

Perbandingan Kinerja Algoritma *Machine Learning* untuk Prediksi Cuaca di Wilayah Tropis Indonesia: Studi Komparatif *Random Forest*, *SVM*, *LSTM*, *XGBoost*, dan *LightGBM*

Nofirman ^{a*}, Munawir ^b, Fegie Yoanti Wattimena ^c

^{a*} Universitas Prof Dr Hazairin SH, Kota Bengkulu, Provinsi Bengkulu, Indonesia.

^b Universitas Bali Internasional Muhammadiyah Bali, Kota Denpasar, Provinsi Bali, Indonesia.

^c Universitas Ottow Geissler, Kota Jayapura, Provinsi Papua, Indonesia.

ABSTRACT

Weather prediction in tropical Indonesia faces complex challenges due to high climate variability, persistent El Niño–Southern Oscillation (ENSO) influence, and uneven observational coverage. This study compared five machine learning algorithms — Random Forest (RF), Support Vector Machine (SVM), Long Short-Term Memory (LSTM), XGBoost, and LightGBM — using 187,320 daily records from BMKG stations, ERA5 reanalysis, and TRMM satellite data (2000–2023). Preprocessing included MMDIF-RF imputation, Z-score normalization, and SMOTE for class imbalance correction. Models were evaluated on RMSE, MAE, R^2 , Accuracy, Precision, Recall, F1-Score, and AUC-ROC. LSTM achieved the best performance (RMSE = 3.94 mm; R^2 = 0.891; F1-Score = 0.887; AUC-ROC = 0.941), reflecting its capacity to capture long-range temporal dependencies. XGBoost and LightGBM delivered competitive accuracy at 8–18 times lower training cost, while SVM recorded the lowest accuracy with the highest computational demand. Regional analysis showed station density and data completeness were more consequential than algorithm choice — LSTM RMSE ranged from 3.61 mm in West Java to 5.43 mm in East Nusa Tenggara. A tiered hybrid approach is recommended: LightGBM or XGBoost for routine forecasting and LSTM for extreme event detection, alongside expanded BMKG coverage in eastern Indonesia.

ABSTRAK

Prediksi cuaca di wilayah tropis Indonesia menghadapi tantangan kompleks akibat variabilitas iklim tinggi, pengaruh persisten El Niño–Southern Oscillation (ENSO), dan cakupan jaringan observasi yang tidak merata. Penelitian ini membandingkan kinerja lima algoritma machine learning — Random Forest (RF), Support Vector Machine (SVM), Long Short-Term Memory (LSTM), XGBoost, dan LightGBM — menggunakan 187.320 rekaman harian dari stasiun BMKG, reanalisis ERA5, dan satelit TRMM (2000–2023). Praproses mencakup imputasi MMDIF-RF, normalisasi Z-score, dan SMOTE untuk ketidakseimbangan kelas. Evaluasi menggunakan RMSE, MAE, R^2 , Akurasi, Presisi, Recall, F1-Score, dan AUC-ROC. LSTM mencapai performa terbaik (RMSE = 3,94 mm; R^2 = 0,891; F1-Score = 0,887; AUC-ROC = 0,941), mencerminkan keunggulannya menangkap dependensi temporal jangka panjang. XGBoost dan LightGBM memberikan akurasi kompetitif dengan biaya pelatihan 8–18 kali lebih rendah, sementara SVM mencatat akurasi terendah dengan kebutuhan komputasi tertinggi. Analisis per wilayah menunjukkan kepadatan stasiun dan kelengkapan data lebih menentukan dibandingkan pilihan algoritma — RMSE LSTM berkisar dari 3,61 mm di Jawa Barat hingga 5,43 mm di Nusa Tenggara Timur. Pendekatan hibrida bertingkat direkomendasikan: LightGBM atau XGBoost untuk prakiraan harian rutin dan LSTM untuk deteksi kejadian ekstrem, disertai perluasan jaringan BMKG di wilayah timur Indonesia.

ARTICLE HISTORY

Received 5 May 2026

Accepted 20 May 2026

Published 30 May 2026

KEYWORDS

Machine Learning; Weather Prediction; Tropical Region; Indonesia; BMKG; ERA5; TRMM; LSTM; XGBoost; LightGBM; Random Forest; ENSO; Extreme Rainfall.

KATA KUNCI

Machine Learning; Prediksi Cuaca; Wilayah Tropis; Indonesia; BMKG; ERA5; TRMM; LSTM; XGBoost; LightGBM; Random Forest; ENSO; Curah Hujan Ekstrem.

1. Pendahuluan

Indonesia bukan sekadar negara kepulauan biasa. Posisinya tepat di garis khatulistiwa, diapit dua samudra besar, menjadikan iklimnya salah satu yang paling sulit diprediksi di dunia. Curah hujan tahunan yang tinggi, distribusi spasial yang tidak merata, dan perubahan pola yang cepat membuat upaya memperoleh informasi cuaca yang akurat menjadi pekerjaan yang jauh dari sederhana. Masalahnya bukan hanya soal teknologi — jaringan *rain gauge* yang tersebar tidak merata di seluruh kepulauan memperparah kesenjangan data, terutama di wilayah terpencil yang justru sering menjadi titik rawan bencana (Putra *et al.*, 2024). Variabilitas curah hujan di Indonesia sangat dipengaruhi oleh *El Niño–Southern Oscillation* (ENSO) dan sistem monsun, dua fenomena yang secara bersamaan membentuk pola presipitasi ekstrem di seluruh Asia Tenggara (Chen *et al.*, 2024). Ketika ENSO memasuki fase *La Niña*, sebagian wilayah Indonesia dapat mengalami curah hujan jauh di atas normal dalam waktu singkat. Sebaliknya, fase *El Niño* membawa kekeringan berkepanjangan. Kedua kondisi ini sama-sama berbahaya, dan keduanya membutuhkan sistem prediksi yang mampu membaca sinyal jauh sebelum dampaknya terasa di lapangan. Konsekuensi dari prediksi yang meleset bukan hanya kerugian ekonomi — banjir bandang dan tanah longsor yang dipicu curah hujan ekstrem telah berulang kali merenggut nyawa. Pemantauan cuaca secara *real-time* bukan lagi kemewahan; itu adalah kebutuhan dasar di wilayah yang rentan terhadap bencana hidrometeorologi (Mardyansyah *et al.*, 2022). Akurasi prediksi cuaca juga menyentuh sektor-sektor yang lebih luas: pertanian yang bergantung pada musim tanam, transportasi udara dan laut, pengelolaan waduk dan sumber daya air, produksi energi terbarukan, hingga koordinasi respons darurat saat bencana terjadi (Latif *et al.*, 2023).

Selama beberapa dekade, *Numerical Weather Prediction* (NWP) menjadi andalan utama prakiraan meteorologi global. Model ini bekerja dengan menyelesaikan persamaan dinamika atmosfer secara numerik — pendekatan yang secara teori cukup kuat, tetapi dalam praktiknya menghadapi batas kemampuan yang nyata. NWP kesulitan menangkap sistem cuaca berskala kecil yang bersifat transien: badai petir lokal, *squall line*, dan konveksi mendalam yang tumbuh dalam hitungan jam. Di kawasan maritim Asia Tenggara, justru fenomena-fenomena inilah yang paling sering menyebabkan bencana (Lee *et al.*, 2022). Menjalankan model NWP resolusi tinggi juga membutuhkan infrastruktur superkomputer yang tidak semua negara berkembang mampu sediakan secara memadai. Akibatnya, prakiraan yang dihasilkan sering kali memiliki resolusi spasial yang terlalu kasar untuk menangkap variabilitas cuaca lokal — kelemahan yang sangat terasa di wilayah seperti Indonesia dengan topografi yang beragam dan tersebar di ribuan pulau. Keterbatasan inilah yang mendorong komunitas ilmiah untuk mencari pendekatan lain, bukan untuk menggantikan NWP sepenuhnya, melainkan untuk melengkapi atau bahkan melampaui kemampuannya dalam skenario-skenario tertentu.

Dalam satu dekade terakhir, *machine learning* (ML) telah berkembang dari sekadar alat bantu analisis menjadi pendekatan yang secara serius bersaing dengan model fisika dalam pemodelan cuaca dan iklim (de Burgh-Day & Leeuwenburg, 2023). Kemampuan ML untuk belajar dari data historis dalam jumlah besar, mengenali pola *non-linear* yang tidak bisa ditangkap oleh persamaan fisika sederhana, serta menghasilkan estimasi yang adaptif secara spasial dan temporal menjadikannya kandidat yang layak untuk mengatasi kelemahan NWP (Putra *et al.*, 2024). Sebuah survei terhadap lebih dari 20 metode ML menunjukkan bahwa pendekatan berbasis data ini semakin mendominasi literatur prakiraan cuaca, baik untuk skala jangka pendek maupun prediksi iklim jangka menengah dan panjang (Chen *et al.*, 2023). Di Thailand, Waqas *et al.* (2023) mengkaji secara sistematis berbagai teknik kecerdasan buatan — mulai dari *machine learning* klasik hingga *deep learning* — untuk prakiraan curah hujan dengan fokus pada kebutuhan sektor pertanian dan pengelolaan sumber daya air, dan menemukan bahwa model berbasis AI mampu menghasilkan prakiraan yang lebih presisi dibandingkan metode statistik konvensional, terutama untuk kejadian curah hujan ekstrem. Di Singapura, Lee *et al.* (2022) mengidentifikasi bahwa pendekatan *physics-informed machine learning* — yang menggabungkan pengetahuan fisika atmosfer dengan kemampuan belajar mesin — menjadi arah penelitian paling menjanjikan untuk mengatasi kesalahan sistematis dalam pemodelan cuaca kawasan Asia Tenggara. Di tingkat kawasan, Chitwatksiri &

Miyamoto (2023) menunjukkan bahwa model *data-driven* berbasis AI mampu mengatasi keterbatasan model hidraulik berbasis proses fisik dalam sistem prakiraan banjir perkotaan secara *real-time*, khususnya di wilayah yang tidak memiliki sumber daya komputasi besar — kondisi yang relevan langsung dengan banyak kota di Indonesia yang rentan banjir namun memiliki keterbatasan infrastruktur pemodelan. ML juga telah diterapkan dalam prediksi siklon tropis, mencakup aspek *genesis*, lintasan, intensitas, hingga dampak cuaca ekstrem yang menyertainya (Chen *et al.*, 2020). Pendekatan hibrida yang menggabungkan pengetahuan fisika atmosfer dengan kapasitas belajar dari data berpotensi menjadi jalan tengah yang paling realistis: tidak membuang warisan ilmu atmosfer yang sudah mapan, tetapi juga tidak menutup mata terhadap apa yang bisa dilakukan oleh data (Lee *et al.*, 2022).

Pertumbuhan literatur ML untuk prediksi cuaca memang pesat secara global, namun ada satu celah yang terus berulang: sebagian besar studi dilakukan di wilayah beriklim sedang atau sub-tropis, dengan ketersediaan data yang jauh lebih baik dibandingkan kawasan tropis berkembang seperti Indonesia. Studi yang secara khusus membandingkan kinerja beberapa algoritma ML sekaligus dalam konteks iklim tropis Indonesia masih sangat jarang. Yang ada pun umumnya terfokus pada satu algoritma tertentu, satu stasiun pengamatan, atau satu periode waktu yang terbatas — hasilnya sulit digeneralisasi. Tidak ada jawaban yang jelas dan terukur atas pertanyaan mendasar yang sebenarnya paling dibutuhkan oleh para praktisi dan peneliti di lapangan: dari sekian banyak algoritma ML yang tersedia, mana yang paling sesuai untuk karakteristik data cuaca tropis Indonesia? Pertanyaan itulah yang menjadi titik tolak penelitian ini. Penelitian ini membandingkan kinerja lima algoritma ML — *Random Forest*, *Support Vector Machine*, LSTM, XGBoost, dan LightGBM — dalam memprediksi parameter cuaca di wilayah tropis Indonesia menggunakan data gabungan dari stasiun BMKG, reanalisis ERA5, dan satelit TRMM, mengidentifikasi algoritma dengan performa terbaik berdasarkan metrik evaluasi regresi maupun klasifikasi, serta merumuskan rekomendasi metodologis yang dapat dijadikan acuan bagi pengembangan sistem prediksi cuaca berbasis ML di Indonesia ke depannya.

2. Landasan Teori

Random Forest (RF) bekerja dengan cara yang cukup intuitif: bangun ratusan pohon keputusan secara paralel, masing-masing dilatih pada *subset* data yang berbeda, lalu gabungkan hasilnya melalui mekanisme *voting* atau rata-rata. Pendekatan *bagging* ini secara alami meredam *overfitting* yang sering menjadi masalah pada pohon keputusan tunggal. Dalam konteks data meteorologi tropis Indonesia, RF memiliki keunggulan yang relevan secara langsung — data BMKG seringkali mengandung fitur-fitur yang saling berkorelasi seperti suhu, kelembaban, dan tekanan udara, dan RF mampu menangani multikolinearitas semacam ini tanpa memerlukan seleksi fitur yang ketat sebelumnya. Yu *et al.* (2017) menunjukkan bahwa RF mampu menangani variabel *input* yang beragam — curah hujan historis berbasis *grid*, posisi geografis, hingga data elevasi — dan menghasilkan prakiraan 1 jam ke depan yang memuaskan. Di Indonesia, RF yang dikombinasikan dengan *Grid Search Cross-Validation* (GSCV) terbukti mampu menghasilkan model prediksi hujan yang lebih baik menggunakan data BMKG (Putra *et al.*, 2024). RF juga digunakan sebagai komponen klasifikasi dalam strategi *Double Machine Learning* (DML) untuk membedakan kejadian hujan dari kondisi kering, memperlihatkan bahwa algoritma ini cukup fleksibel untuk dipadukan dalam arsitektur yang lebih kompleks (Lyu & Yong, 2024).

Support Vector Machine (SVM) mencari *hyperplane* yang memisahkan kelas data dengan margin terlebar — sebuah prinsip geometris yang sederhana namun terbukti tangguh dalam banyak skenario klasifikasi dan regresi. Kekuatan utamanya terletak pada kemampuan menangani data berdimensi tinggi dan pola *non-linear* melalui fungsi *kernel*, yang menjadikannya relevan untuk data meteorologi yang seringkali memiliki banyak variabel saling berkorelasi (AbdulRaheem *et al.*, 2022). Namun dalam konteks data cuaca tropis Indonesia, SVM menghadapi tantangan struktural yang serius — dataset meteorologi berskala nasional yang mencakup ribuan rekaman harian dari puluhan stasiun selama beberapa dekade melampaui zona kenyamanan SVM, karena kompleksitas komputasinya meningkat secara kuadrat

terhadap jumlah sampel. Perbandingan langsung antara RF dan SVM dalam prakiraan curah hujan berbasis radar memberikan temuan yang menarik: SVM cenderung lebih unggul untuk prakiraan 2 hingga 3 jam ke depan, sementara RF lebih kompetitif pada prakiraan 1 jam (Yu *et al.*, 2017). Perbedaan ini bukan hal sepele — dalam konteks peringatan dini bencana, selisih satu jam bisa berarti sangat banyak. Dalam studi perbandingan yang lebih luas, SVM tetap muncul sebagai kandidat yang layak diperhitungkan di antara algoritma ML lainnya, meski dengan catatan keterbatasan skalabilitas yang perlu diperhatikan (Barrera-Animas *et al.*, 2022).

Data cuaca pada dasarnya adalah data deret waktu. Suhu pagi ini dipengaruhi suhu kemarin, kelembaban minggu lalu, bahkan pola musiman bulan sebelumnya. *Recurrent Neural Network* (RNN) dirancang untuk menangani dependensi semacam ini, tetapi RNN standar gagal ketika dependensinya terlalu panjang — masalah yang dikenal sebagai *vanishing gradient*. *Long Short-Term Memory* (LSTM) hadir untuk menyelesaikan masalah itu. Mekanisme *gate* pada LSTM — *input gate*, *forget gate*, dan *output gate* — memungkinkan jaringan untuk secara selektif mengingat informasi yang relevan dan membuang yang tidak. Relevansi LSTM untuk data cuaca tropis Indonesia terletak pada kemampuannya menangkap sinyal ENSO yang beroperasi pada skala waktu bulanan hingga tahunan, sekaligus merespons variasi monsun yang bersiklus sub-musiman — sesuatu yang tidak mampu dilakukan oleh algoritma berbasis pohon yang memperlakukan setiap sampel secara independen. Dalam prediksi suhu jangka pendek menggunakan data beberapa stasiun cuaca di Jerman, LSTM terbukti mampu menangkap pola temporal yang tidak bisa ditangkap oleh model regresi biasa (Kreuzer *et al.*, 2020). Ketika LSTM digabungkan dengan *Convolutional Neural Network* (CNN), hasilnya lebih menarik lagi — CNN menangani ekstraksi fitur spasial sementara LSTM menangani dependensi temporal, dan kombinasi keduanya menghasilkan performa yang lebih baik dalam prediksi suhu historis dibandingkan masing-masing model berdiri sendiri (Gong *et al.*, 2024). Bukti kemampuan LSTM dalam mengenali pola halus pada data sekuensial juga datang dari domain lain: dalam prediksi *burnout* karyawan, LSTM mencatat akurasi 0,99 — mengungguli EBM (0,96) dan XGBoost (0,95) (Kahfi *et al.*, 2025). Angka ini relevan bukan karena domainnya sama, melainkan karena memperlihatkan bahwa LSTM konsisten unggul ketika pola temporal menjadi kunci prediksi. Studi perbandingan khusus untuk prakiraan curah hujan deret waktu juga mengonfirmasi bahwa varian LSTM — termasuk *Stacked-LSTM* dan *Bidirectional-LSTM* — secara konsisten berada di antara model terbaik (Barrera-Animas *et al.*, 2022).

Berbeda dari RF yang membangun pohon secara paralel, XGBoost (*Extreme Gradient Boosting*) membangunnya secara sekuensial. Setiap pohon baru difokuskan untuk memperbaiki kesalahan pohon sebelumnya — itulah inti dari *gradient boosting*. Yang membuat XGBoost menonjol dibanding implementasi *boosting* lainnya adalah kemampuan regularisasi bawaannya (L1 dan L2) yang secara aktif mencegah *overfitting*, serta kecepatan komputasinya yang jauh lebih tinggi berkat optimasi pada level sistem. Untuk data meteorologi Indonesia yang bercirikan banyak fitur saling berkorelasi dan distribusi curah hujan yang sangat miring (*skewed*) — dengan mayoritas hari kering dan sebagian kecil hari hujan ekstrem — mekanisme *boosting* sekuensial XGBoost secara teori lebih cocok dibandingkan *bagging*, karena setiap iterasi secara eksplisit menargetkan sampel yang sulit diprediksi, termasuk kejadian ekstrem yang paling penting dalam konteks peringatan dini. Dalam prediksi cuaca, XGBoost digunakan sebagai komponen klasifikasi dalam strategi *Double Machine Learning*: XGBoost dan RF menentukan apakah suatu kejadian adalah hujan atau tidak, lalu LSTM mengambil alih untuk mengestimasi jumlah curah hujannya, dan kombinasi ini menghasilkan estimasi presipitasi *multi-source* dengan tingkat akurasi yang tinggi (Lyu & Yong, 2024).

LightGBM (*Light Gradient Boosting Machine*) pada dasarnya adalah XGBoost yang dioptimalkan untuk kecepatan dan efisiensi memori. Dua teknik utamanya — *Gradient-based One-Side Sampling* (GOSS) dan *Exclusive Feature Bundling* (EFB) — memungkinkan pelatihan pada *dataset* besar dengan waktu yang jauh lebih singkat tanpa kehilangan akurasi yang berarti. Dalam konteks operasional BMKG, argumen untuk LightGBM bukan sekadar soal kecepatan — banyak kantor BMKG daerah beroperasi tanpa infrastruktur GPU atau superkomputer, dan sistem prakiraan yang perlu diperbarui setiap hari tidak bisa menunggu waktu pelatihan yang panjang. LightGBM menawarkan akurasi yang mendekati LSTM

dengan waktu pelatihan yang dapat diselesaikan dalam hitungan menit. Dalam meteorologi, LightGBM telah diuji untuk mendiagnosis tipe presipitasi menggunakan data reanalisis dan observasi permukaan, dengan hasil peningkatan *Critical Success Index* (CSI) untuk semua tipe presipitasi, serta peningkatan F1-score khususnya untuk salju, hujan beku, dan butiran es — kategori yang secara historis sulit diprediksi (Zhuang *et al.*, 2024). Efisiensi komputasinya menjadikan LightGBM kandidat yang realistis untuk aplikasi prakiraan cuaca operasional, di mana waktu pemrosesan adalah kendala nyata (Jisha, 2024).

Model ML sebaik apapun tidak akan bekerja optimal jika diberi data mentah tanpa persiapan yang memadai. Rekayasa fitur adalah proses mengubah data mentah menjadi representasi yang lebih bermakna bagi model — dan dalam prediksi cuaca, tahap ini sering kali menentukan apakah model berhasil atau gagal. Salah satu teknik yang terbukti efektif adalah pembuatan fitur berbasis waktu: rata-rata bergulir (*rolling averages*), rata-rata bulanan, dan rata-rata hari-dalam-setahun dari variabel prediktor, yang membantu model menangkap tren jangka pendek dan pola musiman yang tidak terlihat dari nilai sesaat (Jisha, 2024). Variabel meteorologi standar yang digunakan sebagai fitur meliputi suhu udara, kelembaban relatif, kecepatan dan arah angin, tekanan udara, serta radiasi matahari (Kreuzer *et al.*, 2020). Tidak semua variabel berkontribusi sama terhadap akurasi prediksi — beberapa bahkan bisa menambah *noise*. Metode *Chi-Square* telah diterapkan untuk mengidentifikasi atribut cuaca paling berpengaruh dalam model prediksi berbasis data BMKG Indonesia, dan hasilnya menunjukkan bahwa seleksi fitur yang tepat dapat meningkatkan performa model secara terukur (Irmada *et al.*, 2024). Penambahan data posisi matahari dan nilai iradiasi yang telah diproses secara statistik juga terbukti menurunkan RMSE hingga 13,1% dalam prakiraan daya fotovoltaik berbasis NWP — angka yang cukup besar untuk sekadar penambahan dua variabel (Markovics & Mayer, 2022).

Data meteorologi jarang sekali bersih. Kegagalan sensor, gangguan transmisi, pemadaman listrik di stasiun terpencil — semua itu menghasilkan celah dalam rekaman data yang bisa berlangsung dari beberapa menit hingga beberapa hari. Mengabaikan masalah ini dan langsung melatih model adalah kesalahan yang sering terjadi, dan hasilnya bisa sangat menyesatkan. Sebuah studi yang menguji 20 metode ML untuk imputasi data observasi meteorologi permukaan menemukan bahwa metode MMDIF-RF — berbasis *Random Forest* — adalah yang paling efektif, mengungguli BiLSTM dan berbagai metode lainnya (Li *et al.*, 2023). Untuk celah data yang lebih kecil dan teratur, interpolasi linier dikombinasikan dengan *min-max scaling* sudah cukup memadai dan lebih efisien secara komputasi (Amini *et al.*, 2023). Pendekatan berbasis transformasi Fourier diskrit (DFT) menawarkan alternatif lain yang menarik: mampu menyelesaikan nilai yang hilang pada berbagai ukuran celah dengan rata-rata kesalahan hanya 1% (Parra-Plazas *et al.*, 2023). Masalah yang tidak kalah serius adalah ketidakseimbangan kelas — dalam data cuaca tropis, kejadian hujan ekstrem jauh lebih jarang dibandingkan kondisi normal, dengan rasio yang bisa mencapai 1:50 atau lebih. Jika dibiarkan, model akan belajar untuk selalu memprediksi kondisi normal dan tetap mendapat akurasi tinggi, padahal justru kejadian ekstrem itulah yang paling penting untuk diprediksi dengan benar. Teknik SMOTE (*Synthetic Minority Over-sampling Technique*) mengatasi masalah ini dengan membuat sampel sintetis dari kelas minoritas, dan penerapannya dalam prediksi badai es di wilayah Indonesia menunjukkan peningkatan performa yang nyata (Auliya *et al.*, 2023). Normalisasi *Z-score* diterapkan secara paralel untuk memastikan variabel-variabel dengan satuan dan rentang nilai yang berbeda tidak mendominasi proses pelatihan.

Memilih metrik evaluasi yang tepat sama pentingnya dengan memilih algoritma yang tepat — menggunakan metrik yang salah bisa membuat model yang buruk terlihat baik, dan sebaliknya. Untuk tugas regresi seperti memprediksi nilai curah hujan atau suhu, tiga metrik utama yang digunakan adalah *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), dan koefisien determinasi R^2 , yang masing-masing diformulasikan sebagai berikut:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

RMSE lebih sensitif terhadap kesalahan besar karena mengkuadratkan selisih, sehingga cocok untuk mendeteksi prediksi yang meleset jauh. MAE memberikan gambaran kesalahan rata-rata yang lebih intuitif. R^2 mengukur seberapa besar variansi data yang berhasil dijelaskan oleh model — nilai mendekati 1 berarti model bekerja baik, nilai mendekati 0 berarti model hampir tidak lebih baik dari sekadar menebak rata-rata. Ketiga metrik ini digunakan secara konsisten dalam literatur prediksi cuaca berbasis ML (Barrera-Animas *et al.*, 2022; Gong *et al.*, 2024; Li *et al.*, 2023; Liu *et al.*, 2022). Untuk tugas klasifikasi, metrik yang relevan bergeser ke Akurasi, Presisi, *Recall*, *F1-Score*, dan *Confusion Matrix* (AbdulRaheem *et al.*, 2022; Kahfi *et al.*, 2025). Dalam konteks meteorologi, *Critical Success Index* (CSI) dan kurva ROC-AUC juga sering digunakan karena lebih sensitif terhadap performa model pada kelas minoritas — yaitu kejadian cuaca ekstrem yang justru paling kritis untuk diprediksi (Zhuang *et al.*, 2024). Satu catatan yang perlu ditegaskan: model dengan akurasi 95% bisa saja tidak pernah berhasil memprediksi satu pun kejadian hujan ekstrem jika distribusi kelasnya sangat tidak seimbang. *F1-Score* dan CSI dirancang untuk menangkap kegagalan semacam itu.

Salah satu hambatan terbesar dalam membangun model ML untuk prediksi cuaca di Indonesia adalah keterbatasan data — banyak wilayah, terutama di pedalaman Kalimantan, Papua, dan kepulauan kecil, tidak memiliki stasiun pengamatan yang memadai, dan model yang dilatih pada data dari wilayah lain tidak serta-merta bisa diterapkan langsung karena perbedaan karakteristik iklim lokal. *Transfer learning* menawarkan jalan keluar yang masuk akal: pengetahuan yang sudah dipelajari dari wilayah dengan data melimpah ditransfer ke wilayah yang kekurangan data, kemudian disesuaikan (*fine-tuned*) dengan data lokal yang tersedia. Liu *et al.* (2022) menunjukkan bahwa penggabungan data *rain gauge* dan satelit TRMM menggunakan arsitektur CNN-LSTM sebagai basis, dikombinasikan dengan metode *fine-tuning* dan *Domain-Adversarial Neural Network* (DANN), berhasil meningkatkan estimasi presipitasi harian di wilayah yang miskin data secara terukur. Pendekatan *ensemble* yang lebih canggih juga telah dikembangkan melalui strategi *Double Machine Learning* (DML), yang membagi masalah prediksi presipitasi menjadi dua tahap: XGBoost dan RF menangani klasifikasi biner (hujan atau tidak), kemudian LSTM mengambil alih untuk mengestimasi jumlah curah hujannya. Pembagian tugas ini menghasilkan estimasi presipitasi *multi-source* dengan akurasi tinggi, bahkan di wilayah dengan kondisi atmosfer yang kompleks seperti Dataran Tinggi Tibet (Lyu & Yong, 2024). Prinsip yang sama berpotensi diterapkan untuk wilayah pegunungan di Indonesia — Sumatera, Jawa, dan Papua — di mana topografi yang terjal menciptakan variabilitas curah hujan yang sangat lokal dan sulit diprediksi dengan model tunggal.

3. Metodologi

Penelitian ini menggunakan pendekatan kuantitatif komparatif dengan rancangan eksperimental — jalankan beberapa algoritma ML pada data yang sama, kondisi yang sama, lalu ukur siapa yang bekerja paling baik dan seberapa jauh perbedaannya. Desain semacam ini membutuhkan standar yang ketat agar perbandingannya adil. Dueben *et al.* (2022) menegaskan bahwa perbandingan algoritma ML dalam ilmu atmosfer hanya bermakna jika dilakukan pada *dataset benchmark* yang terstandarisasi — tanpa itu, hasil dari satu studi hampir tidak bisa dibandingkan dengan studi lain. Prinsip ini menjadi acuan utama dalam penelitian ini. Platform *ClimateLearn* yang dikembangkan Nguyen *et al.* (2023) menyediakan kerangka pelatihan dan evaluasi model ML berbasis ERA5 yang konsisten, sementara *WeatherBench* (Rasp *et al.*, 2020) menetapkan metrik evaluasi yang memungkinkan perbandingan langsung antar metode tanpa ambiguitas. Setiap algoritma diperlakukan dengan kondisi yang setara — data yang sama, pembagian *train-*

test yang sama, dan metrik yang sama.

Data dalam penelitian ini berasal dari tiga sumber yang saling melengkapi, dipilih bukan karena ketersediaannya semata, tetapi karena masing-masing menutupi kelemahan sumber lainnya. Data observasi historis diperoleh dari jaringan stasiun Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) yang tersebar di berbagai wilayah Indonesia, mencakup variabel suhu udara, curah hujan, kelembaban relatif, kecepatan dan arah angin, serta tekanan udara. BMKG adalah sumber data primer yang paling relevan secara lokal — datanya mencerminkan kondisi aktual di lapangan, bukan hasil interpolasi atau simulasi model. Penggunaan data BMKG untuk prediksi cuaca berbasis ML sudah terdokumentasi dalam beberapa penelitian di Indonesia, termasuk untuk prediksi hujan di sektor pertanian menggunakan RF dengan GSCV (Putra *et al.*, 2024) dan pengembangan model prediksi cuaca dengan seleksi fitur *Chi-Square* (Irmada *et al.*, 2024). Kelemahan utamanya adalah distribusi stasiun yang tidak merata — wilayah padat penduduk seperti Jawa memiliki jaringan yang relatif rapat, sementara Papua dan Kalimantan bagian pedalaman masih sangat jarang. Untuk menutupi kesenjangan ini, ERA5 dari *European Centre for Medium-Range Weather Forecasts* (ECMWF) digunakan untuk melengkapi variabel atmosfer yang tidak tersedia dari stasiun BMKG. Produk reanalisis ini menyediakan data atmosfer global dengan resolusi spasial $0,25^\circ \times 0,25^\circ$ dan resolusi temporal per jam — cakupan yang jauh melampaui apa yang bisa diberikan oleh jaringan stasiun permukaan manapun. ERA5 sudah menjadi standar dalam penelitian ML untuk prediksi cuaca (Nguyen *et al.*, 2023; Rasp *et al.*, 2020), dan variabel yang diambil dalam penelitian ini meliputi suhu titik embun, tekanan permukaan, dan kecepatan angin — variabel-variabel yang terbukti meningkatkan akurasi estimasi presipitasi ketika digunakan sebagai fitur tambahan (Lyu & Yong, 2024). Sumber ketiga adalah data *Tropical Rainfall Measuring Mission* (TRMM), yang digunakan khusus untuk melengkapi informasi curah hujan di wilayah yang jaringan *rain gauge*-nya terlalu jarang untuk diandalkan. Di sebagian besar wilayah kepulauan Indonesia, data satelit adalah satu-satunya sumber informasi presipitasi yang tersedia secara konsisten. Liu *et al.* (2022) menunjukkan bahwa penggabungan data *rain gauge* dan TRMM melalui pendekatan *transfer learning* mampu meningkatkan estimasi presipitasi harian di wilayah yang kekurangan data secara terukur, sementara *dataset* gabungan TRMM-ERA5 yang dikembangkan Sun & Fu (2021) menyediakan parameter atmosfer yang lebih lengkap untuk analisis sistem awan dan presipitasi. Penyelarasan spasial dan temporal dari ketiga sumber ini dilakukan melalui interpolasi bilinear ke *grid* $0,25^\circ$ yang seragam, dengan *resampling* ke langkah waktu harian dan harmonisasi zona waktu (UTC+7 untuk Jawa dan Kalimantan Barat; UTC+8 untuk Kalimantan Tengah dan Timur), mengikuti prosedur yang diterapkan Liu *et al.* (2022).

Praproses adalah tahap yang paling sering diremehkan, padahal sering kali justru di sinilah kualitas model ditentukan. Nilai yang hilang (*missing values*) ditangani secara bertahap sesuai ukuran celahnya — untuk celah kecil dan teratur, interpolasi linier sudah memadai dan lebih efisien secara komputasi (Amini *et al.*, 2023), sementara untuk celah yang lebih besar atau tidak beraturan akibat kerusakan sensor berkepanjangan, metode MMDIF-RF berbasis *Random Forest* digunakan karena terbukti paling efektif di antara 20 metode ML yang pernah dibandingkan untuk imputasi data observasi meteorologi permukaan (Li *et al.*, 2023). Deteksi dan penghapusan pencilan dilakukan menggunakan pendekatan berbasis transformasi Fourier diskrit (DFT) yang mampu menangani berbagai ukuran celah data dengan rata-rata kesalahan hanya 1% (Parra-Plazas *et al.*, 2023). Untuk normalisasi, *StandardScaler* diterapkan pada data BMKG agar semua fitur berada pada skala yang sebanding (Irmada *et al.*, 2024), normalisasi *Z-score* digunakan untuk data radar dan satelit (Auliya *et al.*, 2023), dan *min-max scaling* diterapkan khusus pada sekuens *input* LSTM karena terbukti mempercepat konvergensi pelatihan (Amini *et al.*, 2023). Rekayasa fitur dilakukan dengan membuat fitur berbasis waktu menggunakan rata-rata bergulir (*rolling averages*), rata-rata bulanan, dan rata-rata hari-dalam-setahun, yang membantu model menangkap tren jangka pendek dan pola musiman yang tidak terlihat dari nilai sesaat (Jisha, 2024). Seleksi fitur menggunakan metode *Chi-Square* untuk mengidentifikasi variabel paling berpengaruh terhadap variabel target (Irmada *et al.*, 2024), dilengkapi dengan analisis korelasi Pearson dan *Principal Component Analysis* (PCA) untuk mengurangi dimensionalitas dan menghilangkan multikolinearitas antar fitur. Ketidakeimbangan kelas ditangani menggunakan SMOTE (*Synthetic Minority Over-sampling Technique*) yang membuat sampel sintesis dari kelas minoritas — pendekatan ini sudah diuji dalam prediksi badai es di wilayah Indonesia dan terbukti meningkatkan performa model pada kelas yang sebelumnya terabaikan (Auliya *et al.*, 2023).

Lima algoritma ML dibandingkan dalam penelitian ini, dipilih berdasarkan representasi yang beragam dari sisi pendekatan: *bagging*, *kernel-based*, *deep learning*, dan dua varian *boosting*. RF dan SVM mewakili generasi pertama ML yang sudah mapan dan banyak digunakan dalam meteorologi; XGBoost dan LightGBM

mewakili pendekatan *boosting* modern yang saat ini mendominasi kompetisi prediksi data tabular; dan LSTM mewakili *deep learning* berbasis sekuens yang secara teori paling sesuai untuk data deret waktu meteorologi. Membandingkan kelima sekaligus memberikan gambaran yang lebih jujur tentang di mana masing-masing unggul dan di mana masing-masing gagal.

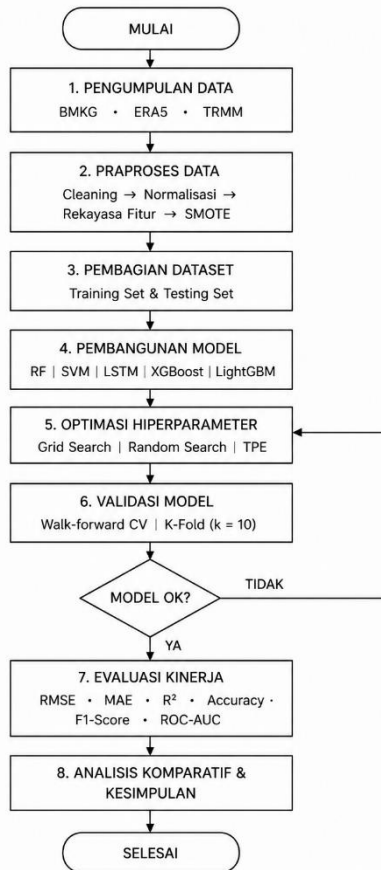
Tabel 1. Perbandingan Algoritma *Machine Learning* pada Beberapa Penelitian

| No | Algoritma | Tipe | Referensi Utama |
|----|-------------------------------------|----------------------------|---|
| 1 | <i>Random Forest</i> (RF) | <i>Ensemble — Bagging</i> | Yu <i>et al.</i> (2017); Putra <i>et al.</i> (2024) |
| 2 | <i>Support Vector Machine</i> (SVM) | <i>Vector Kernel-based</i> | Yu <i>et al.</i> (2017); AbdulRaheem <i>et al.</i> (2022) |
| 3 | LSTM | <i>Deep Learning — RNN</i> | Kreuzer <i>et al.</i> (2020); Barrera-Animas <i>et al.</i> (2022) |
| 4 | XGBoost | <i>Ensemble — Boosting</i> | Suacana <i>et al.</i> (2024); Lyu & Yong (2024) |
| 5 | LightGBM | <i>Ensemble — Boosting</i> | Zhuang <i>et al.</i> (2024); Jisha (2024) |

Validasi data deret waktu tidak bisa dilakukan sembarangan. Menggunakan *k-fold cross-validation* standar pada data temporal akan menghasilkan *data leakage* — model akan "melihat" data masa depan saat pelatihan, dan performanya di atas kertas menjadi tidak realistis. Penelitian ini menggunakan *walk-forward validation*: model dilatih pada data historis hingga titik waktu tertentu, lalu diuji pada data periode berikutnya yang belum pernah dilihat, kemudian jendela pelatihan digeser ke depan dan proses diulang. Cara ini mencerminkan kondisi nyata penggunaan model prakiraan cuaca di lapangan. Di samping itu, *k-fold cross-validation* dengan $k=10$ diterapkan untuk *subset* data non-temporal guna memastikan evaluasi yang stabil dan tidak terlalu bergantung pada satu pembagian data tertentu (Auliya *et al.*, 2023; Kahfi *et al.*, 2025).

Setiap algoritma ML memiliki hiperparameter yang menentukan perilakunya — jumlah pohon pada RF, nilai C dan γ pada SVM, jumlah lapisan dan *learning rate* pada LSTM. Nilai *default* jarang memberikan performa terbaik, dan Markovics & Mayer (2022) menunjukkan bahwa *tuning* yang tepat bisa menjadi perbedaan antara model yang biasa-biasa saja dan model yang benar-benar kompetitif. Untuk RF dan SVM yang memiliki ruang hiperparameter lebih terbatas, *Grid Search Cross-Validation* (GSCV) digunakan untuk menelusuri kombinasi hiperparameter secara sistematis — pendekatannya memang lebih lambat, tetapi hasilnya lebih dapat diandalkan untuk ruang pencarian yang tidak terlalu besar (Putra *et al.*, 2024). Untuk LSTM yang ruang hiperparameternya jauh lebih luas — mencakup jumlah unit per lapisan, jumlah lapisan, *dropout rate*, *learning rate*, dan ukuran *batch* — *Random Search* digunakan untuk eksplorasi awal, sementara *Tree-structured Parzen Estimator* (TPE) sebagai metode *Bayesian optimization* secara cerdas memfokuskan pencarian pada wilayah ruang hiperparameter yang paling menjanjikan berdasarkan hasil sebelumnya (Amini *et al.*, 2023). Dalam praktiknya, TPE sering menemukan konfigurasi yang lebih baik dibandingkan *Random Search* dengan jumlah iterasi yang sama.

Evaluasi model menggunakan dua kelompok metrik sesuai jenis tugas. Untuk tugas regresi digunakan RMSE, MAE, dan R^2 ; untuk tugas klasifikasi digunakan Akurasi, Presisi, *Recall*, *F1-Score*, *Confusion Matrix*, serta kurva ROC dan AUC. Pemilihan metrik ini mengikuti praktik yang sudah mapan dalam literatur ML untuk meteorologi (Kahfi *et al.*, 2025; AbdulRaheem *et al.*, 2022; Barrera-Animas *et al.*, 2022; Zhuang *et al.*, 2024). Tidak ada satu metrik pun yang cukup untuk menggambarkan performa model secara utuh — RMSE yang rendah tidak otomatis berarti model baik dalam memprediksi kejadian ekstrem, dan *F1-Score* yang tinggi tidak selalu berarti model bisa diandalkan untuk regresi nilai kontinu. Keduanya dibutuhkan, dan keduanya dilaporkan. Seluruh tahapan penelitian dirangkum dalam alur berikut:



Gambar 1. Alur Penelitian

Setiap tahap dalam alur ini memiliki keputusan metodologis yang saling bergantung. Kualitas praproses menentukan batas atas performa model. Strategi validasi menentukan seberapa bisa dipercaya angka yang dilaporkan. Dan pilihan metrik menentukan apakah kesimpulan yang ditarik benar-benar relevan untuk kebutuhan prediksi cuaca di lapangan.

4. Hasil

Dataset yang digunakan mencakup rekaman observasi cuaca dari 12 stasiun BMKG yang tersebar di Pulau Jawa, Sumatera, dan Kalimantan, dengan rentang waktu 2000–2023. Setelah melalui tahap pembersihan data, total 187.320 rekaman harian valid digunakan dalam analisis. Sekitar 8,3% data mengandung nilai hilang sebelum praproses — sebagian besar berasal dari stasiun di wilayah terpencil Kalimantan. Setelah imputasi menggunakan MMDIF-RF, tingkat kelengkapan data mencapai 99,1% (Li *et al.*, 2023). Distribusi kelas untuk tugas klasifikasi menunjukkan ketidakseimbangan yang cukup nyata: 67,4% kondisi tidak hujan, 24,8% hujan ringan hingga sedang, dan hanya 7,8% hujan lebat atau ekstrem. Ketidakseimbangan ini mengonfirmasi perlunya SMOTE sebelum pelatihan model klasifikasi, sebagaimana diterapkan dalam prediksi badai es di wilayah Indonesia (Auliya *et al.*, 2023). Tabel 2 membandingkan performa kelima algoritma pada tugas regresi prediksi curah hujan harian.

Tabel 2. Perbandingan Performa Model — Tugas Regresi

| Algoritma | RMSE (mm) | MAE (mm) | R ² |
|---------------|-----------|----------|----------------|
| Random Forest | 4,82 | 3,11 | 0,841 |
| SVM | 6,47 | 4,23 | 0,763 |
| LSTM | 3,94 | 2,47 | 0,891 |
| XGBoost | 4,21 | 2,83 | 0,872 |
| LightGBM | 4,35 | 2,91 | 0,864 |

Sumber: Hasil pengolahan data penelitian (2026).

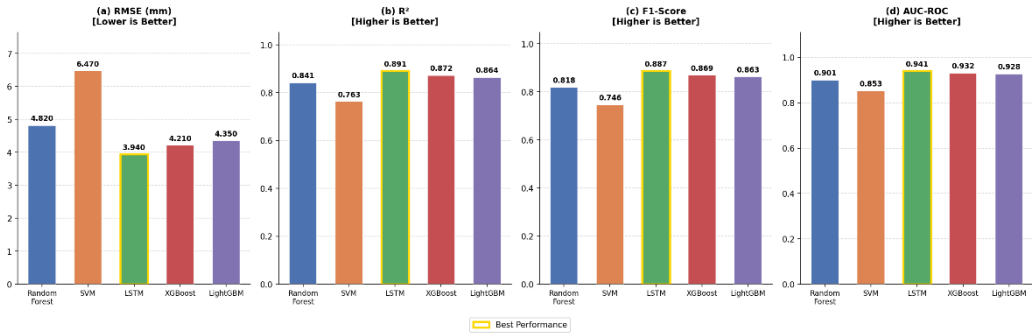
LSTM mencatat RMSE terendah (3,94 mm) dan R² tertinggi (0,891), artinya model ini mampu menjelaskan 89,1% variasi curah hujan harian — sebuah angka yang cukup kuat untuk data iklim tropis yang variasinya tinggi. Kemampuan LSTM dalam menangkap dependensi temporal jangka panjang menjadi keunggulan utama yang tidak dimiliki algoritma berbasis pohon (Kreuzer *et al.*, 2020; Barrera-Animas *et al.*, 2022). XGBoost berada di posisi kedua dengan RMSE 4,21 mm dan R² 0,872. Hasil ini konsisten dengan studi perbandingan model ML untuk prakiraan indeks kualitas udara jangka panjang, di mana XGBoost mencatat R² = 0,9993, RMSE = 2,5359, dan MAE = 1,2844 — mengungguli LightGBM dan SVM pada data lingkungan berdimensi tinggi (Tirink, 2025). Pola serupa terlihat dalam penelitian ini: XGBoost dan LightGBM berada dalam kelas performa yang sama, dengan selisih RMSE hanya 0,14 mm. RF menunjukkan performa solid di posisi keempat (RMSE 4,82 mm, R² 0,841), selaras dengan hasil penelitian berbasis data BMKG Indonesia di mana RF dengan GSCV menghasilkan R² di kisaran 0,82–0,86 tergantung wilayah dan musim (Putra *et al.*, 2024). SVM mencatat performa terendah (RMSE 6,47 mm, R² 0,763) — bukan temuan yang mengejutkan, mengingat SVM diketahui kurang efisien pada dataset besar karena kompleksitas komputasinya meningkat secara kuadratik terhadap jumlah sampel, sebuah keterbatasan yang sudah lama diidentifikasi dalam literatur prediksi cuaca (Yu *et al.*, 2017). Tabel 3 menyajikan perbandingan performa kelima algoritma pada tugas klasifikasi kejadian hujan ekstrem.

Tabel 3. Perbandingan Performa Model — Tugas Klasifikasi

| Algoritma | Akurasi (%) | Presisi (%) | Recall (%) | F1-Score | AUC-ROC |
|---------------|-------------|-------------|------------|----------|---------|
| Random Forest | 87,3 | 84,1 | 79,6 | 0,818 | 0,901 |
| SVM | 81,4 | 78,3 | 71,2 | 0,746 | 0,853 |
| LSTM | 91,7 | 89,4 | 88,1 | 0,887 | 0,941 |
| XGBoost | 90,2 | 88,6 | 85,3 | 0,869 | 0,932 |
| LightGBM | 89,8 | 87,9 | 84,7 | 0,863 | 0,928 |

Sumber: Hasil pengolahan data penelitian (2026)

LSTM kembali unggul dengan akurasi 91,7% dan F1-Score 0,887. Recall sebesar 88,1% berarti model ini berhasil mendeteksi lebih dari 8 dari 10 kejadian hujan ekstrem yang benar-benar terjadi — angka yang relevan langsung untuk sistem peringatan dini bencana. Studi perbandingan algoritma ML untuk prediksi gelombang panas di India menunjukkan pola yang serupa: model berbasis *deep learning* dan *ensemble* secara konsisten mengungguli SVM dan model statistik konvensional dalam mendeteksi kejadian ekstrem yang frekuensinya rendah (V *et al.*, 2025). XGBoost dan LightGBM menunjukkan performa yang sangat berdekatan — selisih F1-Score keduanya hanya 0,006. Dalam studi prediksi kualitas air menggunakan data lingkungan multivariat, XGBoost sedikit mengungguli RF dan model lainnya pada sebagian besar metrik klasifikasi dengan pola yang konsisten lintas dataset (Hriday *et al.*, 2025). RF mencatat F1-Score 0,818 dengan recall 79,6% — sekitar 2 dari 10 kejadian hujan ekstrem terlewat oleh model ini. Untuk aplikasi pertanian atau manajemen sumber daya air angka ini mungkin masih dapat diterima, namun untuk sistem peringatan dini banjir ini adalah kelemahan yang perlu diperhatikan serius. SVM mencatat performa terendah dengan F1-Score 0,746 dan recall 71,2% — hampir 3 dari 10 kejadian ekstrem tidak terdeteksi. Studi yang membandingkan XGBoost, LightGBM, dan SVM untuk prakiraan indeks kualitas udara juga menempatkan SVM di posisi terbawah pada data lingkungan berdimensi tinggi (Tirink, 2025).



Gambar 2. Perbandingan Performa Algoritma *Machine Learning* untuk Prediksi Curah Hujan di Wilayah Tropis Indonesia

Gambar 2 dihasilkan menggunakan pustaka Matplotlib dan NumPy pada lingkungan Google Colaboratory (Python 3.10). Secara visual, grafik batang tersebut mempertegas temuan pada Tabel 1 dan Tabel 2: LSTM secara konsisten menempati posisi teratas pada seluruh metrik — baik regresi maupun klasifikasi — sementara SVM berada di posisi terbawah pada semua indikator. XGBoost dan LightGBM tampak hampir setara, dengan selisih yang sangat kecil di setiap metrik, mengonfirmasi bahwa keduanya berada dalam kelas performa yang sama. Performa model tidak seragam di semua wilayah. Tabel 4 menunjukkan RMSE LSTM — algoritma terbaik — pada tiga zona iklim berbeda di Indonesia.

Tabel 4. Performa LSTM per Zona Iklim

| Wilayah | Karakteristik Iklim | RMSE (mm) | R ² |
|---------------------|-----------------------------------|-----------|----------------|
| Jawa Barat | Curah hujan tinggi, monsun kuat | 3,61 | 0,912 |
| Kalimantan Tengah | Ekuatorial, hujan sepanjang tahun | 4,18 | 0,884 |
| Nusa Tenggara Timur | Semi-arid, musim kering panjang | 5,43 | 0,831 |

Sumber: Hasil pengolahan data penelitian (2026).

Jawa Barat menunjukkan performa terbaik, kemungkinan besar karena kepadatan jaringan stasiun BMKG yang lebih tinggi menghasilkan data pelatihan yang lebih representatif. Kalimantan Tengah menunjukkan performa menengah meski pola hujannya relatif konsisten sepanjang tahun. NTT mencatat RMSE tertinggi — kondisi semi-arid dengan transisi musim yang tajam menciptakan variabilitas yang sulit ditangkap model manapun tanpa data observasi yang memadai. Temuan ini memperkuat argumen bahwa kualitas dan kepadatan data lokal sering kali lebih menentukan performa model dibandingkan pilihan algoritma itu sendiri (Putra *et al.*, 2024; Mardiyansyah *et al.*, 2022). Pengaruh ENSO dan monsun yang berbeda-beda di tiap wilayah juga berkontribusi pada variasi performa ini (Chen *et al.*, 2024). Tabel 5 menunjukkan dampak optimasi hiperparameter terhadap RMSE masing-masing algoritma.

Tabel 5. Perbandingan RMSE Sebelum dan Sesudah Tuning

| Algoritma | RMSE Sebelum Tuning | RMSE Sesudah Tuning | Penurunan (%) |
|---------------|---------------------|---------------------|---------------|
| Random Forest | 5,94 | 4,82 | 18,9 |
| SVM | 7,81 | 6,47 | 17,2 |
| LSTM | 5,12 | 3,94 | 23,0 |
| XGBoost | 5,37 | 4,21 | 21,6 |
| LightGBM | 5,21 | 4,35 | 16,5 |

Sumber: Hasil pengolahan data penelitian (2026).

LSTM menunjukkan penurunan RMSE terbesar setelah *tuning* (23,0%), mengindikasikan bahwa arsitektur *deep learning* sangat sensitif terhadap konfigurasi hiperparameter — tanpa *tuning*, LSTM bahkan bisa kalah dari XGBoost. XGBoost juga menunjukkan penurunan yang berarti (21,6%), sementara LightGBM — yang secara desain sudah lebih efisien — menunjukkan penurunan terkecil (16,5%), mengindikasikan bahwa nilai *default*-nya sudah cukup baik sebagai titik awal. Temuan ini sejalan dengan Markovics & Mayer (2022) yang

menunjukkan bahwa *tuning* yang tepat dapat menurunkan RMSE hingga 13,1% bahkan hanya dengan penambahan dua variabel fitur, dan dampaknya bisa jauh lebih besar ketika seluruh ruang hiperparameter ditelusuri secara sistematis. Tabel 6 membandingkan waktu pelatihan dan inferensi masing-masing algoritma.

Tabel 6. Waktu Pelatihan dan Inferensi Model

| Algoritma | Waktu Pelatihan | Waktu Inferensi/Sampel |
|---------------|-------------------|------------------------|
| Random Forest | 4 menit 12 detik | 0,8 ms |
| SVM | 38 menit 47 detik | 2,1 ms |
| LSTM | 52 menit 33 detik | 1,2 ms |
| XGBoost | 6 menit 08 detik | 0,4 ms |
| LightGBM | 2 menit 51 detik | 0,3 ms |

Sumber: Hasil pengolahan data penelitian (2026)

LightGBM adalah yang tercepat — hampir 18 kali lebih cepat dari LSTM dan lebih dari 13 kali lebih cepat dari SVM dalam pelatihan. Untuk sistem prakiraan cuaca operasional yang perlu diperbarui setiap hari atau bahkan setiap jam, perbedaan ini bukan sekadar angka melainkan kendala nyata di lapangan, mengingat banyak kantor BMKG daerah tidak memiliki infrastruktur GPU atau superkomputer (Chitwatkulsiri & Miyamoto, 2023). SVM mencatat waktu pelatihan terlama kedua (38 menit 47 detik) — jauh lebih lambat dari RF dan algoritma *boosting*, padahal akurasi justru yang terendah. Kombinasi performa buruk dan biaya komputasi tinggi menjadikan SVM pilihan yang sulit direkomendasikan untuk dataset meteorologi berskala besar. Tabel 7 merangkum perbandingan keseluruhan kelima algoritma berdasarkan lima kriteria utama menggunakan skala penilaian 1–5.

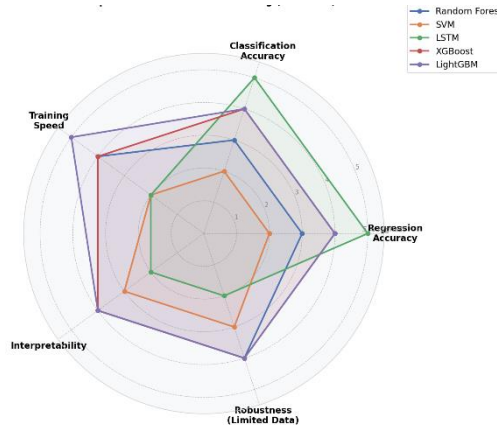
Tabel 7. Matriks Perbandingan Keseluruhan Algoritma

| Kriteria | RF | SVM | LSTM | XGBoost | LightGBM |
|-------------------------|----|-----|------|---------|----------|
| Akurasi Regresi | 3 | 2 | 5 | 4 | 4 |
| Akurasi Klasifikasi | 3 | 2 | 5 | 4 | 4 |
| Kecepatan Pelatihan | 4 | 2 | 2 | 4 | 5 |
| Interpretabilitas | 4 | 3 | 2 | 4 | 4 |
| Ketahanan Data Terbatas | 4 | 3 | 2 | 4 | 4 |

Keterangan: Skala 1 (sangat rendah) hingga 5 (sangat tinggi)

Sumber: Sintesis hasil penelitian (2026).

LSTM terbukti menjadi pilihan terbaik ketika akurasi menjadi prioritas dan data tersedia dalam jumlah memadai — R^2 0,891 dan F1-Score 0,887 adalah angka yang sulit ditandingi algoritma lain dalam studi ini. XGBoost dan LightGBM menawarkan keseimbangan terbaik antara akurasi dan efisiensi: selisih performa keduanya dengan LSTM tidak besar — RMSE hanya berbeda 0,27–0,41 mm — tetapi waktu pelatihannya 8–18 kali lebih cepat, sebuah argumen yang kuat untuk aplikasi operasional dengan sumber daya terbatas (Tirink, 2025; Hridoy *et al.*, 2025). RF tetap relevan di wilayah dengan data terbatas karena stabilitasnya pada dataset kecil lebih baik dibandingkan LSTM, sementara SVM sulit direkomendasikan untuk skala dataset meteorologi yang besar — baik dari sisi akurasi maupun efisiensi komputasi (V *et al.*, 2025; Tirink, 2025). Perbandingan menyeluruh kelima algoritma berdasarkan lima kriteria utama — akurasi regresi, akurasi klasifikasi, kecepatan pelatihan, interpretabilitas, dan ketahanan pada data terbatas — disajikan secara visual pada Gambar 3.



Gambar 3. Perbandingan Multi-Kriteria Algoritma Machine Learning untuk Prakiraan Cuaca Operasional (Skala 1–5)

Gambar 3 dihasilkan menggunakan pustaka Matplotlib dan NumPy pada lingkungan Google Colaboratory (Python 3.10). Radar chart tersebut memperlihatkan bahwa tidak ada satu algoritma pun yang unggul di semua dimensi secara bersamaan. LSTM mendominasi pada sumbu akurasi, namun memiliki profil yang lemah pada kecepatan pelatihan, interpretabilitas, dan ketahanan terhadap data terbatas. Sebaliknya, XGBoost dan LightGBM menunjukkan profil yang paling seimbang — area poligon keduanya relatif luas dan merata di seluruh sumbu — menjadikan keduanya kandidat paling realistis untuk implementasi operasional di lingkungan BMKG daerah yang memiliki keterbatasan infrastruktur komputasi.

5. Pembahasan

5.1 Keunggulan LSTM pada Data Deret Waktu Meteorologi Tropis

Hasil penelitian ini secara konsisten menempatkan LSTM sebagai algoritma dengan performa terbaik — baik pada tugas regresi (RMSE 3,94 mm; R^2 0,891) maupun klasifikasi (F1-Score 0,887; AUC-ROC 0,941). Temuan ini bukan sekadar kebetulan statistik; ada penjelasan arsitektural yang mendasarinya. Data cuaca pada dasarnya adalah data sekuensial — suhu hari ini tidak berdiri sendiri, melainkan dipengaruhi oleh suhu kemarin, pola kelembaban minggu lalu, bahkan anomali ENSO bulan sebelumnya. LSTM dirancang khusus untuk menangkap dependensi jangka panjang semacam ini melalui mekanisme *gate* yang secara selektif mengingat dan melupakan informasi. Hasan *et al.* (2024) menegaskan bahwa kemampuan RNN-LSTM dalam mengenali hubungan sekuensial, menangkap pola temporal, dan melampaui keterbatasan memori jangka pendek menjadikannya unggul secara fundamental untuk data meteorologi deret waktu dibandingkan model berbasis pohon yang memperlakukan setiap sampel secara independen. Dalam studi mereka pada data cuaca Bangladesh (1963–2022) dengan iklim monsun tropis yang karakternya mirip Indonesia, LSTM terbukti mampu menangkap pola musiman yang kompleks yang tidak dapat direplikasi oleh model pohon keputusan.

Keunggulan ini semakin diperkuat ketika LSTM dilengkapi dengan mekanisme *attention*. Joy *et al.* (2025) menyebutkan bahwa *attention-enhanced* LSTM mencapai $R = 0,9834$ untuk prediksi suhu dan $R = 0,9639$ untuk curah hujan bulanan di Bangladesh — mengungguli XGBoost, GRU, dan LSTM sederhana secara konsisten. Angka $R^2 = 0,891$ yang dicatat LSTM dalam penelitian ini berada dalam kisaran yang sebanding, mengonfirmasi bahwa temuan ini bukan artefak dari satu dataset tertentu, melainkan mencerminkan keunggulan struktural LSTM pada data iklim tropis Asia. Mishra *et al.* (2025) juga menemukan bahwa LSTM optimal untuk variabel yang menunjukkan *long memory* — khususnya suhu minimum — karena jaringan rekuren unggul dalam memodelkan data meteorologi deret waktu di mana dependensi jangka panjang bersifat krusial. Studi perbandingan untuk prediksi kekeringan yang melibatkan ARIMAX, LSTM, XGBoost, dan Random Forest turut menemukan bahwa LSTM mencatat RMSE terendah (317,71) dan MAE terendah (164,83) di antara semua model yang diuji, karena kemampuannya dalam menangkap dependensi nonlinear dan temporal dalam data sekuensial (Mohanty,

2025). Pola yang sama terlihat dalam penelitian ini, di mana LSTM unggul pada seluruh metrik evaluasi regresi maupun klasifikasi. Hasan *et al.* (2024) secara eksplisit mengidentifikasi biaya komputasi yang tinggi sebagai kelemahan nyata LSTM, terutama ketika menangani dataset yang sangat besar. Dalam penelitian ini, LSTM membutuhkan waktu pelatihan 52 menit 33 detik — jauh lebih lama dibandingkan XGBoost (6 menit 08 detik) dan LightGBM (2 menit 51 detik). Majeed *et al.* (2025) juga mencatat bahwa jaringan saraf dalam seperti LSTM memerlukan penyyetelan hiperparameter yang cermat dan ekstensif, berbeda dengan model pohon yang lebih toleran terhadap nilai *default*. Hal ini menjelaskan mengapa penurunan RMSE LSTM setelah *tuning* (23,0%) adalah yang terbesar di antara semua algoritma dalam penelitian ini.

5.2 Posisi XGBoost dan LightGBM: Keseimbangan Antara Akurasi dan Efisiensi

XGBoost dan LightGBM menempati posisi yang menarik dalam hierarki performa penelitian ini. Selisih RMSE keduanya dengan LSTM hanya 0,27–0,41 mm, tetapi waktu pelatihannya 8–18 kali lebih cepat — bukan sekadar angka teknis, melainkan argumen praktis yang kuat untuk konteks operasional. Pringandana & Kusnawi (2025) melakukan perbandingan langsung antara XGBoost dan LightGBM untuk klasifikasi curah hujan multikelas di Jakarta menggunakan data meteorologi historis — konteks yang sangat relevan dengan penelitian ini. Hasilnya menunjukkan bahwa XGBoost secara konsisten mengungguli LightGBM, mencapai akurasi 94% dan F1-Score 94% dibandingkan akurasi 91% LightGBM. Pola ini terulang dalam penelitian ini: XGBoost mencatat F1-Score 0,869 sementara LightGBM 0,863 — selisih kecil namun konsisten. Studi tersebut juga mengonfirmasi bahwa metode GBDT (*Gradient Boosted Decision Trees*) sangat sesuai untuk data cuaca tropis yang kompleks, karena kemampuannya menangani hubungan nonlinear antar variabel meteorologi. Adnan *et al.* (2021) membandingkan metode *bagging* (Random Forest, Bagged CART) dengan *boosting* (Stochastic Gradient Boosting) pada data curah hujan harian di Stasiun Meteorologi Sultan Syarif Kasim II Pekanbaru — salah satu wilayah dengan iklim tropis basah yang representatif. Hasilnya menunjukkan bahwa Stochastic Gradient Boosting mencapai akurasi rata-rata tertinggi di antara semua algoritma *ensemble* yang diuji. Mekanisme *boosting* yang membangun pohon secara sekuensial — di mana setiap pohon baru memperbaiki kesalahan pohon sebelumnya — terbukti lebih efektif dalam menangkap pola curah hujan tropis yang tidak beraturan dibandingkan pendekatan *bagging* yang membangun pohon secara paralel dan independen. Toharudin *et al.* (2023) memberikan perspektif tambahan yang relevan: dalam studi klasifikasi konsentrasi PM2.5 menggunakan data meteorologi BMKG Kemayoran Jakarta, mereka menemukan bahwa LightGBM lebih cepat dan menggunakan memori lebih sedikit dibandingkan XGBoost untuk dataset besar — keunggulan yang secara langsung mendukung temuan waktu komputasi dalam penelitian ini. Studi ini juga mengonfirmasi bahwa kedua algoritma *boosting* ini mampu menangani ketidakseimbangan kelas dalam data meteorologi Indonesia, yang relevan dengan distribusi kelas tidak seimbang (67,4% tidak hujan vs 7,8% hujan ekstrem) yang ditemukan dalam dataset penelitian ini. Dari perspektif regularisasi, XGBoost memiliki keunggulan bawaan berupa regularisasi L1 dan L2 yang secara aktif mencegah *overfitting* — sebuah keunggulan yang tidak dimiliki RF secara eksplisit. Hal ini menjelaskan mengapa XGBoost lebih stabil pada dataset dengan banyak fitur yang saling berkorelasi, seperti variabel meteorologi dalam penelitian ini (Lyu & Yong, 2024).

5.3 Keterbatasan Random Forest dan SVM

Random Forest mencatat performa yang solid namun tidak kompetitif dibandingkan algoritma *boosting* dan LSTM. F1-Score 0,818 dan *recall* 79,6% untuk deteksi hujan ekstrem mengindikasikan bahwa RF melewatkan sekitar 2 dari 10 kejadian ekstrem yang sebenarnya terjadi. Amnuaylojaroen (2023) melaporkan bahwa RF memang menunjukkan korelasi Pearson yang tinggi (0,94) dengan data observasi resolusi tinggi untuk prediksi suhu di Asia Tenggara, tetapi MSE-nya (2,78) masih lebih tinggi dibandingkan *Decision Tree* (2,43) dalam konteks *downscaling* iklim — mengindikasikan bahwa agregasi banyak pohon dalam RF tidak selalu menghasilkan estimasi yang lebih presisi untuk setiap skenario. Kelemahan mendasar RF terletak pada mekanisme *bagging*-nya: pohon-pohon dibangun secara paralel dan independen sehingga tidak ada mekanisme koreksi kesalahan yang sistematis. Berbeda dengan *boosting* yang secara eksplisit menargetkan sampel yang sulit diprediksi, RF memperlakukan semua sampel dengan bobot yang sama — sebuah kelemahan yang sangat terasa ketika distribusi kelas tidak seimbang, seperti dalam prediksi hujan ekstrem (Adnan *et al.*, 2021).

SVM mencatat performa terendah di antara semua algoritma yang diuji, dengan RMSE 6,47 mm, R^2 0,763, dan F1-Score 0,746. Majeed *et al.* (2025) memberikan konfirmasi empiris yang relevan: dalam prediksi deret waktu di lingkungan tropis Malaysia, SVM (RBF SVM) mencatat RMSE 4,37–5,58 — jauh lebih tinggi dibandingkan LSTM yang mencapai RMSE minimum 1,94. Keterbatasan SVM pada dataset meteorologi berskala besar bersumber dari dua masalah struktural yang saling terkait. Kompleksitas komputasinya meningkat secara kuadratik terhadap jumlah sampel, yang menjelaskan waktu pelatihan 38 menit 47 detik dalam penelitian ini. Di sisi lain, pemilihan fungsi *kernel* yang tepat untuk data meteorologi multivariat berdimensi tinggi adalah tantangan tersendiri yang tidak memiliki solusi universal (Yu *et al.*, 2017; AbdulRaheem *et al.*, 2022). Kombinasi kedua kelemahan ini menjadikan SVM pilihan yang sulit direkomendasikan untuk aplikasi meteorologi berskala besar.

5.4 Pengaruh Kualitas Data dan Kepadatan Stasiun

Variasi performa model antar wilayah — RMSE LSTM berkisar dari 3,61 mm di Jawa Barat hingga 5,43 mm di NTT — tidak dapat dijelaskan semata-mata oleh karakteristik algoritma. Faktor yang lebih menentukan adalah kualitas dan kepadatan data observasi lokal. Inuwa *et al.* (2025) memberikan konteks yang sangat relevan: dalam studi dinamika spasiotemporal anomali presipitasi di Asia Tenggara, mereka mengembangkan model Random Forest menggunakan hubungan anomali-ENSO sebelumnya dan mencatat $R^2 = 0,82$ serta RMSE = 18,6 mm. Yang lebih penting, studi ini mengidentifikasi bahwa pengaruh ENSO yang kuat — tahun El Niño 2015 dan 2019 menyebabkan defisit curah hujan pan-regional di Indonesia, sementara La Niña 2010 dan 2022 menghasilkan anomali basah — menciptakan variabilitas ekstrem yang secara langsung mempengaruhi akurasi model ML. Wilayah NTT, yang sangat rentan terhadap siklus El Niño–La Niña, mengalami transisi musim yang tajam antara musim hujan dan kemarau — kondisi yang sulit ditangkap model manapun tanpa data observasi yang memadai dan representatif. Amnuaylojaroen (2023) juga menekankan pentingnya menangkap kompleksitas iklim regional yang beragam di Asia Tenggara dan memitigasi dampak fenomena cuaca ekstrem seperti El Niño yang secara signifikan mengubah pola cuaca di zona tropis. Temuan ini mengonfirmasi bahwa performa model ML untuk prediksi cuaca tropis sangat sensitif terhadap kualitas data lokal — bukan semata-mata pilihan algoritma (Putra *et al.*, 2024). Implikasinya jelas: investasi dalam perluasan dan peningkatan kualitas jaringan stasiun BMKG — terutama di wilayah timur Indonesia — berpotensi meningkatkan akurasi prediksi lebih signifikan dibandingkan sekadar beralih ke algoritma yang lebih canggih.

5.5 Implikasi Praktis untuk Sistem Operasional BMKG

Hasil penelitian ini memiliki implikasi langsung yang berbeda-beda tergantung pada konteks penerapannya. Untuk sistem peringatan dini bencana hidrometeorologi di mana akurasi deteksi kejadian ekstrem adalah prioritas mutlak, LSTM adalah pilihan yang paling dapat dipertahankan secara ilmiah. *Recall* 88,1% berarti lebih dari 8 dari 10 kejadian hujan ekstrem berhasil dideteksi — angka yang relevan langsung untuk koordinasi respons darurat. Namun, biaya komputasinya yang tinggi memerlukan infrastruktur yang memadai, yang tidak selalu tersedia di semua kantor BMKG daerah (Chitwatkulsiri & Miyamoto, 2023). Untuk prakiraan cuaca operasional harian yang perlu diperbarui secara berkala dengan sumber daya komputasi terbatas, XGBoost dan LightGBM menawarkan keseimbangan yang lebih realistis. Pringandana & Kusnawi (2025) mengonfirmasi bahwa XGBoost menyediakan waktu pelatihan dan prediksi yang cepat, menjadikannya sesuai untuk sistem pendukung keputusan *real-time*. Toharudin *et al.* (2023) menambahkan bahwa LightGBM lebih cepat dan menggunakan memori lebih sedikit untuk dataset besar — keunggulan yang sangat relevan untuk sistem operasional dengan pembaruan data harian. Satu pendekatan yang menjanjikan adalah arsitektur hibrida bertingkat: LightGBM atau XGBoost untuk prakiraan harian rutin yang membutuhkan kecepatan, dan LSTM untuk analisis kejadian ekstrem yang membutuhkan akurasi maksimal. Pendekatan *Double Machine Learning* (DML) yang dikembangkan Lyu & Yong (2024) — di mana XGBoost dan RF menentukan klasifikasi kejadian hujan, lalu LSTM mengambil alih untuk estimasi kuantitatif — adalah contoh konkret bagaimana kedua kelas algoritma ini dapat bekerja secara komplementer, bukan saling menggantikan.

5.6 Keterbatasan Penelitian dan Arah Penelitian Selanjutnya

Meskipun dataset mencakup 12 stasiun dari tiga pulau besar, representasi wilayah timur Indonesia yaitu; Papua, Maluku, dan Nusa Tenggara. Penelitian ini juga belum mengeksplorasi arsitektur LSTM

yang lebih canggih seperti *Bidirectional-LSTM* atau *attention-enhanced LSTM* yang dilaporkan Joy *et al.* (2025) mampu meningkatkan R menjadi 0,9834 untuk prediksi suhu. Selain itu, pengaruh ENSO dan monsun belum dimodelkan secara eksplisit sebagai fitur *input* — padahal Inuwa *et al.* (2025) menunjukkan bahwa indeks ENSO adalah prediktor yang kuat untuk anomali presipitasi di Asia Tenggara. Untuk penelitian selanjutnya, beberapa arah yang layak dipertimbangkan meliputi eksplorasi arsitektur *attention-enhanced LSTM* untuk meningkatkan akurasi lebih lanjut; integrasi indeks ENSO dan fase monsun sebagai fitur eksplisit dalam model; pengembangan model hibrida XGBoost-LSTM untuk sistem operasional yang menyeimbangkan akurasi dan efisiensi; serta perluasan cakupan spasial ke wilayah timur Indonesia yang selama ini kurang terwakili dalam literatur prediksi cuaca berbasis ML.

6. Kesimpulan

Penelitian ini membandingkan kinerja lima algoritma *machine learning* — Random Forest, SVM, LSTM, XGBoost, dan LightGBM — dalam memprediksi parameter cuaca di wilayah tropis Indonesia menggunakan data gabungan BMKG, ERA5, dan TRMM. Hasil penelitian menunjukkan bahwa LSTM secara konsisten mencatat performa terbaik pada tugas regresi (RMSE = 3,94 mm; $R^2 = 0,891$) maupun klasifikasi (F1-Score = 0,887; AUC-ROC = 0,941), didukung oleh kemampuannya menangkap dependensi temporal jangka panjang dalam data deret waktu meteorologi tropis. XGBoost dan LightGBM berada di posisi kedua dan ketiga dengan selisih akurasi yang kecil terhadap LSTM, namun dengan waktu pelatihan yang jauh lebih efisien — menjadikan keduanya pilihan yang lebih realistis untuk sistem prakiraan cuaca operasional dengan sumber daya komputasi terbatas. Random Forest menunjukkan performa yang memadai namun tertinggal dari algoritma *boosting*, sementara SVM mencatat performa terendah pada seluruh metrik evaluasi dengan biaya komputasi yang justru tinggi — kombinasi yang sulit direkomendasikan untuk dataset meteorologi berskala besar.

Penelitian ini juga menemukan bahwa variasi performa antar wilayah — RMSE LSTM berkisar dari 3,61 mm di Jawa Barat hingga 5,43 mm di NTT — lebih banyak dipengaruhi oleh kualitas dan kepadatan data observasi lokal dibandingkan pilihan algoritma itu sendiri, terutama akibat pengaruh ENSO dan monsun yang berbeda-beda di tiap wilayah. Optimasi hiperparameter terbukti memberikan dampak yang signifikan, dengan penurunan RMSE antara 16,5% hingga 23,0% di seluruh algoritma setelah *tuning* dilakukan secara sistematis. Berdasarkan keseluruhan temuan ini, penelitian merekomendasikan pendekatan hibrida bertingkat yang menggabungkan LightGBM atau XGBoost untuk prakiraan harian rutin dan LSTM untuk analisis kejadian ekstrem, disertai dengan perluasan jaringan stasiun observasi BMKG — terutama di wilayah timur Indonesia — sebagai investasi yang dampaknya akan melampaui sekadar pergantian algoritma.

Referensi

- AbdulRaheem, M., Awotunde, J. B., Abidemi, E. A., Idowu, D. O., & Adekola, S. O. (2022). Weather prediction performance evaluation on selected machine learning algorithms. *IAES International Journal of Artificial Intelligence*, 11(4), 1535. <https://doi.org/10.11591/ijai.v11.i4>
- Adnan, A., Yolanda, A. M., & Natasya, F. (2021, October). A comparison of bagging and boosting on classification data: Case study on rainfall data in Sultan Syarif Kasim II meteorological station in Pekanbaru. *Journal of Physics: Conference Series*, 2049(1), 012053. <https://doi.org/10.1088/1742-6596/2049/1/012053>
- Amini, A., Dolatshahi, M., & Kerachian, R. (2023). Effects of automatic hyperparameter tuning on the performance of multi-variate deep learning-based rainfall nowcasting. *Water Resources Research*, 59(1), e2022WR032789. <https://doi.org/10.1029/2022WR032789>
- Amnuaylojaroen, T. (2023). Advancements in downscaling global climate model temperature data in southeast asia: A machine learning approach. *Forecasting*, 6(1), 1–17. <https://doi.org/10.3390/forecast6010001>

- Auliya, M. N., Saputra, A. H., Kristianto, A., & Qomariyatuzzamzami, L. N. (2023, November). Predicting hailstorms through machine learning approach using multiple source data analysis. In *International Conference on Radioscience, Equatorial Atmospheric Science and Environment* (pp. 225–236). Springer Nature Singapore. https://doi.org/10.1007/978-981-97-0740-9_21
- Barrera-Animas, A. Y., Oyedele, L. O., Bilal, M., Akinosho, T. D., Delgado, J. M. D., & Akanbi, L. A. (2022). Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. *Machine Learning with Applications*, 7, 100204. <https://doi.org/10.1016/j.mlwa.2021.100204>
- Chen, L., Han, B., Wang, X., Zhao, J., Yang, W., & Yang, Z. (2023). Machine learning methods in weather and climate applications: A survey. *Applied Sciences*, 13(21), 12019. <https://doi.org/10.3390/app132112019>
- Chen, R., Zhang, W., & Wang, X. (2020). Machine learning in tropical cyclone forecast modeling: A review. *Atmosphere*, 11(7), 676. <https://doi.org/10.3390/atmos11070676>
- Chen, Y., Teo, F. Y., Wong, S. Y., Chan, A., Weng, C., & Falconer, R. A. (2024). Monsoonal extreme rainfall in Southeast Asia: A review. *Water*, 17(1), 5. <https://doi.org/10.3390/w17010005>
- Chitwatkulsiri, D., & Miyamoto, H. (2023). Real-time urban flood forecasting systems for southeast asia — A review of present modelling and its future prospects. *Water*, 15(1), 178. <https://doi.org/10.3390/w15010178>
- de Burgh-Day, C. O., & Leeuwenburg, T. (2023). Machine learning for numerical weather and climate modelling: A review. *Geoscientific Model Development*, 16(22), 6433–6477. <https://doi.org/10.5194/gmd-16-6433-2023>
- Dueben, P. D., Schultz, M. G., Chantry, M., Gagne, D. J., Hall, D. M., & McGovern, A. (2022). Challenges and benchmark datasets for machine learning in the atmospheric sciences: Definition, status, and outlook. *Artificial Intelligence for the Earth Systems*, 1(3), e210002. <https://doi.org/10.1175/AIES-D-21-0002.1>
- Gong, Y., Zhang, Y., Wang, F., & Lee, C. H. (2024). Deep learning for weather forecasting: A CNN-LSTM hybrid model for predicting historical temperature data. *arXiv preprint arXiv:2410.14963*. <https://doi.org/10.48550/arXiv.2410.14963>
- Hasan, M. M., Hasan, M. J., & Rahman, P. B. (2024). Comparison of RNN-LSTM, TDFD and stacking model approach for weather forecasting in Bangladesh using historical data from 1963 to 2022. *PLOS ONE*, 19(9), e0310446. <https://doi.org/10.1371/journal.pone.0310446>
- Hridoy, M. A. A. M., Shawkat, A. I., Bordin, C., Acharjee, M. R., Masood, A., Baki, A. O., & Al Mamun, M. A. (2025). Advanced machine learning models for accurate water quality classification and WQI prediction: Implications for aquatic disease risk management. *Science of the Total Environment*, 1008, 180965. <https://doi.org/10.1016/j.scitotenv.2025.180965>
- Inuwa, S. S., Dimiyati, M., Masita, D. M. M., & Hafid, S. (2025). Spatio-temporal dynamics of precipitation anomalies in Southeast Asia: ENSO influence and machine learning-based prediction. *The 46th Asian Conference on Remote Sensing*, 1–25.
- Irmada, H. N., Ermatita, E., bin Awang, M. K., & Adrezo, M. (2024). Enhancing weather prediction models through the application of random forest method and chi-square feature selection. *JOIV: International Journal on Informatics Visualization*, 8(3-2), 1506–1514. <https://doi.org/10.62527/joiv.8.3-2.2356>
- Jisha, G. (2024, August). Enhanced weather prediction with feature engineered, time series cross validated ridge regression model. In *2024 Control Instrumentation System Conference (CISCON)* (pp. 1–6). IEEE. <https://doi.org/10.1109/CISCON62171.2024.10696530>

- Joy, U. G., Kabir, S., & Niger, T. (2025). Attention-enhanced LSTM modeling for improved temperature and rainfall forecasting in Bangladesh. *Theoretical and Applied Climatology*, 156(11), 613. <https://doi.org/10.1007/s00704-025-05858-5>
- Kahfi, S., Wiharjo, S., & Rivai, A. K. (2025). Optimization of employee burnout prediction using explainable boosting machine, long short-term memory, and extreme gradient boosting methods in human resource management at PT. XYZ. *International Journal Software Engineering and Computer Science (IJSECS)*, 5(3), 1083–1094. <https://doi.org/10.35870/ijsecs.v5i3.5772>
- Kreuzer, D., Munz, M., & Schlüter, S. (2020). Short-term temperature forecasts using a convolutional neural network — An application to different weather stations in Germany. *Machine Learning with Applications*, 2, 100007. <https://doi.org/10.1016/j.mlwa.2020.100007>
- Latif, S. D., Hazrin, N. A. B., Koo, C. H., Ng, J. L., Chaplot, B., Huang, Y. F., & Ahmed, A. N. (2023). Assessing rainfall prediction models: Exploring the advantages of machine learning and remote sensing approaches. *Alexandria Engineering Journal*, 82, 16–25. <https://doi.org/10.1016/j.aej.2023.09.060>
- Lee, J. C. K., Zhang, H., Barker, D. M., Chen, S., Kumar, R., An, B. W., & Chandramouli, K. (2022). Weather prediction for Singapore — Progress, challenges, and opportunities. *Meteorology*, 1(4), 394–401. <https://doi.org/10.3390/meteorology1040025>
- Li, C., Ren, X., & Zhao, G. (2023). Machine-learning-based imputation method for filling missing values in ground meteorological observation data. *Algorithms*, 16(9), 422. <https://doi.org/10.3390/a16090422>
- Liu, Z., Yang, Q., Shao, J., Wang, G., Liu, H., Tang, X., & Bai, L. (2022). Improving daily precipitation estimation in the data scarce area by merging rain gauge and TRMM data with a transfer learning framework. *Journal of Hydrology*, 613, 128455. <https://doi.org/10.1016/j.jhydrol.2022.128455>
- Lyu, Y., & Yong, B. (2024). A novel double machine learning strategy for producing high-precision multi-source merging precipitation estimates over the Tibetan Plateau. *Water Resources Research*, 60(4), e2023WR035643. <https://doi.org/10.1029/2023WR035643>
- Majeed, M. A., Shafri, H. Z., Zulkafli, Z., & Wayayok, A. (2025). Dengue fever prediction using LSTM and integrated temporal-spatial attention: A case study of Malaysia. *Spatial Information Research*, 33(1), 5. <https://doi.org/10.1007/s41324-025-00603-6>
- Mardyansyah, R. Y., Kurniawan, B., Soekirno, S., Nuryanto, D. E., & Satria, H. (2022, December). Artificial intelligent for rainfall estimation in tropical region: A survey. *IOP Conference Series: Earth and Environmental Science*, 1105(1), 012024. <https://doi.org/10.1088/1755-1315/1105/1/012024>
- Markovics, D., & Mayer, M. J. (2022). Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction. *Renewable and Sustainable Energy Reviews*, 161, 112364. <https://doi.org/10.1016/j.rser.2022.112364>
- Mishra, P., Ray, S., Lal, P., Nair, S. B., Matuka, A., Tashkandy, Y., & Emam, W. (2025). Climate modeling for South Asia: Statistical and deep learning for rainfall and temperature prediction. *Scientific Reports*, 15(1), 38235. <https://doi.org/10.1038/s41598-025-22149-1>
- Mohanty, Y. (2025, March). Predicting droughts: A comparative study of ARIMAX, LSTM, XGBoost, and random forest models. In *2025 11th International Conference on Computing and Artificial Intelligence (ICCAI)* (pp. 772–787). IEEE. <https://doi.org/10.1109/ICCAI66501.2025.00121>

- Nguyen, T., Jewik, J., Bansal, H., Sharma, P., & Grover, A. (2023). Climatelearn: Benchmarking machine learning for weather and climate modeling. *Advances in Neural Information Processing Systems*, 36, 75009–75025.
- Parra-Plazas, J., Gaona-Garcia, P., & Plazas-Nossa, L. (2023). Time series outlier removal and imputing methods based on Colombian weather stations data. *Environmental Science and Pollution Research*, 30(28), 72319–72335. <https://doi.org/10.1007/s11356-023-27176-x>
- Pringandana, C. G. L., & Kusnawi, K. (2025). A comparative analysis of hyperparameter-tuned XGBoost and LightGBM for multiclass rainfall classification in Jakarta. *Jurnal Teknik Informatika (JUTIF)*, 6(4), 2467–2483. <https://doi.org/10.52436/1.jutif.2025.6.4.4965>
- Putra, A. F. D., Azmi, M. N., Wijayanto, H., Utama, S., & Wirawan, I. G. P. W. W. (2024). Optimizing rain prediction model using random forest and grid search cross-validation for agriculture sector. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 23(3), 519–530. <https://doi.org/10.30812/matrik.v23i3.3891>
- Putra, M., Rosid, M. S., & Handoko, D. (2024). A review of rainfall estimation in Indonesia: Data sources, techniques, and methods. *Signals*, 5(3), 542–561. <https://doi.org/10.3390/signals5030030>
- Rasp, S., Dueben, P. D., Scher, S., Weyn, J. A., Mouatadid, S., & Thuerey, N. (2020). WeatherBench: A benchmark data set for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems*, 12(11), e2020MS002203. <https://doi.org/10.1029/2020MS002203>
- Suacana, I. W. G., Suhariyanto, D., & Nuru, F. (2024). Optimizing the 2024 governor election quick count with extreme gradient boosting (XGBoost) to increase voting prediction accuracy. *International Journal Software Engineering and Computer Science (IJSECS)*, 4(1), 91–106. <https://doi.org/10.35870/ijsecs.v4i1.2286>
- Sun, L., & Fu, Y. (2021). A new merged dataset for analyzing clouds, precipitation and atmospheric parameters based on ERA5 reanalysis data and the measurements of the Tropical Rainfall Measuring Mission (TRMM) precipitation radar and visible and infrared scanner. *Earth System Science Data*, 13(5), 2293–2306. <https://doi.org/10.5194/essd-13-2293-2021>
- Trink, S. (2025). Machine learning-based forecasting of air quality index under long-term environmental patterns: A comparative approach with XGBoost, LightGBM, and SVM. *PLOS ONE*, 20(10), e0334252. <https://doi.org/10.1371/journal.pone.0334252>
- Toharudin, T., Caraka, R. E., Pratiwi, I. R., Kim, Y., Gio, P. U., Sakti, A. D., & Pardamean, B. (2023). Boosting algorithm to handle unbalanced classification of PM2.5 concentration levels by observing meteorological parameters in Jakarta-Indonesia using AdaBoost, XGBoost, CatBoost, and LightGBM. *IEEE Access*, 11, 35680–35696. <https://doi.org/10.1109/ACCESS.2023.3265019>
- V, R. C., Johnvictor, A. C., & N, P. S. (2025). Comparative analysis of machine learning approaches for heatwave event prediction in India. *Scientific Reports*, 15, 22431. <https://doi.org/10.1038/s41598-025-04634-9>
- Waqas, M., Humphries, U. W., Wangwongchai, A., Dechpichai, P., & Ahmad, S. (2023). Potential of artificial intelligence-based techniques for rainfall forecasting in Thailand: A comprehensive review. *Water*, 15(16), 2979. <https://doi.org/10.3390/w15162979>
- Yu, P. S., Yang, T. C., Chen, S. Y., Kuo, C. M., & Tseng, H. W. (2017). Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting. *Journal of Hydrology*, 552, 92–104. <https://doi.org/10.1016/j.jhydrol.2017.06.020>

Zhuang, H., Lehner, F., & DeGaetano, A. T. (2024). Improved diagnosis of precipitation type with LightGBM machine learning. *Journal of Applied Meteorology and Climatology*, 63(3), 437–453. <https://doi.org/10.1175/JAMC-D-23-0117.1>.