

Perbandingan Kinerja Model Ensemble Learning untuk Prediksi Kejadian Hujan Esok Hari Berbasis Data Termodinamika BMKG

Comparison of Ensemble Learning Model Performance for Next-Day Rainfall Occurrence Prediction Based on BMKG Thermodynamic Data

Rizal Wahyu Pratama¹, Khulika Malkan^{*2}, Mikhael Setia Budi³, Atika Ratna Dewi⁴

^{1,2,3,4} Program Studi Sains Data, Fakultas Informatika, Telkom University

e-mail: ^{*2}khulika@student.telkomuniversity.ac.id

Abstrak - Prediksi kejadian hujan harian merupakan komponen krusial dalam sistem irigasi pertanian cerdas, namun ketersediaan variabel meteorologi dari sumber resmi yang terbatas menjadi tantangan utama. Penelitian ini mengusulkan sistem prediksi hujan esok hari menggunakan pendekatan ensemble learning berbasis data klimatologi harian SACA&D BMKG Stasiun Meteorologi Kemayoran periode 1993 hingga 2014 dengan total 8.035 observasi. Tiga fitur rekayasa dikonstruksi dari data suhu yaitu Temp_Range sebagai indikator termodinamika serta Day of Year dan Bulan sebagai representasi siklus monsun. Lima algoritma berbasis pohon keputusan dibandingkan dengan penanganan ketidakseimbangan kelas menggunakan pembobotan adaptif. Gradient Boosting terpilih sebagai model terbaik dengan AUC-ROC 0,7088 dan akurasi 66,89 persen setelah optimasi ambang batas klasifikasi. Studi ablasi menggunakan Time-Series Cross Validation lima lipatan membuktikan peningkatan AUC-ROC rata-rata sebesar 0,0754 secara konsisten lintas dekade. Dominasi Day of Year sebagai prediktor terkuat mengkonfirmasi peran pola monsun tahunan dalam menentukan kejadian hujan harian di Jakarta.

Kata kunci – ensemble learning; Gradient Boosting; prediksi hujan harian; rekayasa fitur termodinamika; Time-Series Cross Validation

Abstract - Daily rainfall prediction is a crucial component in smart agricultural irrigation systems, yet the limited availability of meteorological variables from official sources poses a significant challenge. This study proposed a next-day rainfall prediction system using an ensemble learning approach based on daily climatological data from SACA&D BMKG Kemayoran Meteorological Station for the period 1993 to 2014, comprising 8,035 observations. Three engineered features were constructed: Temp_Range as a thermodynamic indicator, and Day of Year and Month as monsoon cycle representations. Five tree-based algorithms were compared with class imbalance addressed through adaptive weighting. Gradient Boosting emerged as the best model with an AUC-ROC of 0.7088 and accuracy of 66.89 percent after classification threshold optimization. Five-fold Time-Series Cross Validation proved consistent average AUC-ROC improvement of 0.0754 across decades. Day of Year dominance as the strongest predictor confirmed the annual monsoon pattern role in determining daily rainfall events in Jakarta.

Keywords – ensemble learning; Gradient Boosting; daily rainfall prediction; thermodynamic feature engineering; Time-Series Cross Validation

I. PENDAHULUAN

Sektor pertanian Indonesia menopang lebih dari 38 juta tenaga kerja dan menyumbang sekitar 13 persen dari Produk Domestik Bruto nasional [1]. Ketergantungan yang sangat besar terhadap pola curah hujan menjadikan sektor ini rentan terhadap ketidakpastian iklim yang terus meningkat akibat perubahan iklim global. Petani di berbagai wilayah, termasuk Banyumas dan sekitarnya, masih mengandalkan pengetahuan turun-temurun untuk menentukan jadwal irigasi dan panen. Ketika perkiraan tersebut meleset, kerugian yang ditanggung dapat mencapai gagal panen total [2].

Prediksi hujan yang akurat untuk satu hari ke depan dapat mengubah cara pengelolaan irigasi secara fundamental. Sistem irigasi pertanian cerdas yang bekerja otomatis berdasarkan prakiraan cuaca menjadi solusi yang relevan dan dibutuhkan [3]. Perkembangan *machine learning* membuka peluang dalam bidang ini karena algoritma mampu mengekstrak pola tersembunyi dari rekam historis klimatologi dengan sumber daya komputasi yang jauh lebih efisien dibandingkan model numerik atmosfer konvensional [4], [5].

Stasiun Meteorologi Kemayoran di Jakarta merupakan salah satu stasiun pengamatan cuaca tertua di Indonesia dengan rekam data lebih dari satu abad. Data hariannya tersedia secara terbuka melalui *Southeast Asian Climate Assessment and Dataset* (SACA&D) yang dikelola bersama BMKG dan *Royal Netherlands Meteorological Institute* (KNMI) [6], [7]. Ketersediaan data jangka panjang yang telah melalui kendali mutu ketat menjadikannya fondasi ideal untuk membangun model prediksi yang dapat divalidasi lintas dekade.

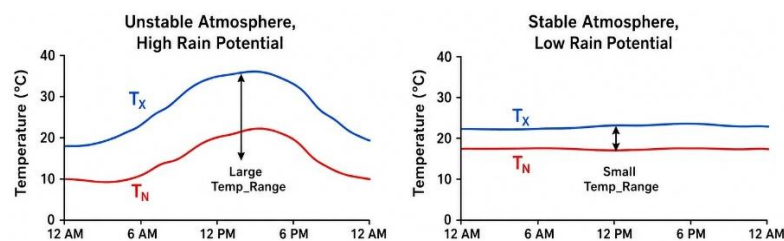
Tantangan utama penelitian ini terletak pada keterbatasan variabel prediktor yang tersedia secara publik, yaitu hanya data suhu harian. Merespons keterbatasan ini, penelitian mengembangkan fitur rekayasa termodinamika berupa *Temp_Range* yaitu selisih suhu maksimum dan minimum harian yang secara meteorologis merupakan indikator ketidakstabilan atmosfer [8], [9]. Fitur temporal *Day of Year* dan Bulan juga ditambahkan untuk menangkap siklus monsun yang dominan di wilayah Jakarta [10], [11]. Dengan landasan rekayasa fitur tersebut, penelitian ini membandingkan lima algoritma *ensemble learning* berbasis pohon keputusan yaitu Random Forest, XGBoost, LightGBM, CatBoost, dan GradientBoosting menggunakan data klimatologi resmi BMKG. Perbandingan dilakukan tidak hanya pada satu evaluasi, melainkan divalidasi secara temporal menggunakan *Time-Series Cross Validation* yang mencakup rentang data dua dekade, sehingga hasil perbandingan yang diperoleh lebih andal dan representatif [12].

II. TINJAUAN TEORI

Penelitian prediksi curah hujan berbasis *machine learning* telah berkembang pesat dengan fokus pada pendekatan berbasis data yang mampu menangkap pola non-linear dalam variabel meteorologi [13]. Sarasa-Cabezuelo [1] menemukan bahwa Random Forest secara konsisten unggul pada klasifikasi biner hujan atau tidak hujan menggunakan dataset cuaca Australia selama sepuluh tahun. Sebuah studi serupa oleh Mdegela et al. [14] menunjukkan keunggulan konsisten algoritma berbasis pohon pada data tropis dengan distribusi kelas tidak seimbang di Tanzania.

Kim et al. [5] membuktikan bahwa penggabungan XGBoost, LightGBM, dan CatBoost dalam arsitektur *ensemble* menghasilkan performa yang lebih stabil dibandingkan model tunggal mana pun. Oprea et al. [15], [16] membandingkan langsung ketiga algoritma *gradient boosting* tersebut untuk prediksi aliran sungai harian dan menemukan bahwa CatBoost unggul pada konfigurasi parameter awal. Dalam konteks Indonesia, Handoko et al. [16] membuktikan XGBoost efektif menangani kompleksitas pola curah hujan menggunakan data BMKG. Sementara itu, Gomez-Skarmeta et al. [17] secara khusus menunjukkan keunggulan Gradient Boosting atas *neural network* pada dataset berukuran terbatas, kondisi yang identik dengan penelitian ini.

Gradient Boosting membangun model secara sekuensial di mana setiap pohon baru memperbaiki kesalahan residual pohon sebelumnya melalui pendekatan *gradient descent* dalam ruang fungsi [4]. Rentang suhu harian atau *Temp_Range* merupakan selisih suhu maksimum dan minimum dalam satu hari pengamatan. Secara termodinamika, rentang suhu yang besar mengindikasikan ketidakstabilan lapisan batas atmosfer yang mendorong konveksi dan pembentukan awan berpotensi hujan [8] yang diilustrasikan pada Gambar 1.



Gambar 1. Hubungan Rentang Suhu Harian dengan Konveksi Atmosfer

Time-Series Cross Validation dengan pendekatan *expanding window* adalah teknik validasi yang menghormati urutan kronologis data dengan selalu menggunakan semua data historis yang tersedia untuk pelatihan kemudian menguji model pada periode berikutnya. Pendekatan ini jauh lebih representatif dibandingkan *k-fold cross validation* konvensional untuk menilai ketahanan model terhadap pergeseran pola iklim jangka panjang [12].

Chen et al. [18] membuktikan bahwa pembobotan kelas adaptif memberikan hasil yang kompetitif tanpa risiko *overfitting* yang muncul pada metode *oversampling* sintetis. Stoean et al. [19] menunjukkan bahwa metode optimasi metaheuristik mampu menemukan konfigurasi model yang optimal lebih efisien dibandingkan pencarian konvensional. Penelitian Ariska et al. [11] dan Mulsandi et al. [10] secara khusus menuliskan pola variasi spasiotemporal curah hujan Indonesia dan sinyal monsun yang sangat dominan di wilayah Jawa, memperkuat justifikasi penggunaan fitur temporal dalam penelitian ini.

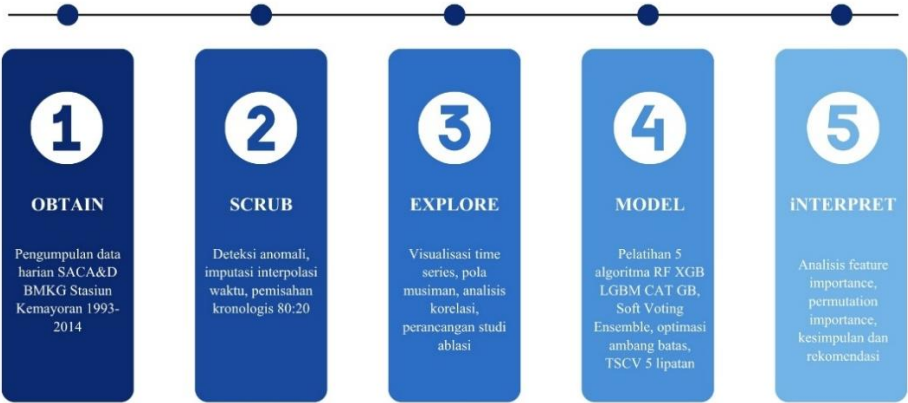
III. METODE

3.1 Data

Penelitian ini menggunakan data klimatologi harian dari SACA&D BMKG Stasiun Meteorologi Kemayoran, Jakarta (STAID: 180) pada koordinat $-6,13^{\circ}\text{LS}$ dan $106,65^{\circ}\text{BT}$ dengan ketinggian 8 meter di atas permukaan laut [6]. Rentang waktu yang digunakan adalah 1 Januari 1993 hingga 31 Desember 2014 menghasilkan 8.035 observasi harian. Pembatasan hingga 2014 dilakukan untuk menjaga homogenitas instrumen pengukuran sebelum transisi besar-besaran ke *Automatic Weather Station* sekitar tahun 2015.

3.2 Kerangka Kerja Penelitian

Seluruh tahapan penelitian mengikuti kerangka kerja OSEMN yang terdiri dari lima fase yaitu *Obtain* untuk pengumpulan data SACA&D BMKG, *Scrub* untuk pembersihan dan imputasi data, *Explore* untuk analisis pola dan korelasi, *Model* untuk pelatihan dan evaluasi lima algoritma *ensemble learning*, serta *iNterpret* untuk analisis kepentingan fitur dan penarikan kesimpulan. Diagram alur lengkap disajikan pada Gambar 3.



Gambar 3. Diagram Alur Penelitian Berbasis OSEMN

3.3 Persiapan Data dan Rekayasa Fitur

Nilai anomali fisik diidentifikasi berdasarkan batas yang tidak mungkin terjadi di iklim tropis Jakarta. Curah hujan di atas 300 mm per hari, suhu maksimum di luar rentang 20 hingga 40°C, suhu minimum di luar 15 hingga 35°C, dan suhu rata-rata di luar 20 hingga 40°C ditandai sebagai kesalahan pencatatan. Seluruh nilai hilang diimputasi menggunakan interpolasi berbasis waktu yang mempertimbangkan jarak temporal antar observasi valid [20].

Tiga fitur rekayasa dikonstruksi dari data yang telah dibersihkan. *Temp_Range* dihitung sebagai selisih TX dan TN pada hari yang sama. Variabel Bulan diekstraksi sebagai bilangan bulat 1 hingga 12. *Day of Year* diekstraksi sebagai posisi hari dalam siklus tahunan dengan rentang 1 hingga 365. Target didefinisikan sebagai status hujan pada hari berikutnya dengan ambang batas 0,5 mm untuk membedakan kejadian hujan efektif dari hujan sangat ringan. Pemisahan data dilakukan secara kronologis dengan rasio 80 berbanding 20 menghasilkan 6.427 hari data latih periode 1993 hingga 2010 dan 1.607 hari data uji periode 2010 hingga 2014.

3.4 Pemodelan dan Evaluasi

Lima algoritma *ensemble learning* berbasis pohon keputusan dikonfigurasi dengan mekanisme penanganan ketidakseimbangan kelas masing-masing. Konfigurasi parameter disajikan pada Tabel 1.

Tabel 1. Konfigurasi Penanganan Ketidakseimbangan Kelas

Model	Parameter	Keterangan
Random Forest	<i>class_weight</i> = <i>balanced</i>	Bobot berbanding terbalik frekuensi kelas
XGBoost	<i>scale_pos_weight</i> = 2,48	Rasio frekuensi kelas mayoritas terhadap minoritas
LightGBM	<i>is_unbalance</i> = <i>True</i>	Penyesuaian bobot sampel otomatis

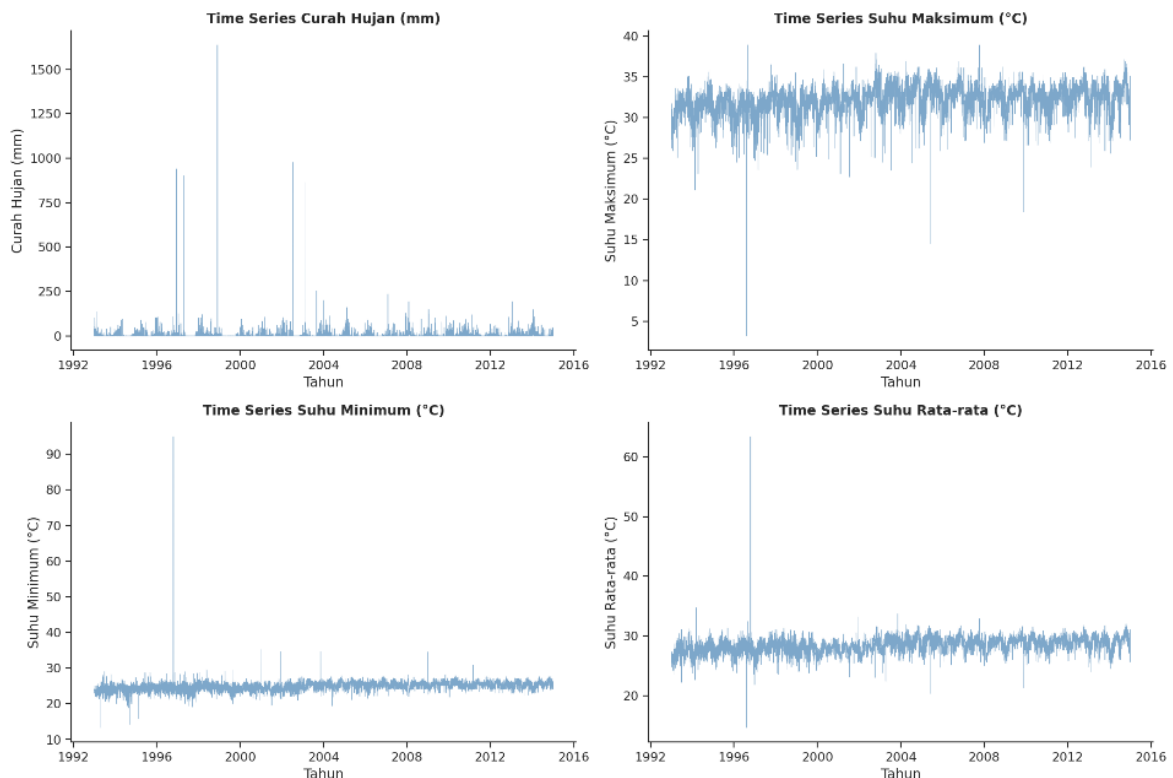
CatBoost	<i>auto_class_weights = Balanced</i>	Invers frekuensi relatif setiap kelas
Gradient Boosting	<i>compute_sample_weight balanced</i>	Bobot sampel eksplisit sebelum pelatihan

Evaluasi dilakukan menggunakan AUC-ROC sebagai metrik primer karena tidak bergantung pada ambang batas klasifikasi dan lebih representatif untuk data tidak seimbang. Akurasi, F1-Macro, dan F1-Weighted digunakan sebagai metrik sekunder. Studi ablasi membandingkan skenario *baseline* yang hanya menggunakan TX, TN, dan TG terhadap skenario usulan dengan tujuh fitur lengkap. Pencarian ambang batas klasifikasi optimal dilakukan secara sistematis pada rentang 0,10 hingga 0,90 dengan langkah 0,01 untuk memaksimalkan F1-Macro [21].

IV. HASIL DAN PEMBAHASAN

4.1 Eksplorasi dan Pembersihan Data

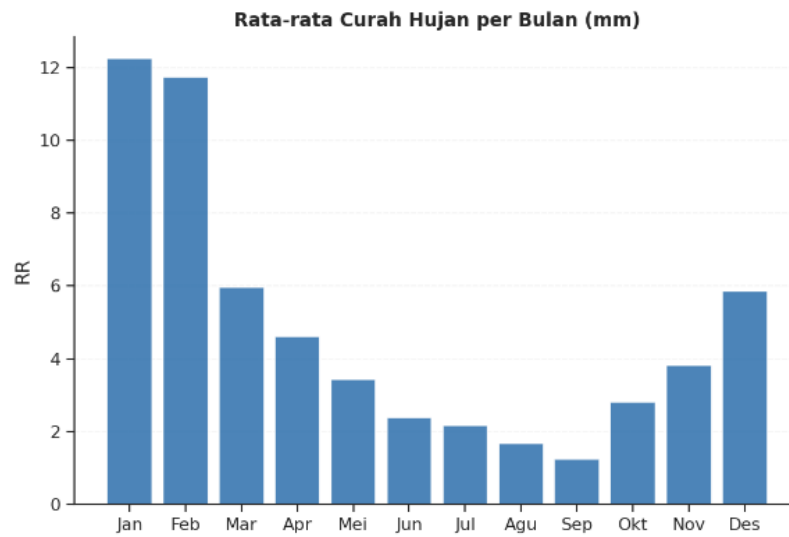
Visualisasi *time series* data mentah pada Gambar 4 mengungkap lima lonjakan curah hujan melampaui 800 mm per hari yang mustahil terjadi secara fisik mengingat rekor harian Jakarta sekitar 340 mm. Variabel suhu menunjukkan tiga nilai ekstrem di bawah 15°C dan satu nilai mendekati 95°C pada suhu minimum yang juga tidak mungkin terjadi di iklim tropis. Total 14 nilai anomali fisik berhasil diidentifikasi dan diimputasi menggunakan interpolasi berbasis waktu [20].



Gambar 4. Visualisasi Time Series Data Mentah Stasiun Kemayoran 1993-2014

Pola musiman pada Gambar 5 menunjukkan karakteristik monsun yang sangat jelas. Curah hujan mencapai puncaknya pada Januari dengan rata-rata 12,25 mm per hari dan mencapai titik terendah pada September dengan hanya 1,24 mm per hari. Pola ini sepenuhnya konsisten dengan karakteristik monsun Asia Tenggara yang mendorong curah hujan tinggi ke wilayah Jawa pada periode Oktober hingga April [10], [11]. Temuan ini memperkuat

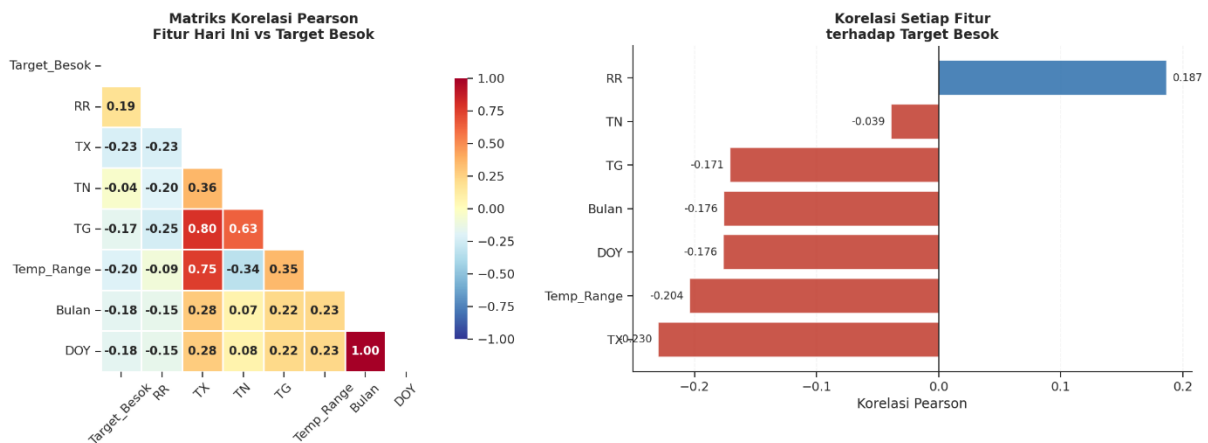
justifikasi penggunaan DOY dan Bulan sebagai fitur temporal dalam model.



Gambar 5. Pola Musiman Rata-rata Curah Hujan Bulanan 1993-2014

4.2 Analisis Korelasi Fitur

Matriks korelasi pada Gambar 6 menunjukkan bahwa curah hujan hari ini memiliki korelasi positif tertinggi terhadap target sebesar 0,187 mengkonfirmasi adanya autokorelasi temporal yang bermakna dalam data curah hujan harian. *Temp_Range* memiliki korelasi negatif sebesar -0,204, artinya hari dengan rentang suhu yang besar cenderung diikuti hari tidak hujan esok harinya. Ini konsisten dengan mekanisme termodinamika di mana rentang suhu tinggi pada musim kemarau mengindikasikan atmosfer yang stabil dan kering [8],[9].



Gambar 6. Matriks Korelasi Fitur terhadap Target Prediksi

4.3 Performa Model

Hasil evaluasi kelima model beserta Soft Voting Ensemble dan Gradient Boosting dengan ambang batas optimal disajikan pada Tabel 2. Gradient Boosting mencatat AUC-ROC tertinggi sebesar 0,7088 mengungguli keempat model lainnya, konsisten dengan temuan Gomez-Skarmeta et al. [17] mengenai keunggulan Gradient Boosting pada dataset berukuran terbatas. Random Forest memiliki akurasi tertinggi namun F1-Macro terendah sebesar 0,5773 mengindikasikan bias kuat terhadap kelas mayoritas sehingga kurang mampu mendeteksi

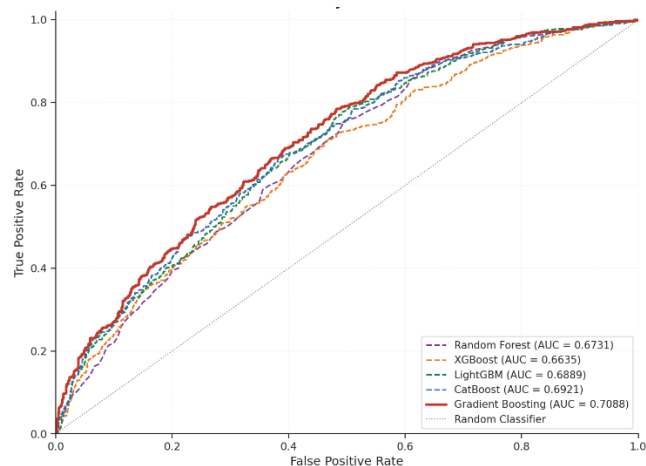
kejadian hujan.

Tabel 2. Perbandingan Performa Seluruh Model pada Data Uji

Model	AUC-ROC	Akurasi	F1-Macro	F1-Weighted
Random Forest	0,6731	0,6652	0,5773	0,6362
XGBoost	0,6635	0,6372	0,6066	0,6401
LightGBM	0,6889	0,6391	0,6206	0,6462
CatBoost	0,6921	0,6360	0,6195	0,6437
Gradient Boosting	0,7088	0,6409	0,6282	0,6492
Soft Voting Ensemble	0,6924	0,6621	0,6099	0,6535
GB + Threshold Optimal	0,7088	0,6689	0,6390	0,6708

Soft Voting Ensemble menghasilkan AUC-ROC 0,6924 yang lebih rendah dari Gradient Boosting yang berdiri sendiri. Fenomena ini terjadi karena kelima model dilatih dari himpunan fitur yang sama sehingga memiliki korelasi kesalahan yang tinggi dan penggabungan tidak menghasilkan diversifikasi yang berarti. Setelah optimasi ambang batas dari 0,50 menjadi 0,54, Gradient Boosting meningkatkan akurasi menjadi 66,89 persen dan F1-Weighted menjadi 0,6708 menjadikannya model final yang dipilih.

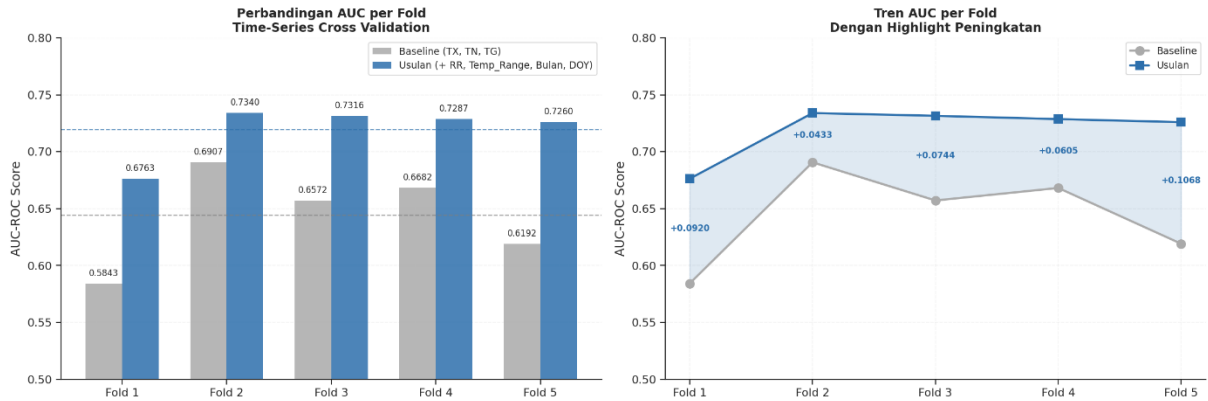
Kurva ROC pada Gambar 7 memperlihatkan Gradient Boosting secara konsisten berada di atas semua model lain di seluruh rentang ambang batas, mengkonfirmasi superioritas model ini tidak hanya pada satu titik ambang batas tertentu.



Gambar 7. Kurva ROC Perbandingan Semua Model

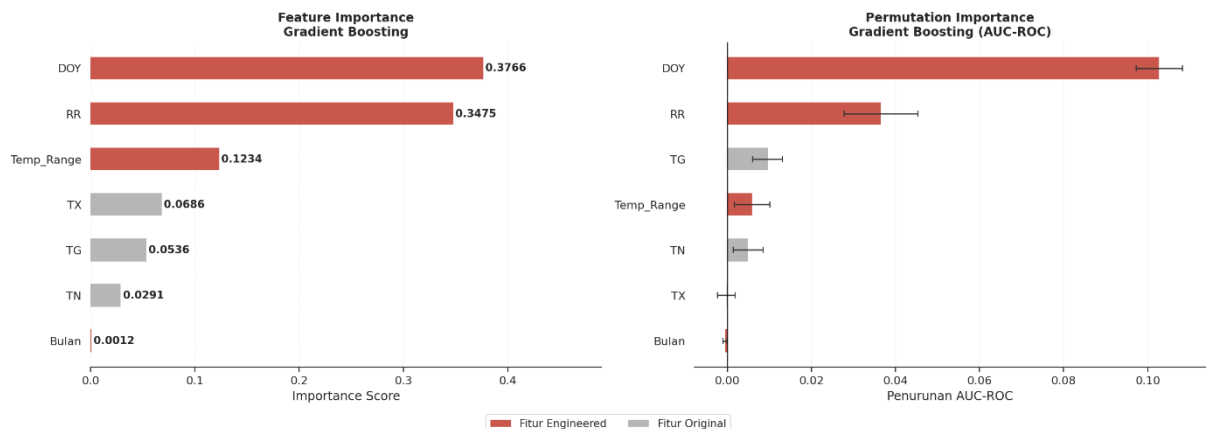
4.4 Studi Ablasi dan Kepentingan Fitur

Studi ablasi menggunakan *Time-Series Cross Validation* lima lipatan pada Gambar 8 membuktikan bahwa rekayasa fitur yang diusulkan meningkatkan rata-rata AUC-ROC sebesar 0,0754 secara konsisten di seluruh lipatan, dari 0,6439 menjadi 0,7193. Konsistensi peningkatan di setiap lipatan tanpa terkecuali membuktikan bahwa fitur rekayasa menangkap pola iklim yang stabil lintas dekade dan bukan sekadar artefak statistik pada periode tertentu [12].



Gambar 8. Studi Ablasi Kontribusi Rekayasa Fitur melalui TSCV

Analisis kepentingan fitur pada Gambar 9 mengungkap *Day of Year* sebagai prediktor terkuat dengan skor 0,3766, diikuti curah hujan hari ini sebesar 0,3475 dan *Temp_Range* sebesar 0,1234. Dominasi DOY mengkonfirmasi bahwa pola siklus monsun tahunan adalah faktor paling determinan dalam prediksi hujan harian di Jakarta, sebuah temuan yang sejalan dengan dokumentasi pola monsun Indonesia oleh Mulsandi et al. [10] dan Ariska et al. [11]. *Temp_Range* yang merupakan fitur rekayasa termodinamika asli penelitian ini menempati posisi ketiga, membuktikan relevansi fisiknya sebagai indikator kondisi atmosfer [8], [9].



Gambar 9. Feature Importance dan Permutation Importance Model Terbaik

V. KESIMPULAN DAN SARAN

Penelitian ini berhasil membuktikan bahwa perbandingan kinerja lima algoritma *ensemble learning* berbasis pohon keputusan menghasilkan temuan yang konsisten dan tervalidasi secara temporal. Dari kelima algoritma yang dibandingkan, Gradient Boosting menjadi model dengan kinerja terbaik yang memperoleh AUC-ROC 0,7088, akurasi 66,89 persen, dan F1-Weighted 0,6708 setelah optimasi ambang batas klasifikasi ke nilai 0,54. Rekayasa fitur termodinamika *Temp_Range* dan temporal *Day of Year* terbukti meningkatkan rata-rata AUC-ROC sebesar 0,0754 secara konsisten melalui *Time-Series Cross Validation* lima lipatan lintas dua dekade, membuktikan kontribusi nyata fitur rekayasa terhadap seluruh model yang dibandingkan. Dominasi *Day of Year* sebagai prediktor terkuat mengonfirmasi peran pola

monsoon tahunan sangat determinan dalam menentukan kejadian hujan harian di Jakarta.

Penelitian selanjutnya direkomendasikan untuk memperluas variabel prediktor dengan kelembapan udara dan tekanan permukaan dari data reanalisis ERA5, mereplikasi metodologi pada jaringan multi-stasiun di seluruh Indonesia, serta mengeksplorasi arsitektur *deep learning* berbasis *transformer* untuk menangkap dependensi temporal jangka panjang yang lebih kompleks.

VI. DAFTAR PUSTAKA

- [1] A. Sarasa-Cabezuelo, "Prediction of Rainfall in Australia Using Machine Learning," *Information*, vol. 13, no. 4, hlm. 163, Mar 2022, doi: 10.3390/info13040163.
- [2] M. Del-Coco, M. Leo, dan P. Carcagnì, "Machine Learning for Smart Irrigation in Agriculture: How Far along Are We?," *Information*, vol. 15, no. 6, hlm. 306, Mei 2024, doi: 10.3390/info15060306.
- [3] E. A. Abioye *dkk.*, "Precision Irrigation Management Using Machine Learning and Digital Farming Solutions," *AgriEngineering*, vol. 4, no. 1, hlm. 70–103, Feb 2022, doi: 10.3390/agriengineering4010006.
- [4] V. Krasnopolsky, "Applying Machine Learning in Numerical Weather and Climate Modeling Systems," *Climate*, vol. 12, no. 6, hlm. 78, Mei 2024, doi: 10.3390/cli12060078.
- [5] J. M. Ahn, J. Kim, dan K. Kim, "Ensemble Machine Learning of Gradient Boosting (XGBoost, LightGBM, CatBoost) and Attention-Based CNN-LSTM for Harmful Algal Blooms Forecasting," *Toxins (Basel)*, vol. 15, no. 10, hlm. 608, Okt 2023, doi: 10.3390/toxins15100608.
- [6] Bambang Setiyo Prayitno *dkk.*, "Daily dataset for climate extremes analyses in Southeast Asia," SACA&D BMKG-KNMI. Diakses: 12 Juni 2026. [Daring]. Tersedia pada: <https://sacad.bmkg.go.id/>
- [7] E. J. M. van den Besselaar, G. van der Schrier, R. C. Cornes, A. S. Iqbal, dan A. M. G. Klein Tank, "SA-OBS: A Daily Gridded Surface Temperature and Precipitation Dataset for Southeast Asia," *J. Clim.*, vol. 30, no. 14, hlm. 5151–5165, Jul 2017, doi: 10.1175/JCLI-D-16-0575.1.
- [8] J. Cruz, M. Belo-Pereira, A. Fonseca, dan J. A. Santos, "Dynamic and Thermodynamic Drivers of Severe Sub-Hourly Precipitation Events in Mainland Portugal," *Atmosphere (Basel)*, vol. 14, no. 9, hlm. 1443, Sep 2023, doi: 10.3390/atmos14091443.
- [9] J. An *dkk.*, "Short-Term Heavy Rainfall Potential Identification Driven by Physical Features: Model Development and SHAP-Based Mechanism Interpretation," *Climate*, vol. 14, no. 1, hlm. 24, Jan 2026, doi: 10.3390/cli14010024.
- [10] A. Mulsandi, Y. Koesmaryono, R. Hidayat, A. Faqih, dan A. Sopaheluwakan, "Detecting Indonesian Monsoon Signals and Related Features Using Space–Time Singular Value Decomposition (SVD)," *Atmosphere (Basel)*, vol. 15, no. 2, hlm. 187, Jan 2024, doi: 10.3390/atmos15020187.
- [11] M. Ariska, Suhadi, Supari, M. Irfan, dan I. Iskandar, "Spatio-Temporal Variations of Indonesian Rainfall and Their Links to Indo-Pacific Modes," *Atmosphere (Basel)*, vol. 15, no. 9, hlm. 1036, Agu 2024, doi: 10.3390/atmos15091036.
- [12] H. Zhang, Y. Liu, C. Zhang, dan N. Li, "Machine Learning Methods for Weather Forecasting: A Survey," *Atmosphere (Basel)*, vol. 16, no. 1, hlm. 82, Jan 2025, doi: 10.3390/atmos16010082.

- [13] L. Chen, B. Han, X. Wang, J. Zhao, W. Yang, dan Z. Yang, "Machine Learning Methods in Weather and Climate Applications: A Survey," *Applied Sciences*, vol. 13, no. 21, hlm. 12019, Nov 2023, doi: 10.3390/app132112019.
- [14] L. Mdegela, E. Municio, Y. De Bock, E. Luhanga, J. Leo, dan E. Mannens, "Extreme Rainfall Event Classification Using Machine Learning for Kikuletwa River Floods," *Water (Basel)*, vol. 15, no. 6, hlm. 1021, Mar 2023, doi: 10.3390/w15061021.
- [15] R. Szczepanek, "Daily Streamflow Forecasting in Mountainous Catchment Using XGBoost, LightGBM and CatBoost," *Hydrology*, vol. 9, no. 12, hlm. 226, Des 2022, doi: 10.3390/hydrology9120226.
- [16] M. Putra, M. S. Rosid, dan D. Handoko, "High-Resolution Rainfall Estimation Using Ensemble Learning Techniques and Multisensor Data Integration," *Sensors*, vol. 24, no. 15, hlm. 5030, Agu 2024, doi: 10.3390/s24155030.
- [17] M. Zambudio Martínez, L. H. M. da Silveira, R. Marin-Perez, dan A. F. S. Gomez, "Development and Comparison of Artificial Neural Networks and Gradient Boosting Regressors for Predicting Topsoil Moisture Using Forecast Data," *AI*, vol. 6, no. 2, hlm. 41, Feb 2025, doi: 10.3390/ai6020041.
- [18] J.-Y. Chiang, Y. Lio, C.-Y. Hsu, C.-L. Ho, dan T.-R. Tsai, "Binary Classification with Imbalanced Data," *Entropy*, vol. 26, no. 1, hlm. 15, Des 2023, doi: 10.3390/e26010015.
- [19] N. Bacanin, C. Stoean, M. Zivkovic, M. Rakic, R. Strulak-Wójcikiewicz, dan R. Stoean, "On the Benefits of Using Metaheuristics in the Hyperparameter Tuning of Deep Learning Models for Energy Load Forecasting," *Energies (Basel)*, vol. 16, no. 3, hlm. 1434, Feb 2023, doi: 10.3390/en16031434.
- [20] R. S. Pontoh, T. Toharudin, B. N. Ruchjana, N. Sijabat, dan M. D. Puspita, "Bandung Rainfall Forecast and Its Relationship with Niño 3.4 Using Nonlinear Autoregressive Exogenous Neural Network," *Atmosphere (Basel)*, vol. 13, no. 2, hlm. 302, Feb 2022, doi: 10.3390/atmos13020302.
- [21] O. Ejike, D. Ndzi, dan M. Z. Shakir, "Comparative Study of Machine Learning-Based Rainfall Prediction in Tropical and Temperate Climates," *Climate*, vol. 13, no. 8, hlm. 167, Agu 2025, doi: 10.3390/cli13080167.