

**ANALISIS KOMPARASI KINERJA ALGORITMA RANDOM FOREST, XGBOOST,
DAN LSTM DALAM MEMPREDIKSI CURAH HUJAN HARIAN
DI STASIUN METEOROLOGI SOEKARNO-HATTA**

**COMPARATIVE ANALYSIS OF RANDOM FOREST, XGBOOST, AND LSTM
PERFORMANCE IN PREDICTING DAILY RAINFALL
AT SOEKARNO-HATTA METEOROLOGICAL STATION**

**Muhammad Aulia Zikri¹⁾, Gracia Nadya Angelina Pangalo²⁾, Muhammad
Fathurrahman Alghazali³⁾, Angelina Serena Gracella Nesty Youwe⁴⁾, Ferdy Indra
Kurniawan⁵⁾, Yosik Norman⁶⁾**

^{1,2,3,4,5)} Program Studi Sarjana Terapan Meteorologi, Sekolah Tinggi Meteorologi Klimatologi dan
Geofisika

⁶⁾ Stasiun Klimatologi Banten, Badan Meteorologi Klimatologi dan Geofisika

Email: ¹⁾auliazikri@stmkg.ac.id, ²⁾graciapangalo28@gmail.com, ³⁾angelyouwe21@gmail.com,
⁴⁾ferdyindra38@gmail.com, ⁵⁾yosik.normal@bmkkg.go.id

ABSTRAK

Curah hujan merupakan salah satu faktor penting bagi keselamatan penerbangan. Hal ini mampu memengaruhi jarak pandang pilot dan gesekan pada runway (*runway friction*). Penelitian ini bertujuan membandingkan kinerja algoritma Random Forest, XGBoost, dan LSTM dalam memprediksi curah hujan harian menggunakan data sinoptik periode 2021 hingga 2025. Analisis dilakukan melalui dua metode, yaitu klasifikasi biner kejadian hujan dan regresi intensitas curah hujan. Hasil penelitian menunjukkan bahwa model LSTM merupakan model paling unggul dengan akurasi klasifikasi tertinggi sebesar 70,11% (AUC: 0,75) serta galat regresi terendah (RMSE: 18,72 mm; MAE: 8,54 mm). Keunggulan LSTM terletak pada kemampuannya dalam menangkap keterikatan temporal dan hubungan nonlinier pada variabel sinoptik. Meskipun demikian, seluruh model masih perlu dievaluasi dalam mendeteksi kejadian hujan lebat akibat adanya ketidakseimbangan data yang menyebabkan bias ke arah kategori hujan ringan. Hal ini menegaskan potensi implementasi *ML* dalam mendukung layanan cuaca penerbangan sekaligus menekankan pentingnya peningkatan teknik pengolahan data untuk memitigasi risiko akibat cuaca ekstrem, khususnya dalam dunia penerbangan.

Kata Kunci: Curah Hujan, LSTM, XGBoost, Random Forest

ABSTRACT

Rainfall is a critical factor for aviation safety, as it significantly impacts pilot visibility and runway friction. This study aims to evaluate and compare the performance of RF (Random Forest), XGBoost, and LSTM (Long Short-Term Memory) algorithms in predicting daily rainfall using synoptic data from 2021 to 2025. The analysis was conducted through two distinct methods: binary classification of rainfall events and regression of rainfall intensity. The results demonstrate that the LSTM model is the most superior, achieving the highest classification accuracy of 70.11% (AUC: 0.75) and the lowest regression errors (RMSE: 18.72 mm; MAE: 8.54 mm). The superiority of LSTM is attributed to its specialized capability to capture temporal dependencies and nonlinear relationships within synoptic variables. Nevertheless, all models require further evaluation for detecting heavy rainfall events due to data imbalance, leading to a noticeable bias toward the light rain category. These findings underscore the potential of implementing ML to support aviation weather services while emphasizing the necessity of advancing data processing techniques to mitigate risks from extreme weather, particularly within the aviation sector

Keywords: rainfall, LSTM, XGBoost, RF

PENDAHULUAN

Faktor cuaca merupakan faktor penting dalam keselamatan dan kinerja operasional penerbangan, khususnya di bandara dengan tingkat lalu lintas udara tinggi seperti Bandara Internasional Soekarno-Hatta. Curah hujan dapat menurunkan jarak pandang pilot serta tingkat gesekan antara ban pesawat dan permukaan landasan pacu, sehingga berpotensi meningkatkan risiko gangguan operasional, termasuk kejadian tergelincir atau *runway excursion* (Supardam et al., 2020). Sebagai salah satu bandara di Indonesia, gangguan operasional akibat cuaca ekstrem, meskipun berskala kecil, dapat memicu keterlambatan penerbangan yang berdampak pada kerugian operasional dan ekonomi. Oleh karena itu, ketersediaan informasi meteorologi yang akurat serta sistem peringatan dini yang akurat menjadi faktor penting dalam mendukung operasional bandara (Putri et al., 2025).

Prediksi curah hujan merupakan permasalahan komputasi yang kompleks karena dinamika atmosfer yang bersifat nonlinier. Metode konvensional seperti *Numerical Weather Prediction* (NWP) memiliki keterbatasan dalam merepresentasikan variabilitas spasial skala mikro dan menghasilkan prediksi yang akurat pada lokasi pengamatan spesifik, seperti stasiun meteorologi (Nitha et al., 2025). Selain itu, model statistik klasik seperti *Autoregressive Integrated Moving Average* (ARIMA) memiliki keterbatasan dalam merepresentasikan hubungan nonlinier yang kompleks antarvariabel meteorologis, sehingga kurang efektif dalam menangkap dinamika intrinsik data atmosfer yang bersifat non-stasioner dan dipengaruhi oleh interaksi variabel yang tidak linier (Kaur et al., 2023). Keterbatasan tersebut menyebabkan penurunan akurasi prediksi, khususnya pada kondisi cuaca ekstrem dan variabilitas musiman yang tinggi. Keterbatasan tersebut berdampak pada

penurunan akurasi prediksi, terutama pada kondisi cuaca ekstrem dan variabilitas musiman yang tinggi. Kompleksitas data atmosfer mendorong pergeseran dari pemodelan fisika murni menuju pendekatan berbasis data historis yang lebih adaptif, seiring meningkatnya ketersediaan data pengamatan meteorologis.

Perkembangan kecerdasan buatan, khususnya *Machine Learning* (ML), menawarkan alternatif untuk mengatasi keterbatasan metode konvensional dalam prediksi cuaca. Algoritma ML mampu memodelkan hubungan nonlinier yang kompleks antarvariabel meteorologis secara langsung dari data tanpa memerlukan formulasi fisika atmosfer secara eksplisit, sehingga lebih efisien secara komputasional (Schultz et al., 2021). Pendekatan ini memanfaatkan data historis jangka panjang untuk melatih model dalam mengidentifikasi pola cuaca berulang serta mendeteksi anomali atmosfer (Zhang et al., 2025). Berbagai studi menunjukkan bahwa pendekatan berbasis data-driven mampu menghasilkan kinerja prediksi yang unggul, bahkan dalam beberapa kasus melampaui model NWP, khususnya untuk *nowcasting* (Das et al., 2024).

Dalam pendekatan *ensemble learning*, *Random Forest* (RF) dan *XGBoost* dikenal luas karena kinerjanya yang baik pada data tabular. Penelitian oleh Iheanacho et al. (2025), menunjukkan bahwa RF efektif dalam menekan risiko *overfitting* melalui mekanisme *bagging*, sehingga menghasilkan prediksi yang stabil meskipun data cuaca mengandung *noise*. Di sisi lain, *Extreme Gradient Boosting* (XGBoost) menggunakan pendekatan *boosting* untuk memperbaiki kesalahan prediksi secara iteratif, sehingga menghasilkan akurasi tinggi dengan efisiensi komputasi yang baik (Chen & Guestrin, 2016). Hasil studi komparasi Sangaji & Sutabri (2025) menunjukkan

bahwa XGBoost cenderung mengungguli model *tree-based* lainnya dalam prediksi curah hujan berdasarkan evaluasi galat. Kedua algoritma, yaitu RF dan XGBoost, merepresentasikan hasil yang unggul, namun efektivitasnya masih perlu diuji lebih lanjut dalam menangani karakteristik cuaca tropis di Indonesia.

Di sisi lain, *Deep Learning* menawarkan pendekatan berbasis jaringan saraf tiruan untuk pemodelan data deret waktu. Algoritma *Long Short-Term Memory* (LSTM) memiliki mekanisme memori internal yang memungkinkan penangkapan ketergantungan temporal jangka panjang yang sering tidak terakomodasi oleh model ML konvensional (Hochreiter & Schmidhuber, 1997). Kemampuan ini penting dalam meteorologi karena kejadian hujan dipengaruhi oleh kondisi atmosfer pada periode sebelumnya. Penelitian oleh Wani et al. (2024) menunjukkan bahwa LSTM memberikan kinerja prediksi yang lebih stabil dan unggul dibandingkan dengan metode statistik tradisional pada data curah hujan dengan variabilitas iklim tinggi. Oleh karena itu, perbandingan LSTM dengan metode ensemble tree seperti RF dan XGBoost penting untuk menilai keseimbangan antara kompleksitas deep learning dan peningkatan akurasi prediksi.

Meskipun telah banyak diterapkan secara terpisah, studi komparatif yang mengevaluasi RF, XGBoost, dan LSTM menggunakan data Stasiun Meteorologi Soekarno-Hatta masih terbatas. Kinerja algoritma berpotensi bervariasi bergantung pada pola distribusi cuaca lokal di wilayah Cengkareng yang dipengaruhi oleh dinamika angin muson. Hingga saat ini, belum terdapat metode yang paling unggul dalam merepresentasikan karakteristik iklim tropis basah di lingkungan bandara, sehingga aspek ini masih menjadi celah penelitian yang perlu dikaji (Ejike et al., 2025). Oleh karena itu, penelitian ini bertujuan mengevaluasi dan membandingkan kinerja ketiga algoritma untuk menentukan model terbaik dalam

prediksi hujan harian, serta memberikan rekomendasi teknis bagi pengembangan sistem peringatan dini otomatis yang mendukung keselamatan penerbangan.

Untuk menempatkan hasil penelitian ini dalam konteks yang lebih luas, diperlukan perbandingan langsung dengan studi-studi yang relevan. Iheanacho et al. (2025) melaporkan bahwa model RF pada data cuaca Nigeria mencapai akurasi klasifikasi sebesar 73,4%, sedikit lebih tinggi dari RF pada penelitian ini (67,93%), yang dapat dikaitkan dengan karakteristik pola hujan konvektif di Nigeria yang lebih teratur dibandingkan dengan dinamika monsun di wilayah Cengkareng. Studi komparasi Sangaji & Sutabri (2025) pada data curah hujan Indonesia menunjukkan XGBoost mengungguli RF pada metrik RMSE, sejalan dengan temuan penelitian ini; namun pada skenario klasifikasi, hasil penelitian ini justru menunjukkan pola sebaliknya di mana RF lebih unggul dari XGBoost.

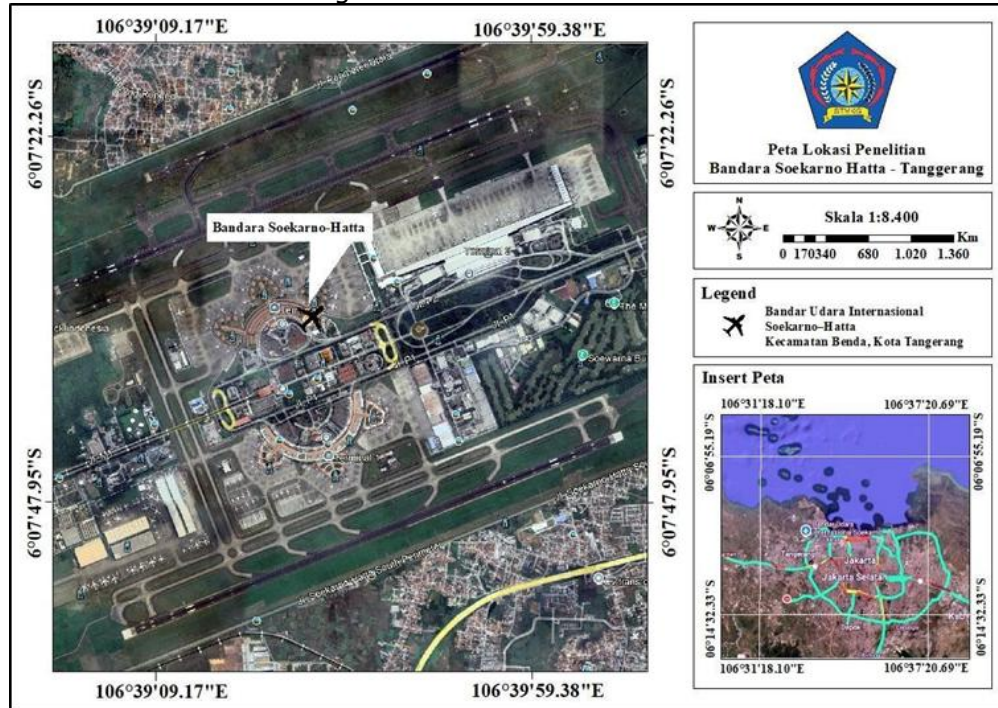
Perbedaan ini kemungkinan besar dipengaruhi oleh karakteristik distribusi data yang berbeda antara lokasi penelitian. Wani et al. (2024) yang menerapkan LSTM pada data Himalaya melaporkan MAE pada kisaran 7–12 mm untuk prediksi curah hujan harian, rentang yang sebanding dengan hasil penelitian ini (MAE LSTM: 8,54 mm), memperkuat validitas arsitektur LSTM yang digunakan. Namun, nilai R^2 dalam penelitian ini jauh lebih rendah dibandingkan dengan studi tersebut, yang kemungkinan mencerminkan perbedaan karakteristik iklim: iklim tropis basah di Soekarno-Hatta memiliki variabilitas konvektif yang lebih tinggi dan tidak terstruktur dibandingkan dengan iklim pegunungan Himalaya. Secara keseluruhan, kontribusi spesifik penelitian ini terletak pada evaluasi komparatif ketiga algoritma dalam konteks iklim tropis bandara di Indonesia, yang mengisi celah penelitian yang belum banyak dikaji dalam literatur sebelumnya (Ejike et al., 2025).

METODE PENELITIAN

Lokasi penelitian

Penelitian ini dilakukan di Bandara Internasional Soekarno-Hatta yang secara administratif berada di Kecamatan Benda, Kota Tangerang, Provinsi Banten. Data meteorologi yang digunakan bersumber dari hasil pengamatan permukaan yang dilakukan oleh Stasiun Meteorologi Kelas I

Soekarno-Hatta. Secara geografis, titik pengamatan stasiun ini terdaftar pada *World Meteorological Organization* (WMO) dengan kode stasiun 96749 serta kode ICAO WIII. Lokasi pengamatan berada pada koordinat $06^{\circ}07'$ Lintang Selatan dan $106^{\circ}39'$ Bujur Timur, dengan ketinggian sekitar 10 meter di atas permukaan laut.



Gambar 1. Lokasi penelitian

Data Penelitian

Data yang digunakan dalam penelitian ini berupa data sekunder hasil pengamatan sinoptik harian yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Stasiun Meteorologi Kelas I Soekarno-Hatta. Data sinoptik permukaan dipilih karena merepresentasikan kondisi cuaca aktual yang diukur secara langsung menggunakan instrumen standar WMO di lingkungan bandara. Penelitian ini menggunakan data historis dengan rentang waktu lima tahun, yaitu dari 1 Januari 2021 hingga 31 Desember 2025. Pemilihan periode tersebut didasarkan pada kebutuhan algoritma ML dan *Deep Learning* yang memerlukan volume data yang memadai untuk mempelajari pola musiman (*seasonality*) dan tren jangka panjang (*trend*) secara optimal (Li & Law, 2024). Data memiliki resolusi temporal harian. Setiap entri merepresentasikan ringkasan parameter cuaca pada tanggal pengamatan. Total data yang digunakan berjumlah 1.836 entri sebelum dilakukan proses pembersihan data (data cleaning). Penelitian ini melibatkan sembilan parameter meteorologi yang terdiri atas delapan variabel prediktor sebagai masukan model dan satu variabel target sebagai keluaran prediksi, dengan rincian variabel, satuan, peran dalam pemodelan, serta notasi matematis yang disajikan pada Tabel 1.

Tabel 1. Variabel data penelitian

No.	Label	Variabel	Satuan	Peran dalam Model	Notasi
1.	TN	Suhu Udara Minimum	°C	Prediktor	X1
2.	TX	Suhu Udara Maksimum	°C	Prediktor	X2
3.	TAVG	Suhu Udara Rata-Rata	°C	Prediktor	X3
4.	RH_AVG	Kelembapan Udara Rata-rata	%	Prediktor	X4
5.	SS	Lamanya Penyinaran Matahari	Jam	Prediktor	X5
6.	FF_X	Kecepatan Angin Maksimum	m/s	Prediktor	X6
7.	DDD_X	Arah Angin Terbanyak	°	Prediktor	X7
8.	FF_AVG	Kecepatan Angin Rata-rata	m/s	Prediktor	X8
9.	RR	Curah Hujan Harian	mm	Target	y

Teknik Pengolahan Data

Data *raw* meteorologi umumnya mengandung noise, *missing values*, serta ketidakkonsistenan format yang berpotensi menurunkan kinerja model prediksi. Oleh karena itu, dilakukan serangkaian tahapan pengolahan data (*data preprocessing*) untuk memperbaiki kualitas dan konsistensi data masukan.

Alur pengolahan data dalam penelitian ini terdiri atas empat tahapan utama, yaitu:

1. Pembersihan Data dan Imputasi

Data sinoptik harian sering menggunakan kode numerik khusus untuk merepresentasikan kondisi tertentu, seperti nilai 8888 yang menunjukkan data tidak terukur dan 9999 yang menandakan ketiadaan data. Pada tahap awal, nilai-nilai tersebut diidentifikasi dan dikonversi ke dalam format *Not a Number* (NaN) agar dapat diproses secara statistik pada tahap analisis selanjutnya. Setelah tahap konversi, penanganan *missing values* dilakukan dengan menerapkan metode interpolasi linier. Pemilihan metode ini didasarkan pada karakteristik data meteorologi sebagai data *time series* yang memiliki ketergantungan temporal, di mana kondisi cuaca pada suatu waktu dipengaruhi oleh kondisi pada periode sebelum dan sesudahnya. Secara matematis, metode interpolasi linear dinyatakan dengan persamaan berikut:

$$y = y_0 + (x - x_0) \frac{y_1 - y_0}{x_1 - x_0}$$

y merepresentasikan nilai *missing value* yang akan diestimasi, x_0, y_0 merupakan titik data valid pada waktu sebelumnya, dan x_1, y_1 merupakan titik data valid pada waktu setelahnya.

2. Transformasi Variabel Target

Untuk mengakomodasi dua skenario pengujian, yaitu klasifikasi dan regresi, variabel target curah hujan (RR) ditransformasikan ke dalam dua format label yang berbeda, sebagai berikut:

- 1) Label Klasifikasi (Biner): Format ini digunakan pada model klasifikasi *RF*, *XGBoost*, dan *LSTM Classifier* untuk membedakan kejadian hujan dan tidak hujan dengan ambang batas 0 mm.

$$y_{class} = \begin{cases} 0, & \text{jika } RR = 0 \text{ mm (tidak hujan)} \\ 1, & \text{jika } RR > 0 \text{ (hujan)} \end{cases}$$

- 2) Label Regresi (Kontinyu): Format ini digunakan pada model regresi untuk memprediksi intensitas curah hujan secara kuantitatif, dengan nilai RR dipertahankan dalam satuan aslinya (milimeter).

3. Normalisasi Data

Variabel prediktor memiliki rentang nilai dan satuan yang berbeda, sehingga perbedaan skala tersebut berpotensi menimbulkan bias dalam proses pembelajaran model, khususnya pada algoritma berbasis *gradient descent* seperti *LSTM*. Oleh karena itu, dilakukan proses normalisasi menggunakan metode *Min-Max Scaling* untuk mentransformasikan seluruh fitur masukan ke dalam rentang $[0, 1]$. Secara matematis, proses normalisasi dinyatakan dengan persamaan sebagai berikut:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Nilai X merupakan nilai aktual, dan X_{min} adalah nilai minimum fitur dan X_{max} adalah nilai maksimum fitur dalam dataset. X merepresentasikan nilai aktual, X_{min} merupakan nilai minimum fitur, dan X_{max} merupakan nilai maksimum fitur dalam dataset.

4. Pembagian Data Latih dan Uji

Pembagian data tidak dilakukan secara acak untuk mencegah kebocoran data masa depan (*look-ahead bias*). Pembagian data dilakukan menggunakan metode *hold-out* sekuensial dengan proporsi sebagai berikut:

- Data Latih (*Training Set*): 80% data awal, digunakan untuk melatih model mengenali pola cuaca.
- Data Uji (*Testing Set*): 20% data akhir digunakan untuk mengevaluasi kinerja model terhadap data yang belum pernah digunakan pada tahap pelatihan maupun validasi sebelumnya.

Namun, metode *hold-out* sekuensial memiliki keterbatasan dibandingkan dengan teknik validasi berbasis *time series* yang lebih canggih, seperti *walk-forward validation* atau *time series cross-validation*. Pendekatan *walk-forward validation*, melatih model secara bertahap dengan jendela waktu yang bergeser

sehingga menghasilkan estimasi kinerja yang lebih representatif terhadap kondisi operasional nyata. Namun, penerapan teknik tersebut pada penelitian ini dibatasi oleh ukuran dataset yang relatif terbatas (1.836 entri setelah *preprocessing*), di mana pemecahan lebih lanjut ke dalam banyak *fold* berisiko menghasilkan set pelatihan yang terlalu kecil untuk melatih model *deep learning* seperti *LSTM* secara optimal. Oleh karena itu, penggunaan *hold-out* sekuensial dengan porsi 80:20 merupakan pendekatan yang paling praktis dan dapat dipertanggungjawabkan secara metodologis dalam penelitian ini (Bergmeir & Benítez, 2012; Li & Law, 2024).

5. Tuning Model

Penentuan *hyperparameter* dilakukan melalui pendekatan empiris berbasis eksperimen yang sistematis. Untuk setiap algoritma, sejumlah kandidat nilai diuji secara bertahap dan dievaluasi berdasarkan kinerja pada data latih. Sebagai contoh, nilai *max_depth* pada *RF* dievaluasi pada rentang $\{5, 10, 15, 20\}$ dan dipilih nilai 10 karena menghasilkan keseimbangan terbaik antara akurasi dan risiko *overfitting*. Nilai *learning_rate* pada *XGBoost* diuji pada rentang $\{0,01; 0,05; 0,1; 0,2\}$, dan nilai 0,1 dipilih karena menghasilkan konvergensi yang stabil. Jumlah neuron *LSTM* diuji pada $\{32, 50, 64, 128\}$ dengan arsitektur satu *hidden layer*, dan 50 neuron dipilih sebagai konfigurasi paling efisien untuk ukuran dataset yang digunakan. Proses ini mengacu pada panduan praktis dari literatur terkait pemodelan data meteorologi berskala menengah (Li & Law, 2024; Wani et al., 2024). Konfigurasi parameter dikunci dengan menetapkan *random_state* = 42 untuk menjaga konsistensi hasil pada setiap percobaan. Pada model *RF*, batas tingkat kompleksitas model ditetapkan (*max_depth* = 10) untuk mengurangi risiko *overfitting* pada data meteorologi yang mengandung *noise*. Sementara itu, pada model *XGBoost*, nilai *learning_rate* ditetapkan sebesar 0,1 guna menyeimbangkan kecepatan konvergensi dan akurasi prediksi.

Sementara itu, arsitektur LSTM dirancang dengan satu *hidden layer* yang terdiri atas 50 neuron dan menggunakan fungsi aktivasi *Rectified Linear Unit* (ReLU) untuk meningkatkan stabilitas proses

pembelajaran pada data deret waktu. Proses pelatihan dilakukan selama 50 *epoch* dengan *batch size* 32 serta menggunakan *optimizer*. Rincian konfigurasi *hyperparameter* disajikan pada Tabel 2.

Tabel 2. Hyperparameter

Algoritma	Parameter	Nilai (Setting)	Fungsi/Deskripsi
Random Forest	<i>n_estimators</i>	100	Jumlah pohon keputusan yang dibangun.
	<i>max_depth</i>	10	Kedalaman maksimum pohon untuk mencegah overfitting.
	<i>random_state</i>	42	Kunci pengacak agar hasil konsisten (reproducible).
	<i>criterion</i>	MSE / Gini	Metrik evaluasi pemisahan node.
XGBoost	<i>n_estimators</i>	100	Jumlah iterasi boosting.
	<i>learning_rate</i>	0.1	Tingkat kecepatan adaptasi model terhadap error (bobot).
	<i>objective</i>	reg:squarederror	Fungsi objektif untuk skenario regresi.
	<i>subsample</i>	0.8	Proporsi data yang digunakan tiap iterasi (mencegah <i>overfitting</i>).
LSTM	<i>Units (Neurons)</i>	50	Jumlah neuron pada Hidden Layer.
	<i>Activation</i>	ReLU	Fungsi aktivasi non-linear (Rectified Linear Unit).
	<i>Optimizer</i>	Adam	Algoritma optimasi pembaruan bobot.
	<i>Epochs</i>	50	Jumlah iterasi pelatihan pada seluruh dataset.
	<i>Batch Size</i>	32	Jumlah sampel data yang diproses sebelum update bobot.
	<i>Loss Function</i>	MSE	Fungsi kerugian yang diminimalkan (untuk regresi).

Teknik Analisis Data

Untuk menilai efektivitas model *ML* dan *Deep Learning* yang dikembangkan, evaluasi dilakukan secara komprehensif melalui pendekatan kuantitatif dan kualitatif. Proses evaluasi kinerja model dibagi ke dalam dua skenario pengujian utama yang disesuaikan dengan tujuan penelitian, yaitu:

1. Evaluasi Model Klasifikasi (Deteksi Kejadian Hujan)
Skenario ini bertujuan untuk mengevaluasi kemampuan model dalam mengklasifikasikan kondisi

cuaca ke dalam dua kelas, yaitu hujan (1) dan tidak hujan (0). Metrik evaluasi yang digunakan meliputi:

- 1) Akurasi (*Accuracy*): mengukur proporsi prediksi yang benar, baik positif maupun negatif, terhadap keseluruhan data. Dalam persamaan yang digunakan, TP merepresentasikan *true positive*, TN *true negative*, FP *false positive*, dan FN *false negative*.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- 2) F1-Score: merupakan *harmonic mean* dari *precision* dan *recall*, serta

sangat relevan untuk digunakan pada data cuaca yang tidak seimbang (*imbalanced dataset*), di mana jumlah kejadian tidak hujan umumnya lebih dominan dibandingkan dengan kejadian hujan.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

- 3) ROC-AUC (*Receiver Operating Characteristic - Area Under Curve*): Kurva *Receiver Operating Characteristic* (ROC) menunjukkan hubungan antara *True Positive Rate* dan *False Positive Rate* pada berbagai nilai ambang (*threshold*). Nilai *Area Under the Curve* (AUC) berada pada rentang 0 hingga 1, di mana nilai yang lebih besar menunjukkan kemampuan model yang lebih baik dalam membedakan kejadian hujan dan tidak hujan.
- 4) *Reliability Diagram* (Kurva Kalibrasi): Digunakan untuk menilai kesesuaian antara probabilitas prediksi dan frekuensi kejadian aktual, serta mengidentifikasi kecenderungan model yang bersifat *over-confident* atau *under-confident* dalam memprediksi peluang hujan.
2. Evaluasi Model Regresi (Estimasi Curah Hujan)
Skenario ini bertujuan untuk mengevaluasi kedekatan antara nilai prediksi intensitas curah hujan (dalam milimeter) dan data observasi aktual. Metrik error yang digunakan adalah:
 - 1) *Root Mean Squared Error* (RMSE): Metrik ini mengukur akar dari rata-rata kuadrat kesalahan prediksi, di mana kesalahan dengan nilai besar diberikan bobot yang lebih tinggi.

Oleh karena itu, metrik ini sensitif dalam mengidentifikasi ketidakmampuan model dalam memprediksi kejadian hujan dengan intensitas tinggi.

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

- 2) *Mean Absolute Error* (MAE): Mengukur rata-rata selisih absolut antara nilai prediksi dan observasi aktual, sehingga memberikan gambaran besaran kesalahan rata-rata yang lebih mudah diinterpretasikan secara fisik.

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

- 3) *Coefficient of Determination* (R^2): Mengukur proporsi variansi data curah hujan yang dapat dijelaskan oleh model. Nilai R^2 yang mendekati 1 menunjukkan bahwa model mampu merepresentasikan variasi data dengan tingkat presisi yang tinggi.
3. Analisis Kategorikal (Tabel Kontingensi): Hasil prediksi regresi selanjutnya dievaluasi secara kategorikal menggunakan tabel kontingensi. Analisis ini memanfaatkan *confusion matrix* multikelas untuk menilai kinerja model dalam mengklasifikasikan kejadian hujan ekstrem yang berpotensi memengaruhi keselamatan penerbangan. Nilai prediksi kemudian dikelompokkan berdasarkan standar intensitas curah hujan BMKG sebagai berikut:

Tabel 3. Klasifikasi curah hujan menurut BMKG

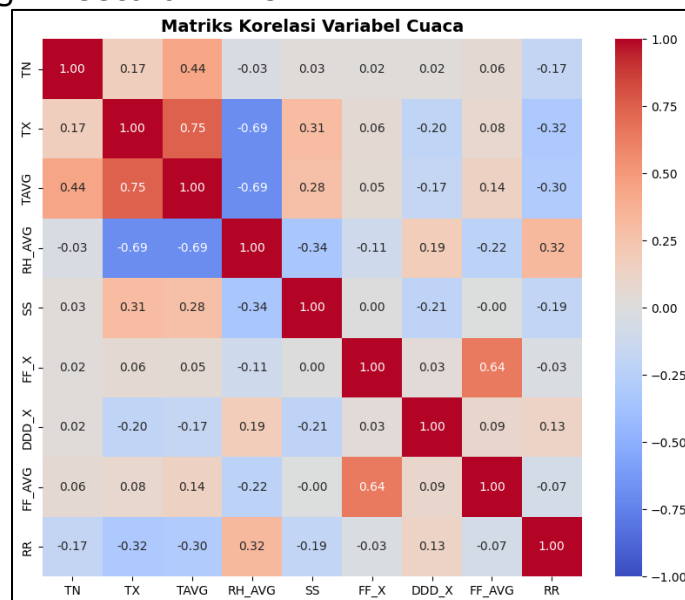
No	Klasifikasi	Rentang (mm /hari)
1.	Hujan ringan	0.5 s.d. < 20
2.	Hujan sedang	20 s.d. < 50
3.	Hujan lebat	50 s.d. < 100
4.	Hujan sangat lebat	100 s.d. < 150
5.	Hujan ekstrem	> 150

HASIL DAN PEMBAHASAN

Matriks Korelasi

Berdasarkan hasil analisis korelasi, variabel kelembapan udara rata-rata (RH_AVG) menjadi faktor dengan pengaruh positif terkuat terhadap curah hujan sebesar 0,32, yang mengindikasikan bahwa peningkatan kadar uap air di atmosfer merupakan pendorong utama terjadinya presipitasi di lokasi penelitian. Sebaliknya, suhu udara maksimum (TX) dan suhu rata-rata (TAVG) menunjukkan korelasi negatif yang signifikan, masing-masing sebesar -0,32 dan -0,30, yang secara fisik

mencerminkan penurunan suhu permukaan saat kondisi udara jenuh dan terjadi hujan. Selain itu, arah angin terbanyak (DDD_X) memberikan kontribusi positif kecil (0,13), sementara lamanya penyinaran matahari (SS) dan suhu minimum (TN) menunjukkan hubungan berbanding terbalik terhadap intensitas curah hujan harian. Secara keseluruhan, rendahnya nilai korelasi linier pada seluruh variabel prediktor (di bawah 0,40) menunjukkan sifat atmosfer yang sangat dinamis dan nonlinier.



Gambar 2. Matriks korelasi

Kinerja Klasifikasi Kejadian Hujan (Hujan vs Tidak Hujan)

Evaluasi model pada skenario pertama berfokus pada kemampuan algoritma dalam mendeteksi kejadian hujan (biner). Berdasarkan hasil rekapitulasi metrik, ditemukan variasi performa yang signifikan antara metode Shallow Learning (RF dan XGBoost) dan Deep Learning (LSTM).

Perbandingan Akurasi dan F1-Score

Tabel rekapitulasi menunjukkan bahwa LSTM memiliki performa terbaik dengan

tingkat akurasi mencapai 0,701 (70,11%) dan F1-Score sebesar 0,730. F1-Score yang tinggi pada LSTM mengindikasikan bahwa model ini lebih seimbang dalam menangani data yang tidak seimbang (imbalanced dataset), di mana kejadian hujan sering kali lebih sedikit dibandingkan dengan hari tidak hujan. RANDOM FOREST menempati posisi kedua dengan akurasi 0,679, sementara XGBoost memiliki akurasi terendah sebesar 0,638.

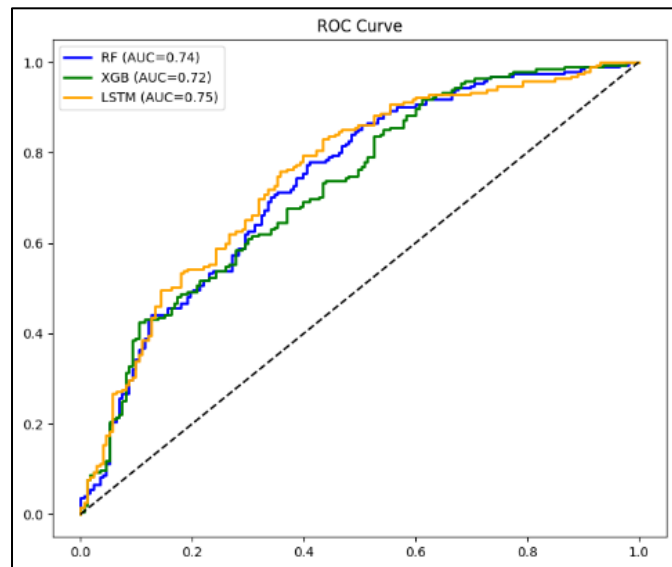
Tabel 4. Perbandingan akurasi dan F-1 score

No	Model	Akurasi	F1 - Score
1	Random Forest	0.679348	0.700508
2	XGBoost	0.638587	0.673219
3	LSTM	0.701087	0.730392

Analisis Kemampuan Separasi (ROC-AUC)

Berdasarkan kurva *Receiver Operating Characteristic* (ROC) pada Gambar 3, model LSTM kembali menunjukkan keunggulan dengan nilai AUC 0,75, disusul oleh RF (0,74) dan XGBoost (0,72). Nilai AUC yang mendekati

1 menunjukkan kemampuan model yang lebih baik dalam memisahkan kelas hujan dan tidak hujan pada berbagai ambang batas probabilitas. Hal ini membuktikan bahwa struktur memori jangka panjang pada LSTM efektif dalam menangkap ketergantungan temporal data cuaca di Stasiun Meteorologi Soekarno-Hatta.

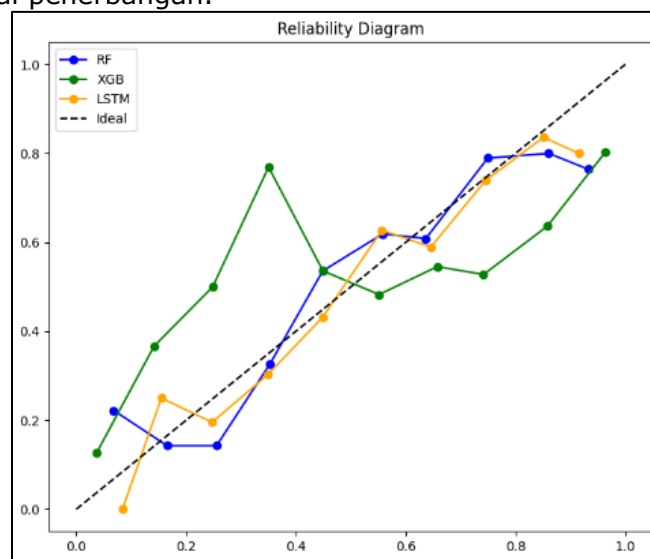


Gambar 3. Grafik ROC-AUC

Diagram Probabilitas

Pada Gambar 4, terlihat bahwa kurva LSTM dan RF berada lebih dekat dengan garis diagonal (ideal) dibandingkan dengan XGBoost. Kurva XGBoost cenderung berada di atas garis ideal pada rentang probabilitas rendah, yang menunjukkan adanya gejala *over-confident* dalam memprediksi peluang hujan. Sebaliknya, LSTM menunjukkan kalibrasi yang lebih stabil, yang sangat krusial untuk sistem peringatan dini otomatis di bandara guna menghindari kesalahan deteksi yang dapat merugikan operasional penerbangan.

Keberhasilan LSTM dalam mencapai akurasi tertinggi berkaitan erat dengan kemampuannya memproses variabel dengan korelasi nonlinier. Meskipun variabel seperti kelembapan (RH_AVG) hanya memiliki korelasi linear 0,32 dan suhu (TX) sebesar -0,32 terhadap curah hujan, LSTM mampu mengekstraksi fitur tersembunyi dari interaksi stokastik antarvariabel tersebut yang gagal ditangkap secara optimal oleh model XGBoost.



Gambar 4. Diagram Reabilitas

Evaluasi Error Numerik

Berdasarkan tabel error numerik pada Tabel 5, dapat dilihat bahwa estimasi intensitas curah hujan secara kuantitatif masih merupakan tantangan besar. LSTM menghasilkan nilai terendah dengan RMSE sebesar 18,72 mm dan MAE sebesar 8,54 mm, mengungguli RF (RMSE=18,92 mm) dan XGBoost (RMSE=19,32 mm). Dalam konteks operasional penerbangan, nilai RMSE sebesar 18,72 mm pada LSTM bermakna bahwa rata-rata estimasi intensitas curah hujan menyimpang hampir 19 mm dari nilai aktual. Pada kondisi hujan sedang dengan intensitas 20–50 mm, simpangan sebesar ini dapat menyebabkan kesalahan klasifikasi kategori curah hujan yang berdampak pada keputusan operasional terkait visibilitas pilot dan kondisi runway.

Nilai RMSE yang jauh lebih besar daripada MAE pada seluruh model menunjukkan adanya kesalahan yang signifikan akibat kegagalan model dalam memprediksi kejadian hujan dengan intensitas ekstrem yang memiliki frekuensi kemunculan lebih jarang dalam dataset. Hal ini perlu dikaji kembali dan ditegaskan bahwa nilai koefisien determinasi (R^2) yang sangat rendah pada kisaran 0,03 hingga 0,09

merupakan temuan yang tidak dapat diabaikan.

Nilai R^2 ini mengindikasikan bahwa model hanya mampu menjelaskan 3–9% variansi total curah hujan harian dari data observasi, yang berarti sekitar 91–97% variabilitas data tidak berhasil ditangkap oleh model. Rendahnya R^2 ini bukan semata-mata mencerminkan kelemahan model, melainkan juga mencerminkan kompleksitas data atmosfer tropis yang bersifat sangat nonlinier, stokastik, dan dipengaruhi oleh proses fisika multiskala yang tidak seluruhnya terekam dalam delapan variabel prediktor yang digunakan.

Penelitian sebelumnya oleh Ejike et al. (2025) yang menemukan bahwa model ML berbasis variabel permukaan pada iklim tropis basah secara konsisten menghasilkan R^2 yang jauh lebih rendah dibandingkan dengan iklim subtropis atau benua, karena variabilitas konvektif skala meso yang dominan di kawasan tropis sulit ditangkap oleh fitur sinoptik harian. Untuk meningkatkan nilai R^2 , penelitian lanjutan perlu mempertimbangkan penambahan variabel prediktor yang lebih representatif seperti indeks stabilitas atmosfer (CAPE, CIN), data kelembapan vertikal dari radiosonde, atau indeks iklim seperti IOD dan ENSO.

Tabel 5. Evaluasi metrik

No	Model	RMSE	MAE	R^2
1	RF	18,9197	9,1073	0,0748
2	XG-Bost	19,3193	8,7936	0,0353
3	LSTM	18,7207	8,5438	0,0942

Analisis Matriks Kontingensi

Pada matriks kontingensi Gambar 5, terlihat adanya bias model yang sangat kuat terhadap kategori hujan ringan. Ketiga algoritma cenderung melakukan prediksi yang berpusat pada kelas "Berawan" dan "Ringan", namun mengalami kesulitan drastis dalam mendeteksi kategori "Sedang" dan "Lebat".

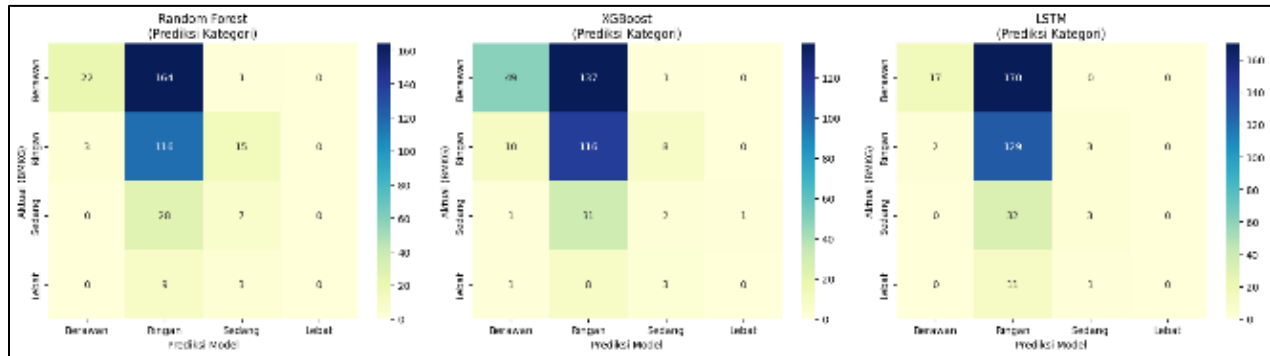
Pada model LSTM, sebagian besar data aktual hujan sedang dan lebat justru diprediksi sebagai hujan ringan. Kondisi ini secara langsung disebabkan oleh ketidakseimbangan data (*imbalanced dataset*), di mana distribusi kelas dalam dataset sangat miring: berdasarkan analisis distribusi data pada periode 2021–2025, hari dengan curah hujan ringan (0,5–20 mm) mendominasi dengan

proporsi lebih dari 60% dari total hari hujan, sementara hari dengan intensitas sedang hingga lebat (≥ 20 mm) hanya mencakup sebagian kecil observasi.

Ketimpangan distribusi ini menyebabkan fungsi loss selama pelatihan didominasi oleh sampel kelas mayoritas, sehingga model mengembangkan kecenderungan sistematis untuk memprediksi intensitas rendah. Untuk mengatasi keterbatasan ini, penelitian lanjutan disarankan untuk mengeksplorasi teknik penanganan imbalanced dataset seperti Synthetic Minority Over-sampling Technique (SMOTE) guna mensintesis sampel sintetis

pada kelas minoritas (He & Garcia, 2009), penggunaan class weighting untuk memberikan bobot lebih besar pada sampel kelas hujan lebat dalam fungsi loss, atau penerapan threshold moving pada output probabilistik model klasifikasi (Chawla et al., 2002).

Ketidakmampuan model dalam memprediksi hujan lebat secara akurat merupakan catatan kritis bagi operasional penerbangan, mengingat dampaknya terhadap visibilitas pilot dan kondisi landasan pacu justru paling signifikan pada kategori hujan yang paling sulit diprediksi ini.

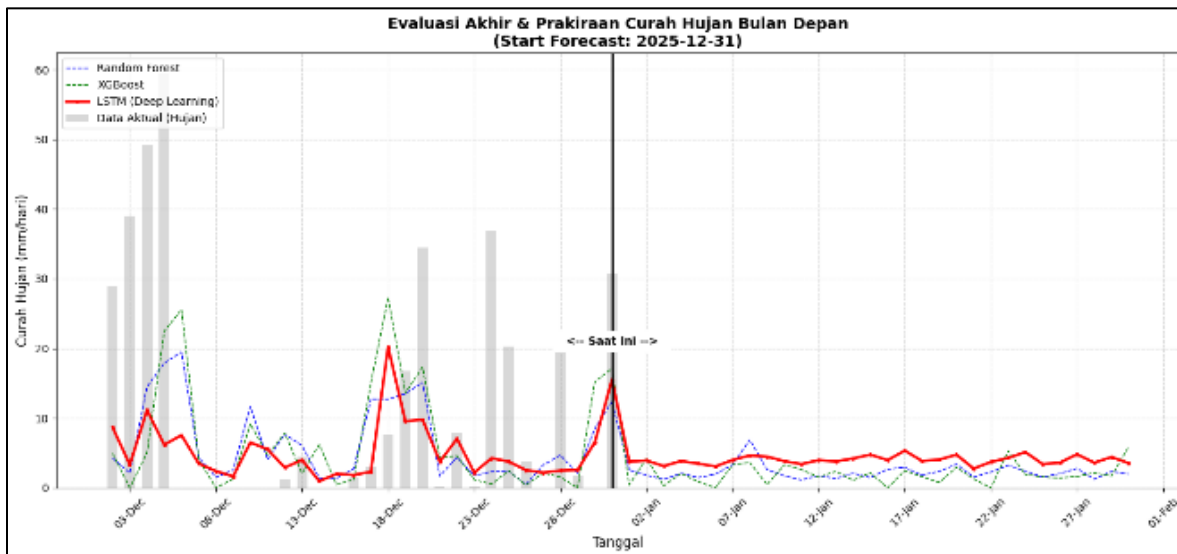


Gambar 5. Matriks kontingensi

Time series prediksi

Berdasarkan grafik time series pada Gambar 6, ketiga model menunjukkan kemampuan yang baik dalam mengikuti tren fluktuasi curah hujan harian, terutama dalam mengenali periode basah dan kering di Stasiun Meteorologi Soekarno-Hatta. Model LSTM (garis merah) memberikan hasil prediksi yang paling stabil dan halus dibandingkan dengan model berbasis pohon keputusan (RF dan XGBoost), khususnya pada fase prakiraan. Hal ini mengindikasikan bahwa struktur memori

pada LSTM efektif dalam menjaga konsistensi prediksi jangka panjang. Meskipun mampu menangkap tren secara umum, ketiga model cenderung melakukan *underestimate* atau gagal mencapai nilai puncak pada kejadian hujan ekstrem. Grafik menunjukkan bahwa saat terjadi lonjakan curah hujan aktual yang tinggi, proyeksi model sering kali berada jauh di bawah nilai observasi. Kendala ini tercermin pada nilai RMSE yang cukup besar, yang dipicu oleh besarnya selisih antara nilai prediksi dan aktual pada hari-hari dengan intensitas hujan lebat.



Gambar 6. Time series prediksi

KESIMPULAN

Penelitian ini berhasil menunjukkan bahwa algoritma Deep Learning LSTM merupakan model yang paling optimal baik untuk skenario klasifikasi kejadian hujan maupun regresi intensitas curah hujan. Pada pengujian klasifikasi, LSTM mencapai tingkat akurasi tertinggi sebesar 70,11% dengan nilai AUC 0,75, mengungguli RF (67,93%) dan XGBoost (63,85%). Keunggulan LSTM didorong oleh kemampuannya dalam menangkap ketergantungan temporal data deret waktu dan mengekstraksi fitur tersembunyi dari variabel atmosfer yang memiliki hubungan non-linear kompleks.

Dalam skenario regresi, LSTM juga mencatatkan nilai kesalahan terendah dengan RMSE 18,72 mm dan MAE 8,54 mm. Meskipun unggul secara komparatif, ketiga model masih menghadapi tantangan besar dalam memprediksi kejadian hujan ekstrem atau hujan lebat. Analisis melalui matriks kontingensi dan verifikasi visual menunjukkan adanya bias yang kuat terhadap kategori hujan ringan, sementara kategori hujan lebat hampir gagal dideteksi secara akurat oleh seluruh model. Hal ini disebabkan oleh ketidakseimbangan data (imbalanced dataset), di mana sampel hujan lebat sangat jarang terjadi dalam periode pengamatan.

Secara keseluruhan, penggunaan algoritma ML terbukti lebih responsif dalam menangani sifat dinamis atmosfer tropis dibandingkan model statistik konvensional. Namun, untuk mendukung sistem peringatan dini keselamatan penerbangan yang lebih presisi, diperlukan pengembangan lebih lanjut pada teknik penanganan ketidakseimbangan data agar model dapat mengenali potensi hujan lebat yang signifikan memengaruhi visibilitas dan kondisi landasan pacu.

DAFTAR PUSTAKA

- Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191, 192–213. <https://doi.org/10.1016/j.ins.2011.12.028>
- BMKG. (2023). *Penyediaan Dan Penyebarluasan Peringatan Dini Cuaca Ekstrem Sesaat (Nowcasting) 1 (Satu) Sampai Dengan 3 (Tiga) Jam Ke Depan Di Lingkungan Badan Meteorologi, Klimatologi, Dan Geofisika* (pp. 1–11). <https://jdih.bmkg.go.id/common/dokumen/2023sopbmkg025.pdf>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling

- technique. *Journal of Artificial Intelligence Research*, 16, 321–357.
<https://doi.org/10.1613/jair.953>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016*(August 2016), 785–794.
<https://doi.org/10.1145/2939672.2939785>
- Das, P., Posch, A., Barber, N., Hicks, M., Duffy, K., Vandal, T., Singh, D., Werkhoven, K. van, & Ganguly, A. R. (2024). Hybrid physics-AI outperforms numerical weather prediction for extreme precipitation nowcasting. *Npj Climate and Atmospheric Science*, 7(1), 1–15.
<https://doi.org/10.1038/s41612-024-00834-8>
- Ejike, O., Ndzi, D., & Shakir, M. Z. (2025). Comparative Study of ML-Based Rainfall Prediction in Tropical and Temperate Climates. *Climate*, 13(8), 1–27.
<https://doi.org/10.3390/cli13080167>
- He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284.
<https://doi.org/10.1109/TKDE.2008.239>
- Hochreiter, S., & Schmidhuber, J. (1997). *Long Short-Term Memory*. 9(8), 1–32.
- Iheanacho, F., Chosen, B., Araoye, T., & Cosmas, O. (2025). *Comparative analysis of RF and LSTM for weather forecasting*. 458–463.
- Kaur, J., Parmar, K. S., & Singh, S. (2023). Autoregressive models in environmental forecasting time series: a theoretical and application review. *Environmental Science and Pollution Research*, 30(8), 19617–19641.
<https://doi.org/10.1007/s11356-023-25148-9>
- Li, W., & Law, K. L. E. (2024). Deep Learning Models for Time Series Forecasting: A Review. *IEEE Access*, 12(Cv), 92306–92327.
<https://doi.org/10.1109/ACCESS.2024.3422528>
- Nitha, V., Pramada, S. K., Praseed, N. S., & Sridhar, V. (2025). Performance Evaluation of Numerical Weather Prediction Models in Forecasting Rainfall Events in Kerala, India. *Atmosphere*, 16(4).
<https://doi.org/10.3390/atmos16040372>
- Putri, N. A., Ramadha, D. B., Kusuma, T. D. P. A., & Prayitno, H. (2025). Pengaruh Faktor Cuaca Terhadap Perencanaan Penerbangan Dan Dampaknya Pada Keterlambatan Penerbangan. *Jurnal Penelitian*, 10(1), 19–27.
<https://doi.org/10.46491/jp.v10i1.1873>
- Sangaji, D., & Sutabri, T. (2025). Analisis XGBoost dan RF untuk Prediksi Curah Hujan dalam Mendukung Mitigasi Karhutla. *Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence)*, 5(1), 13–18.
<https://doi.org/10.55382/jurnalpustakaai.v5i1.905>
- Schultz, M. G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L. H., Mozaffari, A., & Stadtler, S. (2021). Can deep learning beat numerical weather prediction? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194).
<https://doi.org/10.1098/rsta.2020.0097>
- Supardam, D., Fitrianti, R., Yulia, S. N., Penerbangan, P., & Abstrak, M. (2020). The Impact of Lack of Information regarding Runway Conditions after The Rain to Flight Safety. *Jurnal Teknik Dan Keselamatan Transportasi*, 3, 30–34.
- Wani, O. A., Mahdi, S. S., Yeasin, M., Kumar, S. S., Gagnon, A. S., Danish, F., Al-Ansari, N., El-Hendawy, S., & Mattar, M. A. (2024). Predicting rainfall using ML, deep learning, and time series models across an

altitudinal gradient in the North-
Western Himalayas. *Scientific
Reports*, 14(1).
[https://doi.org/10.1038/s41598-024-
77687-x](https://doi.org/10.1038/s41598-024-77687-x)

Zhang, H., Liu, Y., Zhang, C., & Li, N.
(2025). ML Methods for Weather
Forecasting: A Survey. *Atmosphere*,
16(1), 1–34.
<https://doi.org/10.3390/atmos16010>