

Prediksi cuaca Multi-Horizon 1-6 jam pada Wilayah Menggunakan Algoritma LightGBM

Neo Ramadhani^{1*}, Intan Yuniar Purbasari², Fetty Tri Anggraeny³

^{1*}Fakultas Ilmu Komputer

Universitas Pembangunan Nasional
Veteran Jawa Timur

^{1*}22081010299@student.upnjatim.ac.id

² Fakultas Ilmu Komputer

Universitas Pembangunan
Nasional Veteran Jawa Timur

²intanyuniar.if@upnjatim.ac.id

³ Fakultas Ilmu Komputer

Universitas Pembangunan
Nasional Veteran Jawa Timur

³fettyangraeny.if@upnjatim.ac.id

Abstrak— Perubahan cepat dalam kondisi cuaca memiliki dampak signifikan terhadap berbagai sektor dan aktivitas di suatu daerah, sehingga sangat penting untuk memiliki prakiraan cuaca jangka pendek yang tepat dan cepat di tingkat lokal. Penelitian ini menggunakan algoritma LightGBM untuk memprediksi empat parameter cuaca, yaitu suhu, kelembapan, curah hujan, dan kecepatan angin, pada beberapa batas waktu, yaitu 1, 3, dan 6 jam ke depan, di beberapa lokasi dengan karakteristik iklim yang beragam. Data cuaca yang diperoleh setiap menit diubah menjadi format per jam, kemudian dibersihkan dan diproses menjadi fitur terkait waktu, lokasi, serta kondisi cuaca saat ini, serta *rolling statistics* untuk menangkap dinamika sementara, guna melatih model *multi output*. Sebagai pembandingan, kinerja model LightGBM diuji dengan dua *baseline*, yaitu yang pertama adalah Random Forest serta persistence forecast. Kinerja model dievaluasi dengan menggunakan metrik MAE, RMSE, galat ternormalisasi (nRMSE), dan koefisien determinasi (R^2) baik secara individual untuk setiap variabel maupun per batas waktu. Hasil pengujian menunjukkan bahwa model mampu menjelaskan sekitar 65% variasi dari keempat parameter (R^2 makro 0,65; nRMSE 0,49), di mana suhu (R^2 0,98) dan kelembapan (R^2 0,89) adalah parameter yang paling akurat, kecepatan angin tergolong sedang karena menyentuh R^2 0,66, sedangkan curah hujan menjadi yang paling sulit diprediksi (R^2 0,08). LightGBM mengungguli kedua *baseline* pada seluruh parameter termasuk pada curah hujan, yang dimana Random Forest dan persistence forecast justru menghasilkan R^2 negatif sehingga keunggulan tidak dipengaruhi oleh data. Akurasi model mengalami penurunan seiring dengan bertambahnya batas waktu, dari R^2 0,75 pada horizon 1 jam, menjadi 0,54 pada 6 jam, sehingga fokus model adalah pada prakiraan jangka pendek 1–6 jam. Dengan demikian, pendekatan gradient boosting berbasis LightGBM dinilai sangat efisien untuk prakiraan cuaca multi-horizon jangka pendek.

Kata kunci— LightGBM, prakiraan cuaca per-jam, *multi-horizon*, *gradient boosting*

Abstract— Rapid changes in weather conditions have a significant impact on various sectors and activities in an area, making it very important to have accurate and quick short-term weather forecasts at the local level. This study uses the LightGBM algorithm to predict four weather parameters, namely temperature, humidity, rainfall, and wind speed, over several time horizons, specifically 1, 3, and 6 hours ahead, at various locations with diverse climatic characteristics. Weather data obtained every minute is converted into an hourly format, then cleaned and processed into time-related features, location, and current weather conditions, as well as rolling statistics to capture temporary dynamics, in order to train a multi-output model. As a comparison, the performance of the LightGBM model was tested against two baselines, the first being Random Forest and the second being persistence forecast. The model's performance is evaluated using the metrics MAE, RMSE, normalized error (nRMSE), and coefficient of determination (R^2) both individually for each variable and per time boundary. The test results show that the model is able to explain about 65% of the variation in the four parameters (macro R^2 0.65; nRMSE 0.49), where temperature (R^2 0.98) and humidity (R^2 0.89) are the most accurate parameters, wind speed is moderate at R^2 0.66, while rainfall is the most difficult to predict (R^2 0.08). LightGBM outperformed both baselines on all parameters, including rainfall, where Random Forest and persistence forecast produced negative R^2 values, indicating that the superiority was not influenced by the data. The model's accuracy decreases as the time horizon increases, from R^2 0.75 at the 1-hour horizon to 0.54 at the 6-hour horizon, thus the model's focus is on short-term forecasts of 1–6 hours. Therefore, the LightGBM-based gradient boosting approach is considered very efficient for short-term multi-horizon weather forecasting.

Keywords— LightGBM, hourly weather forecast, *multi-horizon*, *gradient boosting*.

I. PENDAHULUAN

Cuaca adalah faktor lingkungan yang memiliki dampak besar terhadap berbagai aktivitas dan sektor di suatu area, seperti pertanian, energi, transportasi, serta kegiatan di luar ruangan. Perubahan suhu, kelembapan, kecepatan angin, dan munculnya hujan dalam

waktu singkat dapat mempengaruhi perencanaan dan keputusan operasional di sektor-sektor tersebut. Oleh karena itu, prakiraan cuaca yang akurat, cepat, dan tersedia untuk beberapa periode waktu sekaligus menjadi suatu kebutuhan yang penting, terutama di tingkat lokal di mana kondisi atmosfer sangat berubah-ubah[1].

Kebutuhan ini mendorong penggunaan pendekatan berbasis data untuk menghasilkan prakiraan multi-horizon, yaitu perkiraan cuaca untuk berbagai selang waktu ke depan dari satu hingga enam jam secara bersamaan sebagai dasar untuk pengambilan keputusan.

Secara tradisional, prakiraan cuaca dibuat menggunakan *Numerical Weather Prediction* (NWP), yang memerlukan sumber daya komputasi yang besar dan cukup lambat dalam menyesuaikan dengan pengamatan lokal. Karena itu, pendekatan berbasis pembelajaran mesin semakin dianggap sebagai pilihan yang lebih efisien dan fleksibel[2]. Dalam analisis data cuaca yang bersifat tabular, algoritma gradient boosting sering kali lebih unggul dibandingkan model linear atau prediktor satu variabel, karena kemampuannya dalam menangkap hubungan non-linear antar variabel[3]. Salah satu varian dari algoritma ini, *Light Gradient Boosting Machine* (LightGBM), memiliki keunggulan dalam hal kecepatan pelatihan, penggunaan memori yang efisien, dan akurasi yang bersaing, dan telah terbukti efektif untuk koreksi serta prediksi parameter suhu dan angin[2]. Beberapa penelitian terbaru bahkan menjadikan model berbasis pohon seperti LightGBM sebagai opsi dengan performa terbaik untuk prakiraan jangka pendek dalam skenario multi-horizon[4]. Oleh karena itu, penelitian ini menggunakan LightGBM untuk memprediksi empat parameter cuaca, suhu, kelembapan, curah hujan, dan kecepatan angin secara bersamaan pada horizon satu hingga 6 jam.

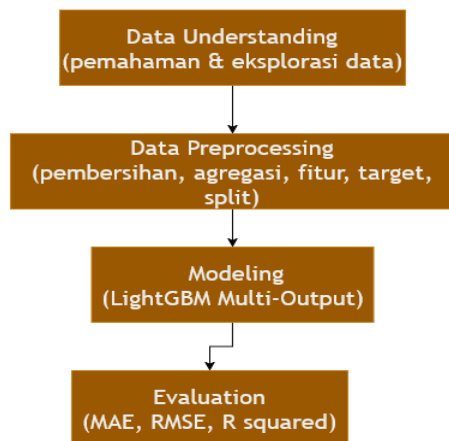
Kendati penggunaan pembelajaran mesin dalam prakiraan cuaca terus meningkat, banyak penelitian masih terbatas pada satu variabel atau satu jangka waktu, dan sering kali memerlukan volume data dan biaya komputasi yang besar ketika menggunakan arsitektur *Deep Learning* [5]. Namun, banyak situasi operasional yang memerlukan model yang sederhana tetapi mampu memprediksi beberapa variabel sekaligus pada berbagai jangka waktu. Tantangan semakin besar karena beberapa parameter khususnya curah hujan memiliki variasi yang tinggi dari

waktu ke waktu, sehingga sulit diprediksi secara akurat[6]. Penelitian komparatif menunjukkan bahwa metode berbasis pohon seperti LightGBM tidak hanya akurat, tetapi juga lebih efisien dalam hal komputasi dibandingkan dengan banyak model pembelajaran mendalam untuk data deret waktu [5], sehingga menjadikannya relevan sebagai dasar pemodelan dalam penelitian ini.

Berdasarkan kebutuhan dan celah yang ada, penelitian ini bertujuan untuk membangun model prakiraan cuaca multi-horizon 1 hingga 6 jam yang menggunakan LightGBM untuk meramalkan empat parameter cuaca secara bersamaan di berbagai wilayah. Pendekatan multi-output untuk memprediksi beberapa variabel cuaca secara bersamaan dengan rekayasa fitur temporal yang bebas dari kebocoran telah terbukti memberikan hasil prediksi yang akurat [7], sementara efektivitas LightGBM juga terlihat pada prediksi curah hujan [8]. Penelitian ini juga membandingkan LightGBM dengan model baseline, yaitu Random Forest dan prediksi persistence yang biasa digunakan untuk prakiraan cuaca jangka pendek. Ini dilakukan agar keunggulan LightGBM dapat dinilai secara objektif, bukan hanya berdasarkan karakteristik data. Kontribusi utama dari penelitian ini meliputi:

- a) penerapan LightGBM multi-output pada data cuaca beresolusi per-jam di berbagai wilayah, yang diperkaya oleh rekayasa fitur *rolling statistics* untuk mengidentifikasi dinamika temporal.
- b) evaluasi kinerja per variabel dan per horizon untuk mengkaji penurunan akurasi seiring dengan bertambahnya horizon waktu; serta
- c) untuk membuktikan keunggulannya secara objektif, membandingkan LightGBM dengan *baseline* Random Forest serta *persistence forecast*.

II. METODOLOGI PENELITIAN



Gambar 1. Diagram alir metodologi penelitian

Penelitian ini mengimplementasikan algoritma Light Gradient Boosting Machine (LightGBM) untuk melakukan prediksi cuaca multi-horizon jangka pendek (hingga 6 jam ke depan) di berbagai wilayah yang memiliki karakteristik iklim yang berbeda. Proses pengembangan model yang disesuaikan menjadi empat tahap utama, yaitu pemahaman data, praproses data, pemodelan data, dan evaluasi. Tahap pemahaman data dilakukan untuk menganalisis struktur dan karakteristik data; praproses data mencakup pembersihan, agregasi, rekayasa fitur, serta penentuan target; pemodelan data membangun model prediksi; dan evaluasi mengukur kinerja model. LightGBM dipilih karena merupakan pohon keputusan gradien boosting yang menggunakan *teknik Gradient-based One-Side Sampling (GOSS) dan Exclusive Feature Bundling (EFB)* sehingga dapat melatih data dalam jumlah besar dengan cepat dan efisien dalam penggunaan memori tanpa mengorbankan akurasi [9].

Data yang digunakan adalah rekaman parameter cuaca historis dari berbagai daerah dengan karakteristik iklim yang berbeda, yang diperoleh melalui Open-Meteo Weather API. Setiap baris data dilengkapi dengan penanda identitas pengamatan (tahun, periode, dan lokasi) serta variabel cuaca utama, yaitu suhu, kelembapan, curah hujan, kecepatan dan arah angin, serta

tekanan udara, ditambahkan dengan variabel turunan seperti suhu terasa dan titik embun. Pada resolusi aslinya, data dicatat per menit di setiap daerah dalam periode pengamatan tertentu sehingga ukurannya dapat mencapai ratusan ribu baris, dengan curah hujan sebagai variabel yang paling tidak seimbang karena didominasi oleh nilai nol.

Pada tahap pemrosesan data, data mentah dibersihkan dari duplikat, nilai yang hilang, dan kolom yang tidak perlu, kemudian diringkas ke dalam resolusi per-jam dengan menghitung rata-rata setiap variabel untuk setiap jam, guna menstabilkan pola dan mengurangi ukuran data. Selanjutnya, fitur waktu dikembangkan dengan mencakup jam dan hari dalam setahun, serta fitur siklik sinus-kosinus untuk menangkap pola harian [10], dan penyandian lokasi serta kondisi cuaca terkini, serta fitur *rolling statistics* untuk menggambarkan dinamika temporal jangka pendek untuk tiap variabel cuaca, dengan rata-rata dan simpangan baku bergerak pada jendela 3 dan 6 jam. Statistik bergerak tersebut dihitung secara terpisah per-wilayah dan hanya menggunakan data selama waktu pengamatan, tanpa memperhitungkan nilai masa depan, sehingga tidak ada kebocoran data. Variabel target ditetapkan. Variabel target dibuat dengan menggeser empat parameter cuaca utama, suhu, kelembapan, curah hujan, dan kecepatan angin ke depan sejauh 1, 3, dan 6 jam, sehingga menghasilkan keluaran multi-output. Data kemudian dibagi secara temporal berdasarkan wilayah dengan proporsi 80:20 [11].

Dalam tahap pemodelan data, LightGBM digunakan dalam kerangka Multi-Output Regressor sehingga satu sistem model dapat memprediksi keempat variabel target pada tiga horizon sekaligus. Konfigurasi utama mencakup 300 pohon (*n_estimators*), tingkat pembelajaran 0,05, serta 63 daun maksimum (*num_leaves*), dengan random seed tetap (42) untuk memastikan hasil yang dapat direproduksi. Dua model dasar dilatih dengan fitur dan skema pembagian data yang sama. Random Forest (*Multi-Output Regressor* berbasis *bagging*) digunakan sebagai

pembandingan berbasis *machine learning*, dan persistence forecast digunakan sebagai tolak ukur minimal yang lazim untuk prakiraan cuaca jangka pendek. Dalam model yang berbasis gradient boosting, penetapan dan pemilihan hyperparameter terbukti berdampak signifikan pada akurasi prediksi [12], sementara tahap praproses dan pengelolaan fitur yang teliti sangat berperan dalam menentukan kualitas input model [13].

Pada tahap evaluasi, kinerja model diukur pada data uji menggunakan empat metrik standar regresi, yaitu *Mean Absolute Error*, *Root Mean Squared Error*, serta koefisien determinasi (R^2). Evaluasi di jalankan tidak secara global, tetapi juga per variabel dan per horizon, sehingga dapat dianalisis variabel mana yang paling sulit diprediksi (misalnya curah hujan) dan bagaimana akurasi cenderung mengalami penurunan seiring bertambahnya waktu. Setelahnya kinerja LightGBM dibandingkan secara langsung dengan kedua *baseline* Random Forest dan prediksi persistence menggunakan metrik yang sama untuk menilai keunggulannya secara objektif.

III. HASIL DAN PEMBAHASAN

Pada bagian ini akan dipaparkan hasil pengujian model LightGBM *multi output* dalam memprediksi empat parameter cuaca suhu, kelembapan, curah hujan, dan kecepatan angin pada tiga horizon waktu 1,3,6 jam, sekaligus membahas artinya. Untuk menunjukkan kinerja model, metrik MAE, RMSE, galat atau error ternormalisasi (nRMSE), dan koefisien determinasi R^2 digunakan baik secara global maupun variabel serta per horizon. Selanjutnya, pola perubahan akurasi seiring bertambahnya rentang waktu prediksi saat rentang waktu dipelajari.

Model LightGBM *multi output* menunjukkan performahor yang baik untuk prakiraan jangka pendek secara keseluruhan model mampu menjelaskan sekitar 65% dari keempat variasi parameter cuaca pada data uji (R^2 0,65 nRMSE

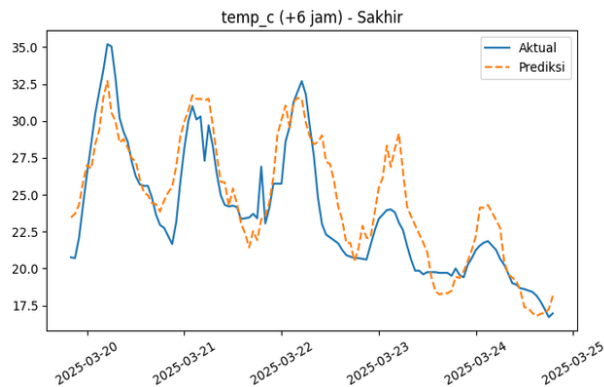
0,49), dengan rincian per variabel yang tersedia dalam table 3.1 dan per horizon yang terdapat pada table 3.2 Pola hasil ini konsisten dengan prinsip suhu (R^2 0,98) dan kelembapan (R^2 0,89) memiliki akurasi tertinggi karena keduanya berfluktuasi dengan halus dan menunjukkan autokorelasi temporal yang kuat, sehingga kondisi beberapa jam mendatang cenderung mirip dengan kondisi saat ini. Kecepatan angin memiliki performa sedang (R^2 0,66) karena lebih bervariasi sedangkan curah hujan menjadi tantangan tersulit (R^2 0,08 nRMSE 0,96) karena nilainya jarang terjadi, didominasi nol, dan sangat tidak menentu, penambahan fitur *rolling statistics* menunjukkan peningkatan kinerja pada curah hujan, menurunkan R^2 dari yang semula mendekati nol pada pemodelan menjadi 0,08. Ini menunjukkan bahwa informasi temporal memang relevan, meskipun prediksi curah hujan masih paling sulit. Akurasi juga mengalami penurunan yang wajar seiring bertambahnya horizon R^2 0,75 (1 jam), 0,66 (3 jam), 0,54(6 jam), sementara nRMSE meningkat dari 0,38 menjadi 0,60. Sementara model hanya menggunakan gambaran sesaat kondisi terkini tanpa data Sejarah. Dengan Batasan horizon maksimum 6 jam, model tetap beroperasi dalam batas andal, sehingga hasilnya lebih kuat dibandingkan apabila dipaksakan melakukan prediksi 24 jam.

Tabel 3. 1. Kinerja model per variabel

Variabel	Matriks			
	MAE	RMSE	nRMSE	R^2
Suhu	0,98	1,36	0,14	0,98
Kelembapan	5,97	8,01	0,32	0,89
Kecepatan angin	2,50	3,25	0,56	0,66
Curah hujan	0,23	0,73	0,96	0,08

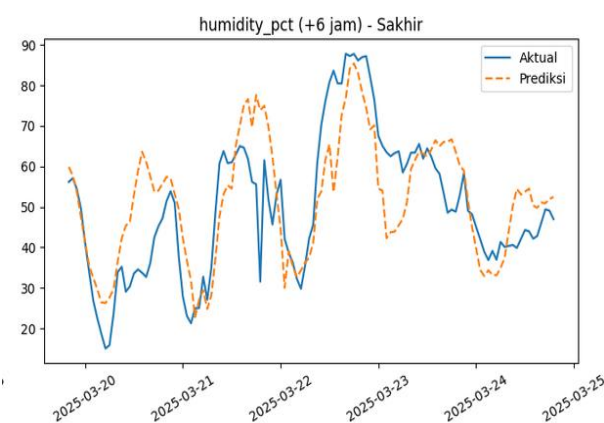
Tabel 3. 2. Kinerja model per horizon waktu

Variabel	Matriks			
	MAE	RMSE	nRMSE	R^2
1 jam	1,28	1,89	0,38	0,75
3 jam	2,53	3,47	0,50	0,66
6 jam	3,45	4,65	0,60	0,54



Gambar 2. Perbandingan nilai aktual dan prediksi suhu ($^{\circ}\text{C}$) pada horizon +6 jam di wilayah Sakhir. Sumbu-x menunjukkan waktu pengamatan data uji (resolusi per-jam; label tanggal hanya penanda harian dari plot), sumbu-y menunjukkan suhu dalam derajat Celsius.

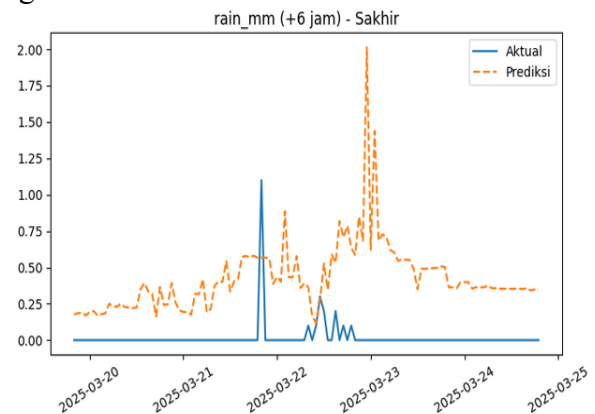
Pada gambar 2, dalam diagram suhu, kurva yang diprediksi hampir sejajar dengan nilai aktual dan berhasil mencerminkan siklus harian fluktuasi dari sekitar 17°C hingga 35°C ; puncak dan lembah harian tercatat dengan selisih yang sangat kecil, hanya kadang-kadang sedikit meleset pada titik puncak tertinggi. Hal ini menunjukkan bahwa suhu adalah variabel yang paling mudah untuk diprediksi (R^2 0,98) karena perubahannya berlangsung halus dan memiliki autokorelasi temporal yang kuat, sehingga kondisi enam jam mendatang masih sangat mirip dengan kondisi saat ini.



Gambar 3. Perbandingan nilai aktual dan prediksi suhu ($^{\circ}\text{C}$) pada horizon +6 jam di wilayah Sakhir. Sumbu-x menunjukkan waktu pengamatan data uji (resolusi per-jam; label tanggal hanya penanda harian dari plot), sumbu-y menunjukkan tingkat kelembapan atau nilai.

Pada gambar 3 diagram kelembapan, prediksi mengikuti fluktuasi aktual yang cukup lebar

(berkisar 15–90%) dengan baik dan mampu menangkap pola naik-turun antarjam. Meski begitu, masih terdapat selisih tipis pada beberapa puncak dan lembah yang tajam, di mana model sedikit melebihi- atau mengurangi-perkiraan nilai ekstrem. Secara keseluruhan akurasi tetap tinggi (R^2 0,89) karena kelembapan seperti suhu berubah relatif mulus dan berkorelasi kuat dengan kondisi terkini.



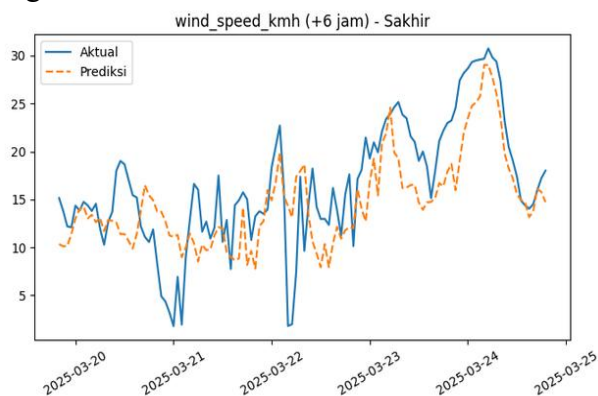
Gambar 4. Perbandingan nilai aktual dan prediksi suhu ($^{\circ}\text{C}$) pada horizon +6 jam di wilayah Sakhir. Sumbu-x menunjukkan waktu pengamatan data uji (resolusi per-jam; label tanggal hanya penanda harian dari plot), sumbu y menunjukkan nilai hujan atau curah.

Selanjutnya pada gambar 4. diagram curah hujan, terdapat perbedaan yang sangat jelas: nilai aktual hampir selalu nol dengan beberapa lonjakan tajam yang jarang terjadi, sementara prediksi model menunjukkan sinyal yang "berisik" dengan nilai kecil (sekitar 0,2–2,0 mm) yang berfluktuasi dan tidak sinkron dengan kejadian nyata model bahkan menghasilkan estimasi hujan ketika keadaan sebenarnya kering dan gagal menangkap lonjakan hujan yang sesungguhnya. Meskipun penambahan fitur *rolling statistics* sedikit memperbaikinya curah hujan menjadi variabel yang paling sulit untuk diprediksi ($R^2 \approx 0,08$). Prediksi curah hujan menjadi sulit disebabkan oleh beberapa faktor:

1. nilai yang jarang dan didominasi oleh nol, membuat data menjadi sangat tidak seimbang;
2. karakteristiknya yang intermiten dan tiba-tiba sehingga tidak berkesinambungan seperti suhu atau kelembapan; dan
3. hampir tidak ada korelasi dengan kondisi saat itu hujan enam jam ke depan lebih

dipengaruhi oleh dinamika atmosfer (pergerakan tekanan dan massa udara) yang hanya sebagian terekam oleh fitur *rolling statistics* jangka pendek.

Akibatnya, model regresi yang berusaha meminimalkan kesalahan rata-rata cenderung "bermain aman" dengan memperkirakan nilai kecil yang bukan nol, sehingga R^2 -nya menjadi sangat rendah.



Gambar 5. Perbandingan nilai aktual dan prediksi suhu ($^{\circ}\text{C}$) pada horizon +6 jam di wilayah Sakhir. Sumbu-x menunjukkan waktu pengamatan data uji (resolusi per-jam; label tanggal hanya penanda harian dari plot), sumbu-y menunjukkan kecepatan angin km/h.

Kecepatan angin pada gambar 3.1. Diagram prediksi dan aktual D, dalam diagram kecepatan angin, model ini cukup baik dalam menangkap tren umum dan peningkatan secara bertahap, termasuk peningkatan yang menuju puncak sekitar 30 km/jam. Namun, model ini cenderung menghaluskan lonjakan yang mendadak, sehingga sebagian puncak dan lembah yang ekstrem tidak sepenuhnya teridentifikasi dan terlihat sedikit tertunda. Kinerjanya dapat dikategorikan sebagai sedang (R^2 0,66) karena kecepatan angin lebih berfluktuasi dan kurang halus dibandingkan suhu maupun kelembapan, meskipun pola umumnya masih cukup jelas.

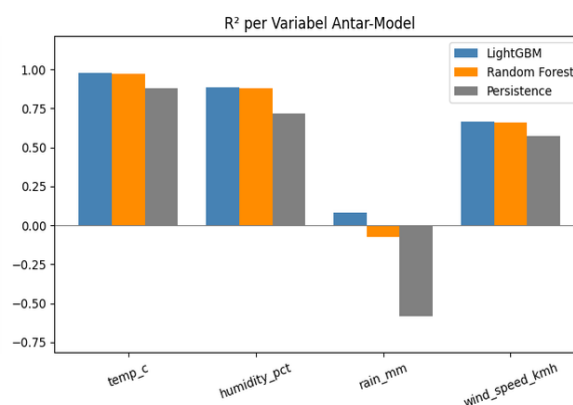
Untuk memastikan bahwa keunggulan LightGBM benar-benar berasal dari kemampuan modelnya dan bukan dari karakteristik dataset, kinerjanya diuji dengan dua model dasar: Forest Random (berbasis *bagging*) dan *Forecasting Persistence* (model naif yang menganggap kondisi H jam ke depan sama dengan kondisi saat ini). Semua tiga model dilatih dan diuji dengan

fitur dan skema pemisahan data yang sama. Hasilnya disajikan dalam Tabel 3.3 dan ini adalah rangkuman hasilnya.

Tabel 3. 3. Koefisien determinasi model LightGBM terhadap Random Forest serta Persistence pada data uji

Variabel	R^2		
	LightGBM	Random Forest	Persistence
Suhu	0,98	0,97	0,88
Kelembapan	0,89	0,88	0,72
Curah hujan	0,08	-0,08	-0,58
Kecepatan angin	0,66	0,66	0,57

LightGBM mengungguli kedua baseline pada seluruh variabel dan secara makro (R^2 makro 0,65 dibandingkan 0,61 pada Random Forest dan 0,40 pada persistence), dan juga mencatat galat ternormalisasi makro terendah (nRMSE). Dalam hal variabel yang mudah diprediksi seperti suhu, kelembapan, dan angin, perbedaan terhadap Random Forest agak kecil, tetapi sangat besar pada curah hujan. LightGBM tetap bernilai positif (R^2 0,08), sedangkan Random Forest (-0,08) dan persistence (-0,58) bernilai negatif, yang menunjukkan bahwa kedua dasar bahkan lebih buruk daripada menebak nilai rata-rata. Dataset tidak dapat diprediksi secara sederhana hanya dengan menganggap kondisi tidak berubah, seperti yang ditunjukkan oleh persistence yang tertinggal jauh (R^2 makro 0,40). Hasil menunjukkan bahwa keunggulan LightGBM adalah nyata dan tidak hanya berasal dari atribut dataset.



Gambar 6. R^2 per variabel antar model dimana sumbu x menunjukkan variabel, sumbu y menunjukkan nilai R^2

Kekurangan dan potensi perbaikan Secara umum, kendala utama model adalah prediksi curah hujan yang lebih sulit dan kurang akurat pada horizon yang lebih jauh. Kinerja telah diperbaiki dengan menambah fitur *rolling statistics*, terutama pada curah hujan di mana R^2 -nya naik dari nol menjadi 0,08. Namun, peningkatan ini hanya sedikit karena sifat hujan yang intermiten dan didominasi oleh nol. Perbaikan lain termasuk penambahan fitur-fitur tambahan lainnya.

IV. KESIMPULAN

Penelitian ini berhasil mengembangkan model *Light Gradient Boosting Machine* (LightGBM) yang mampu memprediksi empat parameter secara bersamaan yaitu suhu, kelembapan, curah hujan, serta kecepatan angin, pada horizon pendek selama 1, 3, 6 jam. Model ini memanfaatkan data dari Open-Meteo yang telah diolah menjadi resolusi perjam, serta dilakukan rekayasa fitur berdasarkan waktu, lokasi dan kondisi cuaca saat ini, dan *rolling statistics* dengan pemsiahan data temporal yang bebas dari kebocoran. Secara keseluruhan, model ini dapat menjelaskan sekitar 65% dari variasi keempat parameter tersebut (R^2 makro yang menghasilkan 0,65 serta nRMSE 0,49). Prediksi suhu (R^2 0,98) dan kelembapan (R^2 0,89) menunjukkan akurasi tertinggi karena sifatnya yang berubah dan lancar serta memiliki autokorelasi yang kuat. Kecepatan angin memiliki tingkat akurasi sedang (R^2 0,66) sedangkan curah hujan menjadi yang paling sulit diprediksi (R^2 0,08) akibat kondisi yang jarang terjadi, didominasi nol, serta tidak menentu. Akurasi model menurun seiring dengan bertambahnya horizon waktu (R^2 0,75 pada 1 jam menjadi 0,54 pada 6 jam). Dibandingkan dengan model *baseline*, LightGBM unggul pada semua parameter, termasuk persistence yang memiliki nilai R^2 negatif dan curah hujan pada Random Forest. Dengan demikian, dapat disimpulkan bahwa LightGBM merupakan pendekatan yang efisien dan akurat untuk prakiraan cuaca multi-horizon jangka pendek, terutama pada parameter yang kontinuitasnya tinggi, ada beberapa keterbatasan yang masih perlu ditingkatkan

dalam penelitian berikutnya. Penambahan fitur *rolling statistics* pada penelitian ini telah memperbaiki kinerja, terutama pada curah hujan, namun peningkatannya masih terbatas, oleh karena itu penelitian lanjutan seperti penambahan fitur-fitur baru atau mungkin menggunakan model-model terbaru serta mendapatkan data yang lebih akurat dengan cakumungkin akan membantu lebih jauh dalam penelitian ini.

REFERENSI

- [1] M. Zanchi, S. Zapperi, S. Bocchi, O. Drofa, S. Davolio, and C. A. M. La Porta, "Improving localized weather predictions for precision agriculture: A Time-Series Mixer approach for hazardous event detection," *Environmental Modelling & Software*, vol. 191, p. 106509, Jun. 2025, doi: 10.1016/j.envsoft.2025.106509.
- [2] R. Tang, Y. Ning, C. Li, W. Feng, Y. Chen, and X. Xie, "Numerical forecast correction of temperature and wind using a single-station single-time spatial lightgbm method," *Sensors*, vol. 22, no. 1, Jan. 2022, doi: 10.3390/s22010193.
- [3] B. Carrera and K. Kim, "Comparison analysis of machine learning techniques for photovoltaic prediction using weather sensor data," *Sensors (Switzerland)*, vol. 20, no. 11, Jun. 2020, doi: 10.3390/s20113129.
- [4] K. Blazakis, S. Perrakis, E. Karapidakis, F. Mavromatakis, and Y. Katsigiannis, "ML-Based Multi-Horizon Wind Speed and Wind Direction Forecasting for Aviation and Energy Applications in Coastal Crete," *Electronics (Switzerland)*, vol. 14, no. 24, Dec. 2025, doi: 10.3390/electronics14244856.
- [5] T. Hall and K. Rasheed, "A Survey of Machine Learning Methods for Time Series Prediction," Jun. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/app15115957.
- [6] V. S. Monego, J. A. Anochi, and H. F. de Campos Velho, "South America Seasonal Precipitation Prediction by Gradient-Boosting Machine-Learning Approach," *Atmosphere (Basel)*, vol. 13, no. 2, Feb. 2022, doi: 10.3390/atmos13020243.
- [7] A. Kumar, J. Farmania, and P. P. Suresh, "Weather Prediction using Machine Learning Regression Models."
- [8] "LightGBM's effectiveness in Rainfall Prediction." [Online]. Available: <https://ssrn.com/abstract=4397719>
- [9] G. Ke *et al.*, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," 2017. [Online]. Available: <https://www.researchgate.net/publication/378480234>
- [10] S. Khazem and H. Kanso, "Cyclical Temporal Encoding and Hybrid Deep Ensembles for Multistep Energy Forecasting," Dec. 2025, [Online]. Available: <http://arxiv.org/abs/2512.03656>
- [11] S. Albelali and M. Ahmed, "Hidden Leaks in Time Series Forecasting: How Data Leakage Affects LSTM Evaluation Across Configurations and Validation Strategies".
- [12] I. Y. Purbasari, A. Priharyoto Bayuseno, R. Rizal Isnanto, and T. I. Winarni, "Readmission Risk Prediction After Total Hip Arthroplasty Using Machine Learning and Hyperparameter Optimized with Bayesian Optimization." [Online]. Available: www.ijacsa.thesai.org
- [13] F. T. Anggraeny, Y. V. Via, and R. Mumpuni, "Image preprocessing analysis in handwritten Javanese character recognition," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 2, pp. 860–867, Apr. 2023, doi: 10.11591/eei.v12i2.4172.