktual dan prediksi menjadi lebih jelas dan interval kepercayaan 95% terlihat jauh lebih lebar, meskipun model masih mampu menangkap pola umum perubahan kecepatan angin. Secara keseluruhan, grafik ini menunjukkan bahwa semakin panjang horizon prediksi, semakin besar ketidakpastian prediksi yang tercermin dari pelebaran interval kepercayaan.

Diskusi

Hasil penelitian menunjukkan bahwa panjang horizon prediksi merupakan faktor utama yang mempengaruhi tingkat kesalahan prediksi kecepatan angin. Nilai kesalahan meningkat secara konsisten ketika horizon prediksi diperpanjang dari 1 langkah menjadi 7 langkah dan 30 langkah. Fenomena ini umum terjadi pada peramalan deret waktu karena hubungan statistik antara variabel historis dan nilai masa depan cenderung melemah seiring bertambahnya jarak waktu prediksi. Temuan ini sejalan dengan penelitian dalam literatur peramalan energi yang menunjukkan bahwa ketidakpastian prediksi meningkat secara signifikan pada horizon yang lebih panjang [11][18]. Selain itu, beberapa penelitian dalam jurnal energi juga melaporkan bahwa prediksi kecepatan angin jangka panjang tetap menjadi tantangan karena kompleksitas dinamika atmosfer dan variasi musiman [5][4].

Perbandingan strategi Recursive dan Direct menunjukkan bahwa pendekatan Recursive cenderung lebih stabil pada horizon menengah dan panjang dalam dataset ini. Hal ini kemungkinan disebabkan oleh adanya ketergantungan temporal jangka pendek yang kuat pada data kecepatan angin, sehingga model satu langkah yang akurat dapat digunakan secara iteratif untuk menghasilkan prediksi pada horizon berikutnya. Hasil ini juga dilaporkan dalam beberapa studi multistep forecasting yang menunjukkan bahwa pendekatan Recursive dapat mempertahankan stabilitas prediksi ketika data memiliki struktur temporal yang kuat [12][19]. Selain itu, kemampuan model ensemble berbasis pohon seperti LightGBM dalam menangkap hubungan nonlinier antarvariabel juga berkontribusi terhadap kestabilan model pada prediksi multi-langkah [9].

Penelitian ini juga menunjukkan bahwa penyertaan variabel eksogen meteorologi memberikan kontribusi positif terhadap kinerja prediksi. Penambahan suhu minimum dan kelembaban relatif secara konsisten menurunkan nilai MAE pada seluruh horizon prediksi, dengan pengaruh yang lebih besar pada horizon menengah dan panjang. Hasil ini mendukung penelitian sebelumnya yang menunjukkan bahwa faktor atmosfer seperti suhu dan kelembaban memiliki hubungan fisik dengan dinamika angin dan dapat meningkatkan akurasi model prediksi angin ketika digunakan sebagai variabel input tambahan [10]. Variabel eksogen memungkinkan model untuk menangkap dinamika atmosfer yang tidak sepenuhnya tercermin dalam data kecepatan angin historis, sehingga meningkatkan representasinya terhadap proses fisik yang mendasari pembentukan angin [20][21]. Penelitian lain dalam bidang energi terbarukan juga menunjukkan bahwa integrasi variabel meteorologi dalam model machine learning dapat meningkatkan performa prediksi angin dan energi angin [6].

Selain itu, hasil penelitian menunjukkan bahwa tuning hiperparameter memiliki pengaruh yang bergantung pada horizon prediksi. Pada horizon pendek, tuning hiperparameter tidak memberikan peningkatan yang signifikan, bahkan pada beberapa kasus meningkatkan nilai kesalahan. Dalam

beberapa skenario, tuning hiperparameter justru meningkatkan nilai kesalahan, yang mengindikasikan overfitting struktural. Fenomena ini dapat dijelaskan melalui perspektif trade-off bias-varians, di mana peningkatan kompleksitas model LightGBM tanpa kontrol yang memadai dapat meningkatkan varians dan menurunkan generalisasi [22][23]. Hal ini menunjukkan bahwa kompleksitas model tambahan tidak selalu diperlukan ketika pola temporal jangka pendek sudah cukup kuat untuk menghasilkan prediksi yang akurat. Sebaliknya, pada horizon panjang, tuning hiperparameter memberikan peningkatan kinerja yang signifikan. Hal ini menunjukkan bahwa prediksi jangka panjang memerlukan konfigurasi model yang lebih optimal untuk mengontrol kompleksitas model dan menjaga keseimbangan bias-varians. Temuan ini sejalan dengan penelitian dalam machine learning forecasting yang menunjukkan bahwa pengaturan hiperparameter yang tepat menjadi semakin penting ketika kompleksitas masalah prediksi meningkat [16].

Kesimpulan

Penelitian ini menganalisis pengaruh konfigurasi hiperparameter dan variabel eksogen terhadap prediksi kecepatan angin multi-langkah menggunakan algoritma LightGBM dengan strategi Recursive dan Direct. Hasil penelitian menunjukkan bahwa panjang horizon prediksi merupakan faktor utama yang mempengaruhi tingkat kesalahan prediksi, dimana nilai kesalahan meningkat secara konsisten dari horizon 1-langkah hingga 30-langkah. Penyertaan variabel eksogen berupa suhu minimum dan kelembaban relatif terbukti mampu meningkatkan kinerja model pada seluruh horizon, dengan dampak yang lebih besar pada horizon menengah dan panjang. Selain itu, tuning hiperparameter menunjukkan bahwa pengaruhnya bergantung pada horizon. Tuning memberikan manfaat yang lebih jelas untuk prediksi jangka panjang.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa kinerja prediksi kecepatan angin tidak hanya dipengaruhi oleh pemilihan algoritma, tetapi juga oleh kombinasi strategi prediksi, variabel eksogen, konfigurasi hiperparameter, serta panjang horizon prediksi. Oleh karena itu, penelitian selanjutnya dapat mengeksplorasi penggunaan variabel meteorologi tambahan serta membandingkan model LightGBM dengan pendekatan pembelajaran mesin atau deep learning lainnya untuk meningkatkan akurasi prediksi pada horizon yang lebih panjang.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih atas dukungan dana dari Pusat Penelitian dan Penerbitan (Puslitpen) Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) UIN Syarif Hidayatullah Jakarta. Penelitian ini merupakan hasil penelitian yang didanai Badan Layanan Umum (BLU) UIN Syarif Hidayatullah Jakarta No. UN.01/KPA/1068/2024.

Referensi

- [1] N. Fitriyati, *Prediksi Kecepatan Angin Menggunakan Filter Kalman*. Tangerang Selatan: UIN Jakarta Press, 2024.
- [2] F. Rahman, T. E. Sutanto, and N. Fitriyati, "Web Traffic Anomaly Detection using Stacked Long Short-Term Memory," *Inpr. Indones. J. Pure Appl. Math.*, vol. 3, no. 2, pp. 112–121, 2021, doi: 10.15408/inprime.v3i2.21879.
- [3] R. A. Fahrezi, M. Y. Wijaya, and N. Fitriyati, "Prediksi Harga Penutupan Saham Bank Central Asia: Implementasi Algoritma Long Short-Term Memory Dan Perbandingannya Dengan Support Vector Machine," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 5, no. 1, pp. 452–464, 2024, doi: 10.46306/lb.v5i1.582.
- [4] S. S. Band *et al.*, "A two-stage deep learning-based hybrid model for daily wind speed forecasting," *Heliyon*, vol. 11, no. 1, Jan. 2025, doi: 10.1016/j.heliyon.2024.e41026.
- [5] M. Paula, C. Marilaine, F. José Nuno, and C. Wallace, "Predicting Long-Term Wind Speed in Wind Farms of Northeast Brazil: A Comparative Analysis Through Machine Learning Models," *IEEE Lat. Am. Trans.*, vol. 18, no. 11, pp. 2011–2018, 2020, doi: 10.1109/TLA.2020.9398643.
- [6] J. Marc *et al.*, "Optimizing Wind Power Forecasting Using Machine Learning : A Comparative Study with Emphasis on LightGBM

- for Predictive Maintenance,” *Int. J. Adv. Eng. Res. Sci.*, vol. 12, no. 6, pp. 69–80, 2025, doi: <https://dx.doi.org/10.22161/ijaers.126.7> Optimizing.
- [7] J. Park and E. A. Hwang, “A Two-Stage Multistep-Ahead Electricity Load Forecasting Scheme Based on LightGBM and Attention-BiLSTM,” *Sensors*, vol. 21, p. 7697, 2021, doi: <https://doi.org/10.3390/s21227697>.
- [8] Y. Shi, G. Ke, Z. Chen, S. Zheng, and T.-Y. Liu, “Quantized Training of Gradient Boosting Decision Trees,” in *36th Conference on Neural Information Processing Systems (NeurIPS 2022)*, 2023.
- [9] G. Ke *et al.*, “LightGBM: A Highly Efficient Gradient Boosting Decision Tree,” in *31st Conference on Neural Information Processing Systems (NIPS 2017)*, 2017, pp. 1–9.
- [10] I. U. Siloko and U. O.O., “A statistical study of wind speed and its connectivity with relative humidity and temperature in Ughelli, Delta State, Nigeria,” *Sci. World J.*, vol. 18, no. 3, pp. 404–413, 2023, doi: [10.4314/swj.v18i3.13](https://doi.org/10.4314/swj.v18i3.13).
- [11] S. Ben Taieb, G. Bontempi, A. F. Atiya, and A. Sorjamaa, “A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition,” *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7067–7083, 2012, doi: <https://doi.org/10.1016/j.eswa.2012.01.039>.
- [12] S. Ben Taieb and R. J. Hyndman, “Recursive and direct multi-step forecasting : the best of both worlds Recursive and direct multi-step forecasting : the best of both worlds,” 2012.
- [13] T. McElroy, “When are Direct Multi-step and Iterative Forecasts Identical ?,” *J. of Forecasting, J. Forecast.*, vol. 34, pp. 315–336, 2015, doi: [10.1002/for.2321](https://doi.org/10.1002/for.2321).
- [14] S. Ben Taieb and G. Bontempi, “Recursive multi-step time series forecasting by perturbing data,” in *11th IEEE International Conference on Data Mining, ICDM 2011, Vancouver, BC, Canada, December 11-14, 2011*, 2011, pp. 695–704. doi: [10.1109/ICDM.2011.123](https://doi.org/10.1109/ICDM.2011.123).
- [15] A. Lazcano, J. E. Sandubete, and M. A. Jaramillo-Morán, “A comparative framework for multi-horizon time series forecasting: Neural networks with adaptive preprocessing,” *Mach. Learn. with Appl.*, vol. 22, p. 100781, 2025, doi: <https://doi.org/10.1016/j.mlwa.2025.100781>.
- [16] J. Gong, Z. Qu, Z. Zhu, H. Xu, and Q. Yang, “Ensemble models of TCN-LSTM-LightGBM based on ensemble learning methods for short-term electrical load forecasting,” *Energy*, vol. 318, p. 134757, 2025, doi: <https://doi.org/10.1016/j.energy.2025.134757>.
- [17] J. A. Ilemobayo *et al.*, “Hyperparameter Tuning in Machine Learning: A Comprehensive Review,” *J. Eng. Res. Reports*, vol. 26, no. 6, pp. 388–395, 2024, doi: [10.9734/jerr/2024/v26i61188](https://doi.org/10.9734/jerr/2024/v26i61188).
- [18] R. J. and Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice. (3rd ed.)*. 2021. [Online]. Available: <https://otexts.com/fpp3/>
- [19] R. Green, H. Day, Z. A. Abdallah, and T. M. S. Filho, “Epistemic Error Decomposition for Multi-step Time Series Forecasting : Rethinking Bias – Variance in Recursive and Direct Strategies,” *arXiv*, vol. 2511.11461, 2025.
- [20] Y. Zhang, J. Wang, and X. Wang, “Review on probabilistic forecasting of wind power generation,” *Renew. Sustain. Energy Rev.*, vol. 32, pp. 255–270, 2014, doi: <https://doi.org/10.1016/j.rser.2014.01.033>.
- [21] H. Liu, H.-Q. Tian, C. Chen, and Y. Li, “A hybrid statistical method to predict wind speed and wind power,” *Renew. Energy*, vol. 35, no. 8, pp. 1857–1861, 2010, doi: <https://doi.org/10.1016/j.renene.2009.12.011>.
- [22] P. Probst, A.-L. Boulesteix, and B. Bischl, “Tunability : Importance of Hyperparameters of Machine Learning Algorithms,” *J. Mach. Learn. Res.*, vol. 20, pp. 1–32, 2019.
- [23] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, “Understanding deep learning (still) requires rethinking generalization,” *Commun. ACM*, vol. 64, no. 3, pp. 107–115, Feb. 2021, doi: [10.1145/3446776](https://doi.org/10.1145/3446776).