

Analisis Komparatif Model SARIMAX, XGBoost, dan LSTM untuk Peramalan Curah Hujan Bulanan di Kota Makassar

Mohammad Zahid, Rahmawati, Andi Seppewali, dan Bintang Guntur



Volume 14, Issue 2, Pages 434–456, August 2026

Diterima 21 Mei 2026, Direvisi 17 Juli 2026, Disetujui 23 Juli 2026, Diterbitkan 1 Agustus 2026

To Cite this Article : M. Zahid, R. Rahmawati, A. Seppewali, dan B. Guntur, "Analisis Komparatif Model SARIMAX, XGBoost, dan LSTM untuk Peramalan Curah Hujan Bulanan di Kota Makassar", *Euler J. Ilm. Mat. Sains dan Teknol.*, vol. 14, no. 2, pp. 434–456, 2026, <https://doi.org/10.37905/euler.v14i2.39072>

© 2026 by author(s)

JOURNAL INFO • EULER : JURNAL ILMIAH MATEMATIKA, SAINS DAN TEKNOLOGI

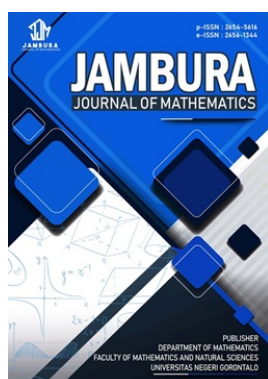


	Homepage	: http://ejurnal.ung.ac.id/index.php/euler/index
	Journal Abbreviation	: Euler J. Ilm. Mat. Sains dan Teknol.
	Frequency	: Three times a year
	Publication Language	: English (preferable), Indonesia
	DOI	: https://doi.org/10.37905/euler
	Online ISSN	: 2776-3706
	License	: Creative Commons Attribution-NonCommercial 4.0 International License
	Publisher	: Department of Mathematics, Universitas Negeri Gorontalo
	Country	: Indonesia
	OAI Address	: http://ejurnal.ung.ac.id/index.php/euler/oai
	Google Scholar ID	: QF_r-gAAAAJ
	Email	: euler@ung.ac.id

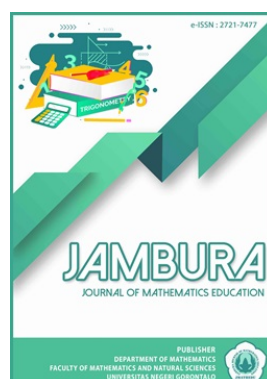
JAMBURA JOURNAL • FIND OUR OTHER JOURNALS



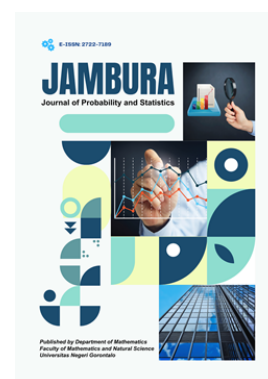
Jambura Journal of Biomathematics



Jambura Journal of Mathematics



Jambura Journal of Mathematics Education



Jambura Journal of Probability and Statistics

Analisis Komparatif Model SARIMAX, XGBoost, dan LSTM untuk Peramalan Curah Hujan Bulanan di Kota Makassar

Mohammad Zahid^{1,*}, Rahmawati¹, Andi Seppewali¹, Bintang Guntur¹

¹Jurusan Matematika, Universitas Sulawesi Barat, Majene 91414, Indonesia

ARTICLE HISTORY

Diterima 21 Mei 2026
Direvisi 17 Juli 2026
Disetujui 23 Juli 2026
Diterbitkan 1 Agustus 2026

KATA KUNCI

Curah Hujan
SARIMAX
XGBoost
LSTM
Peramalan
Time Series

KEYWORDS

Rainfall
SARIMAX
XGBoost
LSTM
Forecasting
Time Series

ABSTRAK. Curah hujan merupakan salah satu unsur meteorologi yang penting dalam berbagai sektor, seperti pertanian, pengelolaan sumber daya air, dan mitigasi bencana hidrometeorologi. Penelitian ini bertujuan membandingkan performa model Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX), Extreme Gradient Boosting (XGBoost), dan Long Short-Term Memory (LSTM) dalam peramalan curah hujan bulanan di Kota Makassar. Data yang digunakan merupakan data meteorologi bulanan periode Januari 2000 hingga Desember 2024 yang diperoleh dari NASA POWER dengan total 300 observasi. Variabel yang digunakan meliputi curah hujan, suhu udara, kelembapan, kecepatan angin, tekanan udara, dan radiasi matahari. Tahapan penelitian meliputi data preprocessing, analisis eksplorasi data, uji stasioneritas menggunakan Augmented Dickey–Fuller (ADF), pemodelan SARIMAX, XGBoost, dan LSTM, serta evaluasi model menggunakan Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), dan koefisien determinasi (R^2). Hasil penelitian menunjukkan bahwa model XGBoost menghasilkan performa terbaik dengan nilai RMSE sebesar 2.5441 dan R^2 sebesar 0.8063, sedangkan model SARIMAX menghasilkan nilai MAPE terkecil sebesar 39.2614%. Sementara itu, model LSTM menghasilkan performa yang belum optimal dengan nilai RMSE sebesar 5.2125 dan R^2 sebesar 0.1646. Hasil penelitian menunjukkan bahwa pendekatan machine learning berbasis boosting lebih efektif dalam menangani hubungan nonlinear pada data curah hujan bulanan dibandingkan model statistik klasik maupun deep learning pada jumlah data terbatas.

ABSTRACT. Rainfall is one of the important meteorological elements in various sectors, such as agriculture, water resource management, and hydrometeorological disaster mitigation. This study aims to compare the performance of Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX), Extreme Gradient Boosting (XGBoost), and Long Short-Term Memory (LSTM) models in monthly rainfall forecasting in Makassar City. The data used were monthly meteorological data from January 2000 to December 2024 obtained from NASA POWER with a total of 300 observations. The variables used include rainfall, temperature, humidity, wind speed, pressure, and solar radiation. The research stages consisted of data preprocessing, exploratory data analysis, stationarity testing using the Augmented Dickey–Fuller (ADF) test, SARIMAX, XGBoost, and LSTM modeling, and model evaluation using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). The results showed that the XGBoost model achieved the best performance with an RMSE value of 2.5441 and an R^2 value of 0.8063, while the SARIMAX model produced the lowest MAPE value of 39.2614%. Meanwhile, the LSTM model showed less optimal performance with an RMSE value of 5.2125 and an R^2 value of 0.1646. The results indicate that the boosting-based machine learning approach is more effective in handling nonlinear relationships in monthly rainfall data compared to classical statistical and deep learning models on limited datasets.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License. **Editorial of EULER:** Department of Mathematics, Universitas Negeri Gorontalo, Jln. Prof. Dr. Ing. B. J. Habibie, Bone Bolango 96554, Indonesia.

1. Pendahuluan

Curah hujan merupakan salah satu unsur meteorologi yang memiliki peranan penting dalam berbagai sektor kehidupan, seperti pertanian, pengelolaan sumber daya air, mitigasi banjir, hingga perencanaan infrastruktur [1]. Intensitas curah hujan yang tinggi dan tidak me-

*Penulis Korespondensi.

entu dapat menyebabkan berbagai bencana hidrometeorologi, sedangkan curah hujan yang rendah dalam jangka waktu tertentu dapat memicu kekeringan. Oleh karena itu, peramalan curah hujan yang akurat menjadi sangat penting untuk mendukung pengambilan keputusan yang tepat dalam berbagai bidang.

Kota Makassar sebagai salah satu kota metropolitan di Indonesia memiliki karakteristik iklim tropis dengan pola curah hujan musiman yang dipengaruhi oleh berbagai faktor atmosfer dan oseanografi. Variabilitas curah hujan di wilayah ini dipengaruhi oleh perubahan suhu udara, kelembapan, tekanan udara, kecepatan angin, serta radiasi matahari. Kondisi tersebut menyebabkan pola curah hujan bersifat kompleks, dinamis, dan cenderung *nonlinear*, sehingga membutuhkan metode peramalan yang mampu menangkap pola musiman maupun hubungan antarvariabel meteorologi [1].

Metode statistik deret waktu masih banyak digunakan dalam peramalan curah hujan karena memiliki kemampuan yang baik dalam menangkap pola tren dan musiman [2]. Berbagai metode statistik telah dikembangkan untuk meningkatkan akurasi peramalan, mulai dari metode pemulusan (*smoothing*) seperti *Double Moving Average* dan *Exponential Moving Average* [3, 4] hingga model autoregresif seperti *Autoregressive Integrated Moving Average* (ARIMA) dan *Seasonal Autoregressive Integrated Moving Average* (SARIMA). Salah satu pengembangan dari model SARIMA adalah *Seasonal Autoregressive Integrated Moving Average with Exogenous Variables* (SARIMAX), yang mengakomodasi pengaruh variabel eksogen sehingga faktor-faktor eksternal dapat diintegrasikan ke dalam proses pemodelan [5, 6]. Berbagai penelitian menunjukkan bahwa model deret waktu dengan variabel eksogen mampu menghasilkan prediksi yang lebih akurat karena dapat menangkap pola musiman sekaligus pengaruh variabel meteorologi secara simultan [7]. Selain itu, metode deret waktu telah banyak diterapkan dalam berbagai permasalahan peramalan di Indonesia, baik pada bidang meteorologi maupun sektor lainnya, dan menunjukkan kinerja yang baik dalam memodelkan fenomena temporal [1, 8].

Seiring berkembangnya teknologi komputasi, pendekatan *machine learning* dan *deep learning* mulai banyak diterapkan dalam bidang klimatologi dan meteorologi. Salah satu metode *machine learning* yang memiliki performa tinggi pada data *nonlinear* adalah *Extreme Gradient Boosting* (XGBoost). Metode ini mampu menangani hubungan kompleks antarvariabel dan memiliki kemampuan generalisasi yang baik pada berbagai jenis data [9, 10]. Penelitian-penelitian menggunakan *ensemble learning* berbasis *gradient boosting* menunjukkan bahwa pendekatan ini efektif dalam menangani data berdimensi tinggi dengan hubungan *nonlinear* [10, 11]. Selain itu, beberapa penelitian juga menunjukkan bahwa metode *hybrid machine learning* mampu memberikan akurasi prediksi yang lebih tinggi dibandingkan metode tunggal [12].

Di sisi lain, metode *deep learning* seperti *Long Short-Term Memory* (LSTM) juga banyak digunakan pada data deret waktu karena mampu mempelajari ketergantungan jangka panjang (*long-term dependency*) pada data sekuensial [13, 14]. Model LSTM dirancang untuk mengatasi kelemahan *Recurrent Neural Network* (RNN) tradisional dalam mempertahankan informasi historis pada urutan data yang panjang. Penelitian-penelitian sebelumnya menunjukkan bahwa model berbasis *deep learning* seperti LSTM mampu meningkatkan akurasi prediksi pada data dengan pola temporal yang dinamis [13]. Selain itu, berbagai implementasi jaringan saraf tiruan juga terbukti efektif dalam menangani data *nonlinear* dan fluktuatif pada berbagai domain [15].

Selain itu, pendekatan *ensemble learning* berbasis pohon keputusan juga diterapkan pada permasalahan prediksi lain, seperti *quantile regression forest* [16]. Penelitian Khaulasari dan Maulana Akbar pada *Euler: Jurnal Ilmiah Matematika, Sains dan Teknologi* menunjukkan bahwa model GSTARIMAX mampu digunakan untuk memodelkan data deret waktu yang dipengaruhi

variasi kalender dan faktor eksternal [6]. Hal ini menunjukkan bahwa pengembangan metode prediksi deret waktu terus mengalami kemajuan, baik melalui pendekatan statistik maupun *machine learning* [12, 16].

Meskipun berbagai penelitian mengenai peramalan curah hujan telah banyak dilakukan, sebagian besar penelitian masih berfokus pada penerapan satu metode tertentu tanpa melakukan perbandingan secara komprehensif antara pendekatan statistik, *machine learning*, dan *deep learning* menggunakan himpunan data yang sama. Selain itu, penelitian yang membandingkan performa model SARIMAX, XGBoost, dan LSTM pada data curah hujan bulanan di Kota Makassar masih sangat terbatas. Padahal, karakteristik curah hujan di wilayah tropis Indonesia yang dipengaruhi oleh pola musiman serta hubungan antarvariabel meteorologi yang bersifat kompleks dan *nonlinear* memerlukan evaluasi berbagai pendekatan pemodelan untuk memperoleh metode peramalan yang paling sesuai.

Berdasarkan kesenjangan penelitian tersebut, penelitian ini melakukan analisis komparatif terhadap model SARIMAX, XGBoost, dan LSTM menggunakan data meteorologi bulanan Kota Makassar dengan memanfaatkan variabel eksogen berupa suhu udara, kelembapan, kecepatan angin, tekanan udara, dan radiasi matahari. Kinerja ketiga model dievaluasi menggunakan indikator *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), dan koefisien determinasi (R^2). Hasil penelitian ini diharapkan dapat memberikan gambaran mengenai keunggulan dan keterbatasan masing-masing pendekatan serta menjadi referensi dalam pemilihan model peramalan curah hujan yang lebih akurat untuk mendukung pengambilan keputusan pada bidang meteorologi, pengelolaan sumber daya air, dan mitigasi bencana hidrometeorologi.

2. Metode

Penelitian ini menggunakan pendekatan kuantitatif dengan metode komparatif untuk membandingkan performa model statistik, *machine learning*, dan *deep learning* dalam peramalan curah hujan bulanan di Kota Makassar. Model yang dibandingkan meliputi *Seasonal Autoregressive Integrated Moving Average with Exogenous Variables* (SARIMAX), *Extreme Gradient Boosting* (XGBoost), dan *Long Short-Term Memory* (LSTM). Ketiga model dipilih karena merepresentasikan tiga pendekatan yang berbeda dalam pemodelan deret waktu, yaitu model statistik, *ensemble machine learning*, dan *deep learning*, sehingga memungkinkan dilakukan evaluasi terhadap kemampuan masing-masing model dalam menangkap pola linier, musiman, maupun hubungan *nonlinear* antarvariabel meteorologi [6, 7, 9, 13, 14]. Tahapan penelitian meliputi pengumpulan data, *data preprocessing*, pembangunan model, evaluasi performa, serta analisis komparatif terhadap hasil peramalan.

2.1. Data Penelitian

Data yang digunakan merupakan data meteorologi bulanan Kota Makassar periode Januari 2000 hingga Desember 2024 yang diperoleh dari NASA POWER dengan total 300 observasi. Variabel penelitian terdiri atas curah hujan (*rainfall*) sebagai variabel target, serta suhu udara (*temperature*), kelembapan (*humidity*), kecepatan angin (*wind speed*), tekanan udara (*pressure*), dan radiasi matahari (*solar radiation*) sebagai variabel prediktor. Pemilihan variabel-variabel meteorologi tersebut didasarkan pada perannya dalam memengaruhi dinamika pembentukan curah hujan melalui interaksi proses atmosfer. Selanjutnya, data dibagi menjadi data *training* dan *testing* dengan rasio 80:20 untuk proses pelatihan dan pengujian model.

2.2. Tahapan Penelitian

Tahapan penelitian diawali dengan pengumpulan data meteorologi bulanan Kota Makassar, kemudian dilakukan *data preprocessing* yang meliputi pemeriksaan *missing value*, penanganan data apabila diperlukan, serta analisis statistik deskriptif untuk memahami karakteristik data. Selanjutnya, data dibagi menjadi data *training* dan *testing* dengan rasio 80:20, kemudian dilakukan analisis eksplorasi data (*Exploratory Data Analysis/EDA*) untuk mengidentifikasi pola, tren, musiman, dan hubungan antarvariabel.

Tahap berikutnya adalah pengujian stasioneritas data menggunakan uji *Augmented Dickey-Fuller* (ADF). Apabila data belum memenuhi asumsi stasioner, dilakukan proses *differencing* hingga diperoleh data yang stasioner. Untuk model SARIMAX, penentuan orde model dilakukan berdasarkan analisis plot *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF). Selanjutnya, dilakukan pembangunan model SARIMAX, XGBoost, dan LSTM menggunakan data *training*. Kinerja masing-masing model kemudian dievaluasi menggunakan *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), dan koefisien determinasi (R^2). Pada tahap akhir, dilakukan analisis komparatif terhadap hasil evaluasi untuk menentukan model yang memberikan performa peramalan terbaik pada data curah hujan bulanan di Kota Makassar.

2.3. Model Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX)

Model *Seasonal Autoregressive Integrated Moving Average with Exogenous Variables* (SARIMAX) merupakan pengembangan dari model *Autoregressive Integrated Moving Average* (ARIMA) dan *Seasonal Autoregressive Integrated Moving Average* (SARIMA) dengan penambahan variabel eksogen [6, 17]. Model ini digunakan untuk memodelkan data deret waktu yang memiliki pola tren, musiman, serta dipengaruhi oleh faktor eksternal. Bentuk umum model SARIMAX dinyatakan sebagai

$$\text{SARIMAX}(p, d, q)(P, D, Q)_s, \quad (1)$$

dengan p menyatakan orde *autoregressive* (AR), d orde *differencing* nonmusiman, q orde *moving average* (MA), P orde *seasonal autoregressive*, D orde *seasonal differencing*, Q orde *seasonal moving average*, dan s adalah periode musiman.

Model SARIMA secara umum dirumuskan sebagai berikut [17]:

$$\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B^s)^DY_t = \Theta_Q(B^s)\theta_q(B)\varepsilon_t, \quad (2)$$

dengan B menyatakan operator *backshift*, $\Phi_P(B^s)$ dan $\Theta_Q(B^s)$ berturut-turut merupakan polinomial *seasonal autoregressive* dan *seasonal moving average*, sedangkan $\phi_p(B)$ dan $\theta_q(B)$ merupakan polinomial *autoregressive* dan *moving average* nonmusiman.

Penambahan variabel eksogen menghasilkan model SARIMAX yang secara umum dirumuskan sebagai berikut:

$$Y_t = c + \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{k=1}^m \beta_k X_{k,t} + \varepsilon_t, \quad (3)$$

dengan Y_t adalah variabel curah hujan, $X_{k,t}$ adalah variabel eksogen ke- k , ϕ_i adalah parameter *autoregressive*, θ_j adalah parameter *moving average*, β_k adalah koefisien variabel eksogen, c adalah konstanta, dan ε_t adalah residual model.

Tahapan pembentukan model SARIMAX pada penelitian ini meliputi:

1. Pengujian stasioneritas data.

2. Melakukan proses *differencing* apabila data belum stasioner.
3. Mengidentifikasi parameter model menggunakan plot *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF).
4. Mengestimasi parameter model.
5. Melakukan uji diagnostik residual.
6. Melakukan peramalan (*forecasting*) curah hujan.

Uji stasioneritas dilakukan menggunakan uji *Augmented Dickey–Fuller* (ADF) yang dirumuskan sebagai berikut [18]:

$$\Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \sum_{i=1}^p \delta_i \Delta Y_{t-i} + \varepsilon_t, \quad (4)$$

dengan ΔY_t menyatakan data hasil perbedaan (*differencing*), α adalah konstanta, β adalah koefisien tren, γ adalah parameter yang digunakan untuk menguji keberadaan akar unit, δ_i adalah koefisien lag, dan ε_t adalah galat acak.

Tahap diagnostik model dilakukan menggunakan uji *Ljung–Box* untuk mengetahui apakah residual telah memenuhi asumsi *white noise*. Statistik uji *Ljung–Box* dirumuskan sebagai berikut [19]:

$$Q = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k}, \quad (5)$$

dengan n adalah jumlah observasi, h adalah banyaknya lag yang diuji, dan $\hat{\rho}_k$ adalah autokorelasi residual pada lag ke- k .

Model SARIMAX dipilih karena mampu menangkap pola tren dan musiman pada data curah hujan bulanan sekaligus mengakomodasi pengaruh variabel meteorologi eksternal, sehingga diharapkan menghasilkan akurasi peramalan yang lebih baik dibandingkan model deret waktu tanpa variabel eksogen.

2.4. Model Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) merupakan algoritma *ensemble learning* berbasis *gradient boosting* yang membangun pohon keputusan secara bertahap untuk meminimalkan kesalahan prediksi [9]. Metode ini dikenal memiliki performa yang tinggi pada data *nonlinear* serta mampu menangani hubungan kompleks antarvariabel [10]. Prediksi model XGBoost dinyatakan sebagai

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad (6)$$

dengan $\hat{y}_i$ adalah nilai prediksi, f_k adalah fungsi pohon keputusan ke- k , x_i adalah vektor masukan, dan K adalah jumlah pohon keputusan.

Fungsi objektif XGBoost dirumuskan sebagai berikut [9]:

$$\text{Obj} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k), \quad (7)$$

dengan $l(\cdot)$ merupakan fungsi kerugian (*loss function*) yang mengukur selisih antara nilai aktual dan nilai prediksi, sedangkan $\Omega(f_k)$ merupakan fungsi regularisasi.

Fungsi regularisasi pada XGBoost diberikan oleh

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2, \quad (8)$$

dengan T adalah jumlah daun pada pohon keputusan, w adalah bobot setiap daun, sedangkan γ dan λ merupakan parameter regularisasi yang digunakan untuk mengendalikan kompleksitas model.

Untuk mendukung transparansi dan reproduktibilitas penelitian, konfigurasi *hyperparameter* utama model XGBoost ditetapkan sebagai berikut: *learning rate* (η) sebesar 0.05, *max_depth* sebesar 5, *n_estimators* sebanyak 500, *subsample* sebesar 0.8, *colsample_bytree* sebesar 0.8, dan *min_child_weight* sebesar 3. Parameter regularisasi pada pers. (8) menggunakan $\lambda = 1.0$ dan $\gamma = 0$, sedangkan regularisasi L1 diterapkan melalui *reg_alpha* sebesar 0.1. Fungsi objektif yang digunakan adalah *reg:squarederror*. Proses pelatihan model menerapkan mekanisme *early stopping* berdasarkan galat pada data validasi dengan *early_stopping_rounds* sebesar 30 serta *random_state* sebesar 42 untuk menjamin reproduktibilitas. Berdasarkan konfigurasi tersebut, model terbaik diperoleh pada iterasi ke-277.

Model XGBoost dipilih karena memiliki kemampuan generalisasi yang baik, relatif tahan terhadap *overfitting*, serta efektif dalam memodelkan hubungan *nonlinear* pada data meteorologi yang memiliki karakteristik kompleks [9, 10].

2.5. Model Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) merupakan salah satu metode *deep learning* yang dikembangkan dari *Recurrent Neural Network* (RNN) untuk mempelajari pola temporal pada data deret waktu [14]. Model LSTM dirancang untuk mengatasi permasalahan *vanishing gradient* yang sering terjadi pada RNN konvensional, sehingga mampu mempertahankan informasi penting dalam ketergantungan jangka panjang (*long-term dependency*). Struktur utama LSTM terdiri atas *forget gate*, *input gate*, *cell state*, dan *output gate*.

Persamaan *forget gate* diberikan oleh

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f), \quad (9)$$

dengan f_t menyatakan keluaran *forget gate*, W_f adalah matriks bobot, b_f adalah bias, h_{t-1} adalah *hidden state* pada waktu sebelumnya, dan x_t adalah data masukan pada waktu ke- t . Persamaan *input gate* dinyatakan sebagai

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i), \quad (10)$$

sedangkan kandidat memori dirumuskan sebagai

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c). \quad (11)$$

Pembaruan *cell state* dilakukan menggunakan persamaan

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \quad (12)$$

dengan $\odot$ menyatakan operasi perkalian elemen demi elemen (*element-wise multiplication*). Persamaan *output gate* diberikan oleh

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), \quad (13)$$

dan *hidden state* pada waktu ke- t dihitung menggunakan

$$h_t = o_t \odot \tanh(C_t). \quad (14)$$

Pada penelitian ini, model LSTM menggunakan data sekuens dengan panjang *look-back* selama 12 bulan. Sebelum proses pelatihan, seluruh variabel dinormalisasi untuk meningkatkan stabilitas proses pembelajaran serta mempercepat konvergensi model [13, 15]. Arsitektur

model terdiri atas dua lapisan rekuren, yaitu lapisan *Bidirectional LSTM* pertama dengan 64 unit (*return_sequences = True*) dan lapisan LSTM kedua dengan 32 unit. Masing-masing lapisan diikuti oleh *Batch Normalization* dan *Dropout* dengan *dropout rate* sebesar 0.2 untuk mengurangi risiko *overfitting*. Kedua lapisan rekuren menggunakan fungsi aktivasi *tanh* dengan aktivasi rekuren *sigmoid*, serta regularisasi L2 sebesar 1×10^{-4} pada bobot *kernel*. Lapisan keluaran terdiri atas *Dense Layer* dengan 16 unit menggunakan aktivasi *ReLU*, kemudian diikuti oleh satu neuron keluaran (*Dense(1)*).

Proses optimasi dilakukan menggunakan algoritma Adam dengan *learning rate* sebesar 0.001 dan fungsi *loss* berupa *Huber loss*, yang lebih robust terhadap keberadaan *outlier* dibandingkan *Mean Squared Error*. Selain itu, *Mean Absolute Error* (MAE) digunakan sebagai metrik pemantauan selama proses pelatihan. Model dilatih selama maksimum 200 *epoch* dengan *batch size* sebesar 16. Untuk mencegah *overfitting*, diterapkan mekanisme *early stopping* dengan *patience* sebesar 25 *epoch* berdasarkan nilai *validation loss* (*val_loss*) serta opsi *restore_best_weights*. Penyesuaian *learning rate* dilakukan menggunakan skema *ReduceLROnPlateau* dengan faktor 0.5, *patience* 10 *epoch*, dan *minimum learning rate* sebesar 1×10^{-6} . Berdasarkan konfigurasi tersebut, proses pelatihan berhenti pada *epoch* ke-109 dengan model terbaik diperoleh pada *epoch* ke-84 (lihat Subbab 3.6).

Model LSTM dipilih karena memiliki kemampuan untuk mempelajari ketergantungan temporal jangka panjang serta menangkap hubungan *nonlinear* yang kompleks pada data curah hujan bulanan.

2.6. Skema Input Data dan Keadilan Perbandingan Antar Model

Sebagaimana dijelaskan pada pers. (1), (6), dan (9), ketiga model yang dibandingkan dalam penelitian ini menggunakan representasi data masukan yang berbeda sesuai dengan karakteristik arsitekturnya. Model SARIMAX menggunakan data deret waktu asli yang telah memenuhi asumsi stasioneritas melalui proses *differencing*, XGBoost menggunakan fitur tambahan berupa variabel *lag*, sedangkan LSTM menggunakan data sekuens dengan panjang *look-back* selama 12 bulan. Perbedaan representasi data tersebut bukan merupakan perlakuan yang disengaja untuk menguntungkan model tertentu, melainkan konsekuensi dari mekanisme internal masing-masing model dalam merepresentasikan ketergantungan temporal, sebagaimana lazim diterapkan pada penelitian peramalan variabel meteorologi yang membandingkan pendekatan statistik, *machine learning*, dan *deep learning* [11, 20–23].

Model SARIMAX memiliki struktur *autoregressive* dan *moving average* yang secara inheren mampu merepresentasikan autokorelasi, tren, serta pola musiman melalui komponen AR, MA, dan komponen musiman $(P, D, Q)_s$, sehingga variabel eksogen dapat dimasukkan secara langsung pada waktu ke- t tanpa memerlukan rekayasa fitur tambahan [20]. Sebaliknya, XGBoost merupakan model *ensemble* berbasis pohon keputusan yang tidak memiliki mekanisme memori terhadap urutan data. Oleh karena itu, ketergantungan temporal direpresentasikan melalui proses *feature engineering*, yaitu dengan membentuk fitur *lag* dari variabel target maupun variabel prediktor. Pendekatan ini merupakan praktik baku dalam penerapan model berbasis pohon pada data deret waktu dan telah banyak digunakan pada penelitian peramalan curah hujan [22, 23]. Adapun LSTM merupakan arsitektur RNN yang secara khusus dirancang untuk memproses data sekuensial melalui mekanisme *memory cell* dan sistem *gate*. Oleh sebab itu, penggunaan jendela *look-back* tetap selama 12 bulan, yang disesuaikan dengan periode musiman ($s = 12$) pada model SARIMAX, merupakan representasi masukan yang sesuai dengan prinsip kerja model LSTM dan tidak memerlukan pembentukan fitur *lag* secara eksplisit [14].

Meskipun menggunakan representasi data masukan yang berbeda, perbandingan perfor-

ma ketiga model tetap dilakukan secara adil berdasarkan beberapa pertimbangan. Pertama, seluruh model dibangun menggunakan himpunan data dasar yang sama, yaitu data meteorologi bulanan Kota Makassar periode Januari 2000 hingga Desember 2024 dengan variabel target dan variabel prediktor yang identik. Kedua, pembagian data *training* dan *testing* dilakukan secara kronologis tanpa proses pengacakan dengan rasio dasar 80:20. Perbedaan jumlah observasi efektif pada masing-masing model semata-mata disebabkan oleh kebutuhan pembentukan fitur *lag* maupun jendela *look-back*, bukan karena penggunaan periode pengujian yang berbeda secara substansial. Ketiga, seluruh model dievaluasi menggunakan data curah hujan pada skala asli, tanpa menggunakan data hasil *differencing* maupun normalisasi, sehingga nilai *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), dan koefisien determinasi (R^2) dapat dibandingkan secara langsung. Keempat, representasi data masukan pada masing-masing model mengikuti praktik yang umum digunakan dalam literatur peramalan curah hujan yang membandingkan pendekatan statistik, *machine learning*, dan *deep learning* [22]. Dengan demikian, perbedaan skema input pada penelitian ini mencerminkan representasi data yang paling sesuai bagi karakteristik masing-masing model dan tidak memberikan keuntungan yang tidak proporsional terhadap salah satu model.

3. Hasil dan Pembahasan

3.1. Data Preprocessing

Tahap awal penelitian dilakukan melalui proses *data preprocessing* untuk memastikan kualitas data sebelum dilakukan pembangunan model. Data yang digunakan merupakan data meteorologi bulanan Kota Makassar periode Januari 2000 hingga Desember 2024 dengan total 300 observasi. Variabel penelitian terdiri atas curah hujan (*rainfall*), suhu udara (*temperature*), kelembapan (*humidity*), kecepatan angin (*wind speed*), tekanan udara (*pressure*), dan radiasi matahari (*solar radiation*).

Berdasarkan hasil pemeriksaan data, tidak ditemukan *missing value* pada seluruh variabel sehingga data dapat langsung digunakan pada tahap analisis selanjutnya. Setelah itu, dilakukan analisis statistik deskriptif untuk memberikan gambaran umum mengenai karakteristik masing-masing variabel penelitian. Hasil statistik deskriptif disajikan pada Tabel 1.

Tabel 1. Statistik deskriptif data penelitian.

Variabel	Mean	Std	Min	Q1	Median	Q3	Max
Rainfall	5.1645	4.6676	0.0000	1.2050	3.9400	7.9150	22.0500
Temp	27.4387	0.7166	25.9600	26.9075	27.3400	27.9225	29.5000
Humidity	78.5382	6.1245	62.2600	73.9200	79.7850	83.2475	88.2300
Wind Speed	2.5705	0.5136	1.5700	2.1475	2.6100	2.9700	3.8600
Pressure	99.7842	0.0929	99.5200	99.7200	99.7800	99.8500	100.0000
Solar	5.3484	0.7569	3.4716	4.8441	5.2518	5.8667	7.1374

Berdasarkan Tabel 1, variabel curah hujan memiliki nilai rata-rata sebesar 5.1645 mm/bulan dengan simpangan baku sebesar 4.6676 mm/bulan. Nilai maksimum curah hujan mencapai 22.0500 mm/bulan, sedangkan nilai minimum sebesar 0.0000 mm/bulan. Perbedaan yang cukup besar antara nilai minimum dan maksimum serta besarnya simpangan baku menunjukkan bahwa curah hujan di Kota Makassar memiliki variasi yang tinggi selama periode pengamatan.

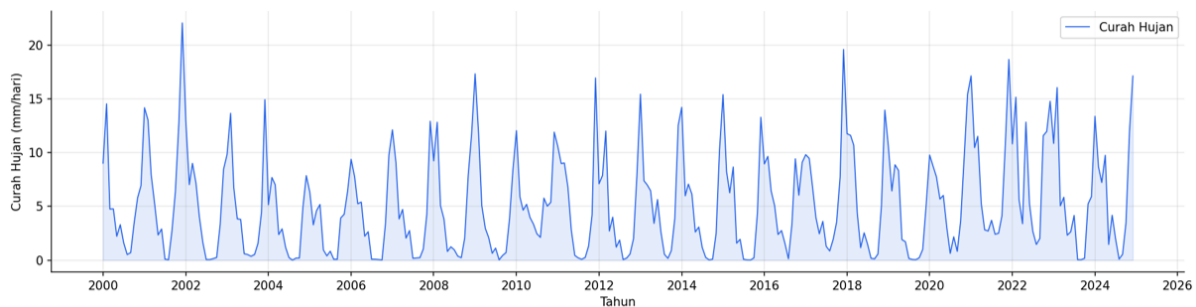
Sementara itu, variabel meteorologi lainnya menunjukkan tingkat variasi yang relatif lebih rendah dibandingkan curah hujan. Suhu udara memiliki rata-rata sebesar 27.4387°C dengan simpangan baku 0.7166°C, sedangkan kelembapan rata-rata mencapai 78.5382%. Kecepatan angin, tekanan udara, dan radiasi matahari juga memperlihatkan fluktuasi yang relatif

stabil selama periode pengamatan. Karakteristik tersebut mengindikasikan bahwa variabilitas utama dalam data penelitian berasal dari perubahan curah hujan, sementara variabel meteorologi lainnya berpotensi berperan sebagai variabel eksogen dalam proses pemodelan.

Selanjutnya, data dibagi secara kronologis menjadi data *training* dan data *testing* dengan rasio 80:20, sehingga diperoleh 240 observasi untuk pelatihan model dan 60 observasi untuk pengujian model. Pembagian secara kronologis dilakukan untuk mempertahankan struktur temporal data deret waktu dan menghindari terjadinya *data leakage* selama proses evaluasi model.

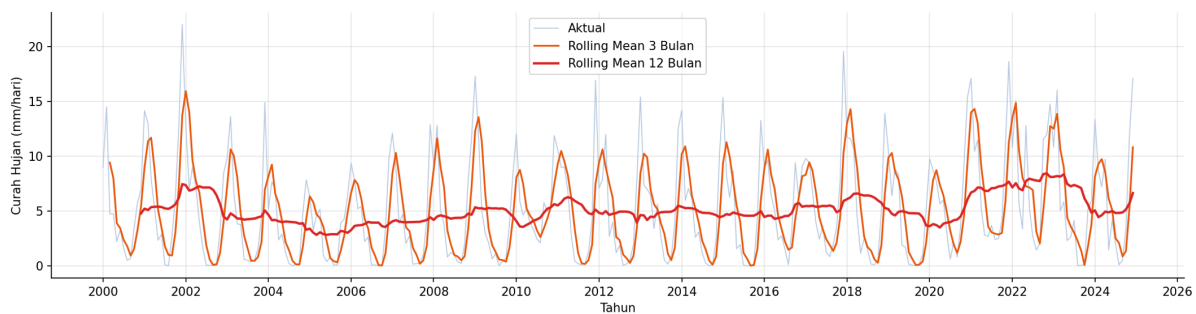
3.2. Exploratory Data Analysis (EDA)

Tahap *Exploratory Data Analysis* (EDA) dilakukan untuk memahami karakteristik data sebelum proses pemodelan. Analisis ini bertujuan mengidentifikasi pola deret waktu, tren, komponen musiman, serta hubungan antarvariabel meteorologi yang digunakan dalam penelitian. Selain itu, EDA juga berfungsi untuk memberikan gambaran awal mengenai karakteristik data sehingga pemilihan metode pemodelan dapat dilakukan secara lebih tepat. Hasil visualisasi pada tahap ini meliputi plot deret waktu, *rolling mean*, dekomposisi deret waktu, dan *heatmap correlation*.



Gambar 1. Plot deret waktu curah hujan bulanan Kota Makassar.

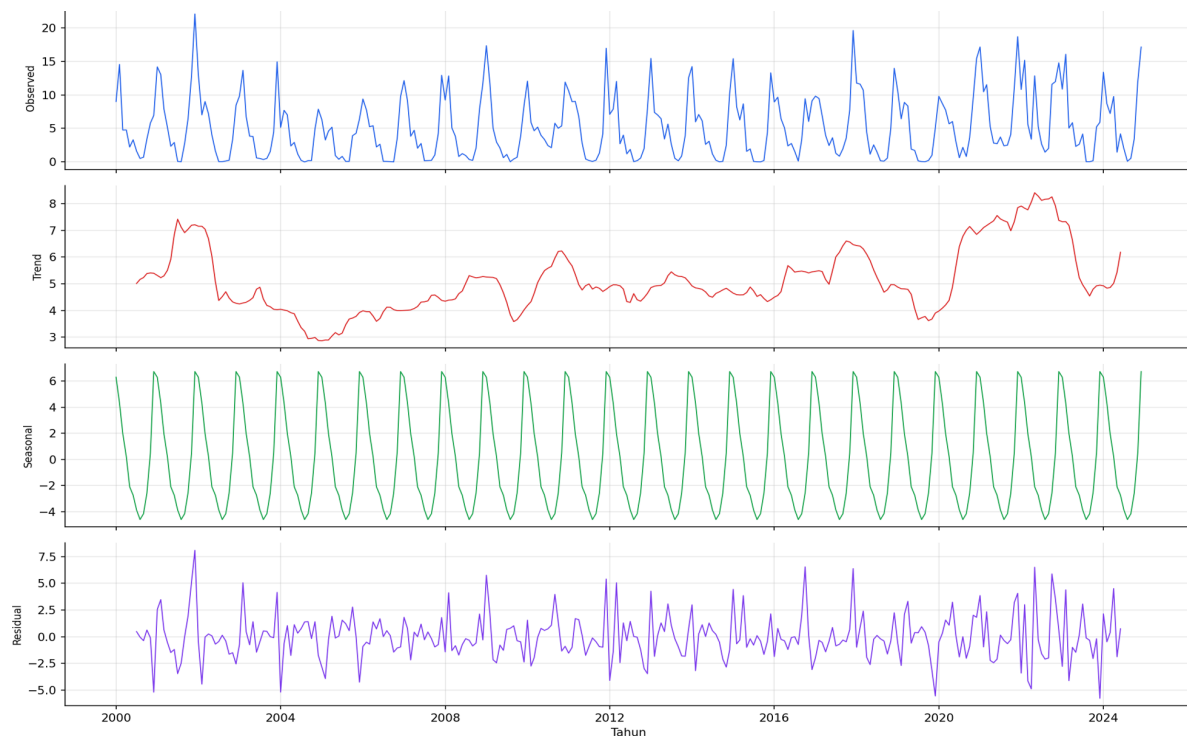
Berdasarkan **Gambar 1**, data curah hujan bulanan di Kota Makassar menunjukkan pola musiman yang berulang secara konsisten setiap tahun. Curah hujan cenderung meningkat pada awal dan akhir tahun, sedangkan nilai yang lebih rendah terjadi pada pertengahan tahun. Pola tersebut mencerminkan karakteristik iklim tropis musiman di wilayah Sulawesi Selatan yang dipengaruhi oleh pergantian musim hujan dan musim kemarau. Keberadaan pola musiman yang kuat mengindikasikan bahwa komponen musiman perlu dipertimbangkan dalam proses pemodelan. Hasil ini sejalan dengan penelitian Fitriati dkk. [1] yang melaporkan adanya pola musiman dominan pada data meteorologi di wilayah Indonesia Timur.



Gambar 2. Plot *rolling mean* curah hujan bulanan.

Gambar 2 menunjukkan bahwa nilai *rolling mean* bergerak secara periodik mengikuti pola musiman tahunan. Meskipun terjadi fluktuasi antarperiode, tidak terlihat adanya kecende-

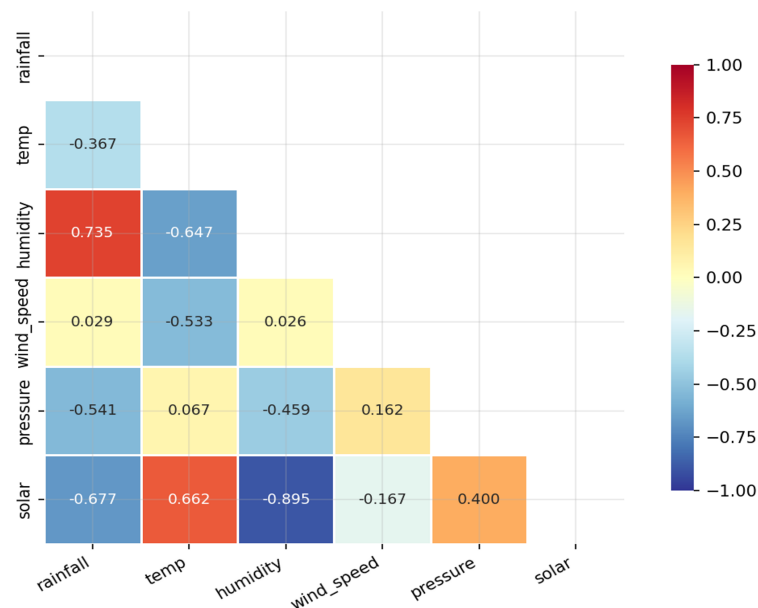
rungan peningkatan maupun penurunan yang signifikan selama periode pengamatan tahun 2000 hingga 2024. Hal ini mengindikasikan bahwa data memiliki tren jangka panjang yang relatif stabil, sementara komponen musiman menjadi karakteristik utama yang mendominasi variasi curah hujan.



Gambar 3. Dekomposisi deret waktu curah hujan bulanan.

Hasil dekomposisi deret waktu pada **Gambar 3** memperlihatkan bahwa data curah hujan dapat dipisahkan menjadi tiga komponen utama, yaitu tren, musiman, dan residual. Komponen tren menunjukkan perubahan yang relatif kecil sepanjang periode pengamatan, sedangkan komponen musiman memiliki pola yang konsisten dengan amplitudo yang relatif tetap setiap tahun. Adapun komponen residual masih memperlihatkan fluktuasi yang tidak beraturan, yang mengindikasikan adanya pengaruh faktor-faktor lain di luar pola musiman, seperti variabilitas iklim regional maupun fenomena atmosfer berskala global, misalnya *El Niño–Southern Oscillation* (ENSO).

Hubungan antarvariabel meteorologi ditunjukkan melalui *heatmap correlation* pada **Gambar 4**. Hasil analisis menunjukkan bahwa variabel kelembapan (*humidity*) memiliki korelasi positif yang relatif kuat terhadap curah hujan. Sebaliknya, variabel radiasi matahari (*solar radiation*) menunjukkan korelasi negatif yang cukup signifikan, yang mengindikasikan bahwa peningkatan intensitas radiasi matahari cenderung diikuti oleh penurunan curah hujan. Variabel suhu udara, kecepatan angin, dan tekanan udara memperlihatkan tingkat korelasi yang relatif lebih rendah terhadap curah hujan. Temuan ini menunjukkan bahwa kelembapan dan radiasi matahari berpotensi menjadi variabel eksogen yang penting dalam proses pemodelan. Hasil tersebut juga konsisten dengan penelitian Fitriati dkk. [1] yang melaporkan bahwa perubahan kelembapan dan radiasi matahari berkaitan erat dengan variasi pola curah hujan di wilayah Indonesia.



Gambar 4. Heatmap correlation antarvariabel meteorologi.

3.3. Uji Stasioneritas

Sebelum dilakukan pemodelan SARIMAX, data curah hujan diuji stasioneritasnya menggunakan uji *Augmented Dickey–Fuller* (ADF). Pengujian ini bertujuan untuk mengetahui apakah data telah memenuhi asumsi stasioneritas yang menjadi salah satu prasyarat utama dalam pemodelan deret waktu. Suatu data dikatakan stasioner apabila memiliki nilai rata-rata dan variansi yang relatif konstan sepanjang waktu serta tidak mengandung akar unit (*unit root*) [18]. Hasil uji ADF disajikan pada Tabel 2.

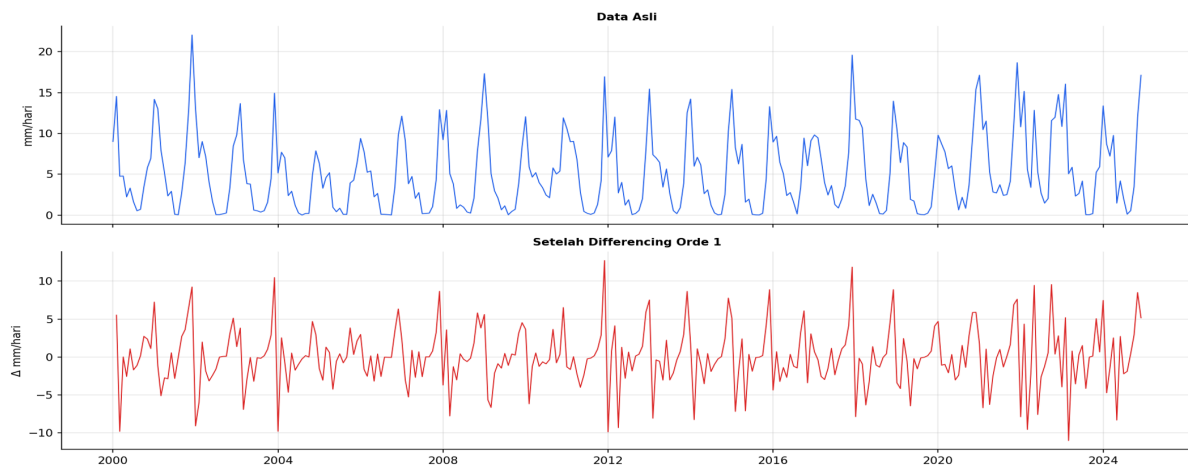
Tabel 2. Hasil uji *Augmented Dickey–Fuller* (ADF).

Data	ADF Statistic	p-value	Keterangan
Level	-2.524973	0.109516	Tidak stasioner
Differencing orde 1	-15.575281	0.000000	Stasioner

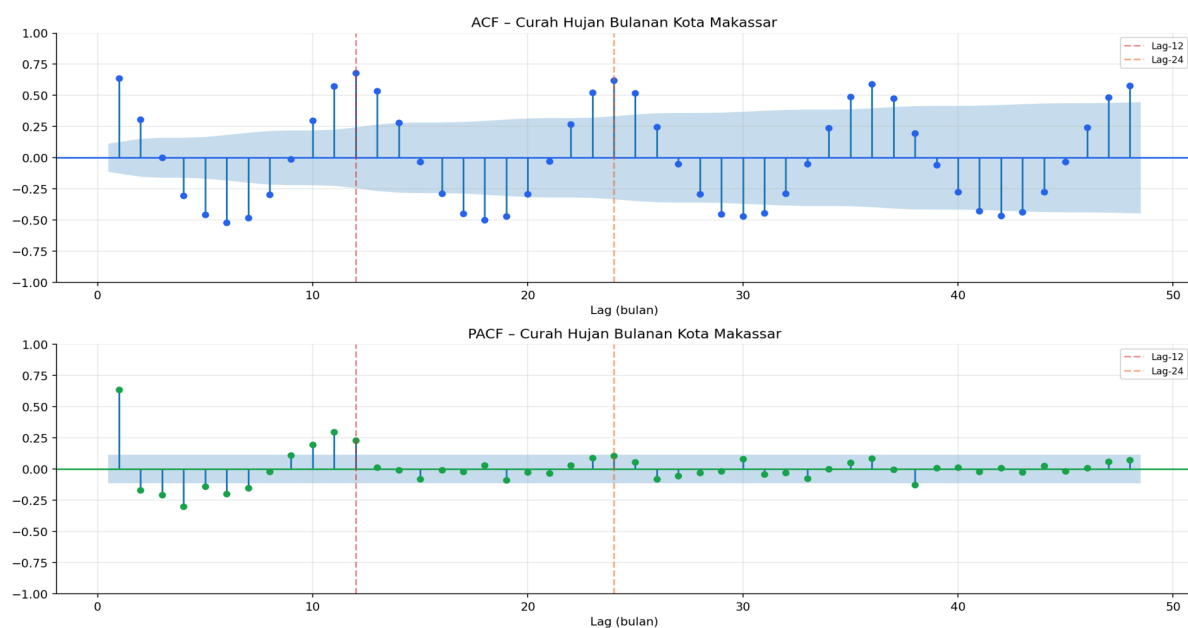
Berdasarkan Tabel 2, data curah hujan pada tingkat *level* memiliki nilai *p-value* sebesar 0.109516, yang lebih besar daripada taraf signifikansi 0.05. Dengan demikian, hipotesis nol yang menyatakan adanya akar unit tidak dapat ditolak, sehingga data belum memenuhi asumsi stasioneritas. Untuk mengatasi kondisi tersebut dilakukan proses *differencing* orde pertama. Setelah proses tersebut, nilai *p-value* menurun menjadi 0.000000 (< 0.05), sehingga hipotesis nol ditolak dan data dinyatakan telah stasioner [18]. Oleh karena itu, data hasil *differencing* digunakan pada proses identifikasi dan estimasi model SARIMAX.

Perubahan karakteristik data sebelum dan sesudah proses *differencing* ditunjukkan pada Gambar 5. Sebelum dilakukan *differencing*, data memperlihatkan fluktuasi musiman yang cukup besar dengan rata-rata yang berubah mengikuti waktu. Setelah dilakukan *differencing* orde pertama, data berfluktuasi di sekitar nilai nol dengan variasi yang lebih stabil. Hasil tersebut menunjukkan bahwa proses *differencing* berhasil menghilangkan komponen tren dan menghasilkan data yang memenuhi asumsi stasioneritas, sehingga layak digunakan dalam pemodelan SARIMAX.

Identifikasi orde awal model dilakukan menggunakan plot *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF) sebagaimana disajikan pada Gambar 6. Plot ACF me-



Gambar 5. Plot data curah hujan sebelum dan sesudah proses *differencing*.



Gambar 6. Plot Autocorrelation Function (ACF) dan Partial Autocorrelation Function (PACF).

nunjukkan pola periodik dengan lonjakan autokorelasi yang signifikan pada kelipatan lag ke-12, yaitu lag ke-12, 24, 36, dan 48. Pola tersebut mengindikasikan adanya komponen musiman tahunan yang kuat pada data curah hujan bulanan. Sementara itu, plot PACF memperlihatkan lonjakan signifikan pada lag pertama yang kemudian menurun secara cepat (*cut-off*), sehingga mengindikasikan keberadaan komponen *autoregressive* nonmusiman berorde rendah.

Berdasarkan karakteristik plot ACF dan PACF tersebut, ditetapkan beberapa kandidat model SARIMAX yang selanjutnya dievaluasi berdasarkan kriteria pemilihan model. Hasil identifikasi menunjukkan bahwa model SARIMAX(1, 1, 1)(1, 1, 1)₁₂ memberikan performa terbaik dan selanjutnya digunakan pada tahap estimasi parameter sebagaimana dibahas pada Subbab 3.4.

3.4. Hasil Pemodelan SARIMAX

Berdasarkan hasil identifikasi menggunakan plot *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF) pada Gambar 6, diperoleh model SARIMAX terbaik, yaitu SARIMAX(1, 1, 1)(1, 1, 1)₁₂. Model tersebut dibangun dengan memanfaatkan variabel eksogen berupa suhu udara (*temperature*), kelembapan (*humidity*), kecepatan angin (*wind speed*), te-

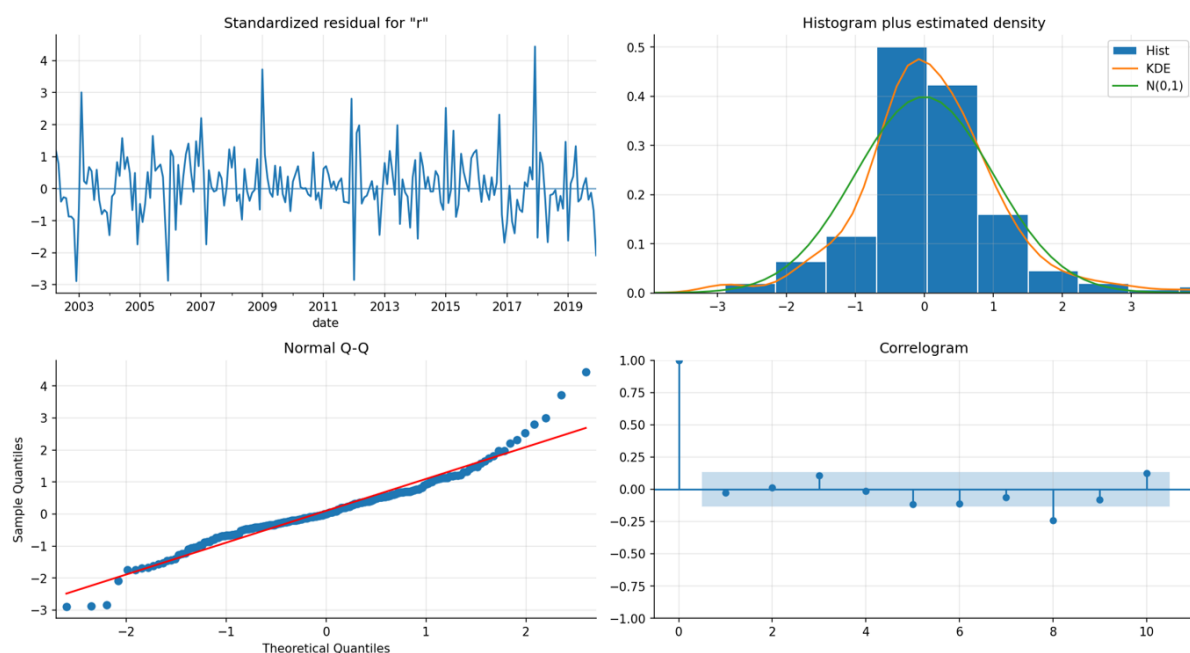
kanan udara (*pressure*), dan radiasi matahari (*solar radiation*). Hasil estimasi parameter model disajikan pada Tabel 3.

Tabel 3. Hasil estimasi parameter model SARIMAX.

Parameter	Koefisien	<i>p</i> -value
Temp	-0.1700	0.749
Humidity	-0.0025	0.979
Wind Speed	-0.0185	0.967
Pressure	0.8473	0.693
Solar	-3.7984	0.000
AR(1)	0.1911	0.008
MA(1)	-1.0874	0.000
SAR(12)	-0.1930	0.000
SMA(12)	-0.8683	0.000

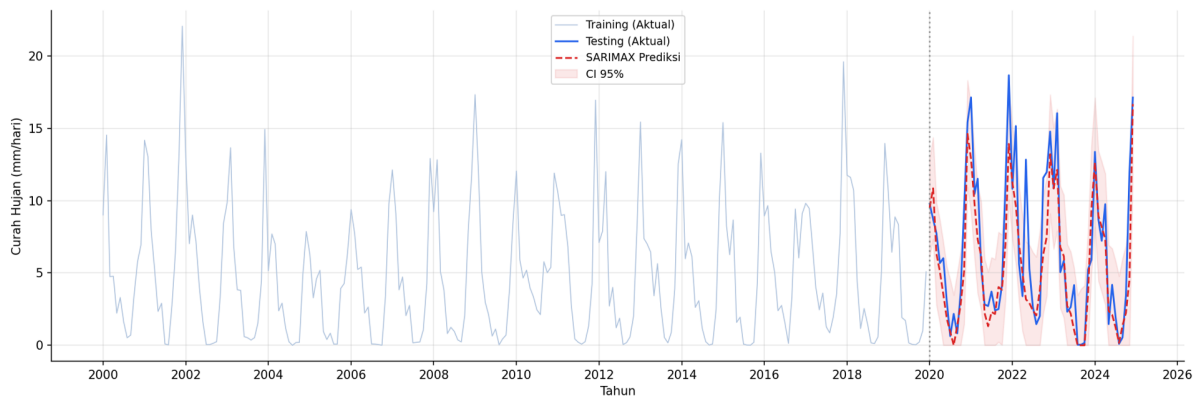
Berdasarkan Tabel 3, variabel radiasi matahari (*solar radiation*) merupakan satu-satunya variabel eksogen yang berpengaruh signifikan terhadap curah hujan dengan nilai *p*-value lebih kecil dari 0.05. Koefisien bernilai negatif menunjukkan bahwa peningkatan intensitas radiasi matahari cenderung diikuti oleh penurunan curah hujan. Sebaliknya, variabel suhu udara, kelembapan, kecepatan angin, dan tekanan udara belum menunjukkan pengaruh yang signifikan secara parsial pada taraf signifikansi 5%. Selain itu, seluruh parameter autoregresif dan musiman memiliki nilai *p*-value yang signifikan, sehingga mengindikasikan bahwa komponen autoregresif, *moving average*, dan musiman berperan penting dalam merepresentasikan dinamika curah hujan bulanan.

Setelah proses estimasi parameter, dilakukan uji diagnostik residual menggunakan uji *Ljung–Box*. Hasil pengujian menunjukkan nilai *Prob(Q)* sebesar 0.70, yang lebih besar dari taraf signifikansi 0.05. Dengan demikian, hipotesis nol tidak ditolak sehingga residual model memenuhi asumsi *white noise*. Hasil tersebut menunjukkan bahwa model SARIMAX telah mampu menangkap informasi yang terkandung dalam data dan layak digunakan untuk proses peramalan [19].



Gambar 7. Diagnostik residual model SARIMAX.

Hasil diagnostik residual model disajikan pada **Gambar 7**. Grafik residual menunjukkan penyebaran yang acak di sekitar nilai nol tanpa membentuk pola tertentu, sehingga tidak mengindikasikan adanya pelanggaran asumsi model. Selain itu, histogram residual memperlihatkan distribusi yang mendekati normal, sedangkan *Q-Q plot* menunjukkan sebagian besar titik berada di sekitar garis diagonal. Korelogram ACF residual juga tidak menunjukkan autokorelasi yang signifikan pada setiap lag. Temuan tersebut memperkuat hasil uji *Ljung-Box*, yang menyatakan bahwa residual telah memenuhi asumsi *white noise* [19].



Gambar 8. Perbandingan nilai aktual dan hasil prediksi model SARIMAX.

Perbandingan antara nilai aktual dan hasil prediksi model SARIMAX pada data pengujian ditunjukkan pada **Gambar 8**. Secara umum, model mampu mengikuti pola musiman curah hujan dengan baik, terutama dalam merepresentasikan periode musim hujan dan musim kemarau. Namun demikian, masih terdapat deviasi prediksi pada beberapa kejadian curah hujan ekstrem. Kondisi tersebut merupakan karakteristik umum model statistik linier, yang cenderung mengalami keterbatasan dalam memodelkan fluktuasi ekstrem maupun hubungan yang bersifat *nonlinear*.

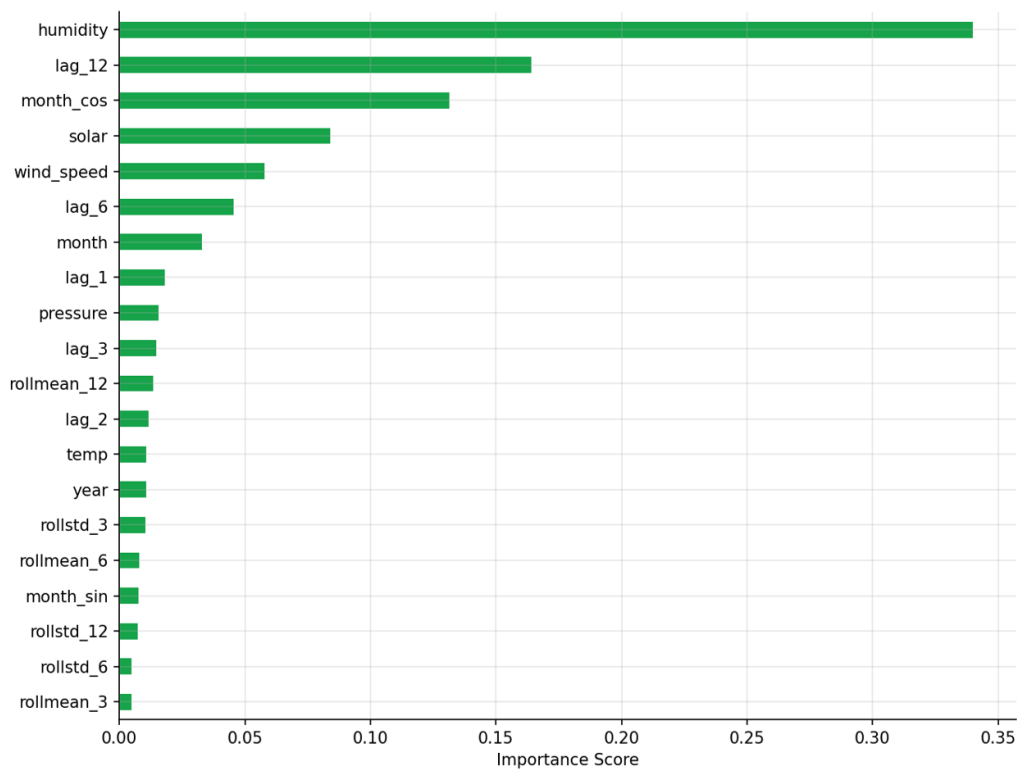
3.5. Hasil Pemodelan *Extreme Gradient Boosting* (XGBoost)

Model *Extreme Gradient Boosting* (XGBoost) dibangun menggunakan variabel meteorologi serta fitur *lag* dari variabel curah hujan. Pembentukan fitur *lag* bertujuan untuk merepresentasikan ketergantungan temporal sehingga model dapat memanfaatkan informasi dari periode sebelumnya dalam memprediksi curah hujan pada periode berikutnya [9, 10]. Pada penelitian ini dihasilkan sebanyak 20 fitur yang digunakan sebagai masukan model. Setelah pembentukan fitur, tersedia sebanyak 230 observasi sebagai data *training* dan 58 observasi sebagai data *testing*.

Proses pelatihan dilakukan menggunakan algoritma *gradient boosting*, yaitu dengan membangun pohon keputusan secara bertahap untuk meminimalkan kesalahan prediksi. Berdasarkan mekanisme *early stopping*, model terbaik diperoleh pada iterasi ke-277. Hasil tersebut menunjukkan bahwa setelah pembentukan 277 pohon keputusan, peningkatan performa model mulai konvergen sehingga proses pelatihan dihentikan untuk menghindari *overfitting*.

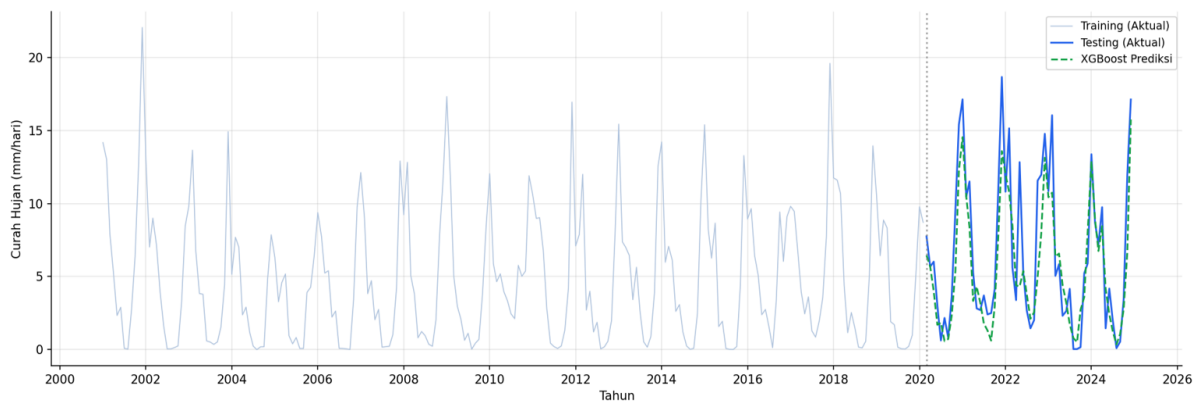
Secara umum, model XGBoost mampu mengikuti pola fluktuasi curah hujan dengan baik. Kemampuan tersebut menunjukkan bahwa pendekatan *boosting* efektif dalam memodelkan hubungan yang kompleks dan *nonlinear* antara curah hujan dengan variabel-variabel meteorologi [11]. Visualisasi hasil pemodelan disajikan pada **Gambar 9** dan **Gambar 10**.

Berdasarkan **Gambar 9**, fitur *lag* curah hujan (*rainfall_lag*) memiliki tingkat kepentingan (*feature importance*) paling tinggi dibandingkan fitur lainnya. Selanjutnya, variabel radiasi matahari dan kelembapan juga memberikan kontribusi yang cukup besar terhadap proses prediksi.



Gambar 9. Feature importance model XGBoost.

Dominasi fitur *lag* menunjukkan bahwa informasi curah hujan pada periode sebelumnya merupakan prediktor utama dalam memodelkan curah hujan periode berikutnya. Hasil tersebut sesuai dengan karakteristik data deret waktu yang memiliki ketergantungan temporal yang kuat.



Gambar 10. Perbandingan nilai aktual dan hasil prediksi model XGBoost.

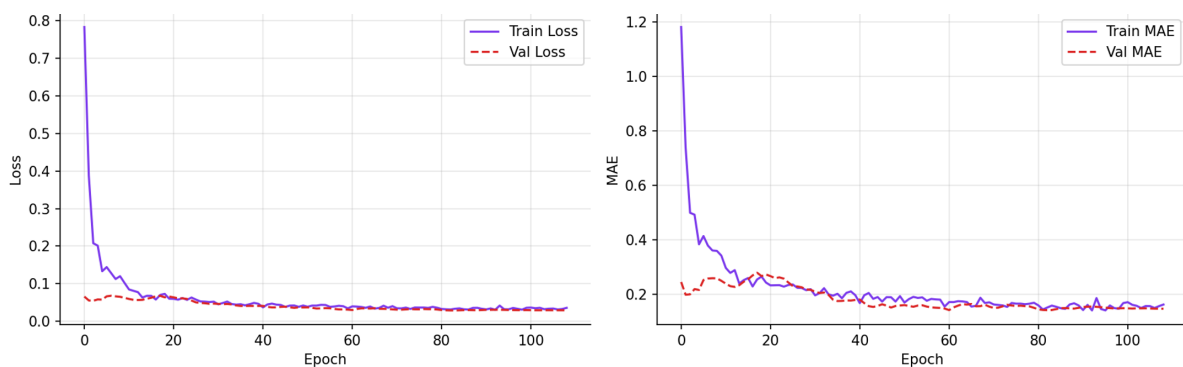
Perbandingan antara nilai aktual dan hasil prediksi model XGBoost pada data pengujian ditunjukkan pada **Gambar 10**. Secara umum, model mampu mengikuti pola musiman curah hujan dengan lebih baik dibandingkan model SARIMAX, terutama pada perubahan nilai curah hujan yang terjadi secara bertahap. Meskipun demikian, masih terdapat beberapa penyimpangan prediksi pada kejadian curah hujan yang sangat tinggi. Secara keseluruhan, hasil visualisasi menunjukkan bahwa model XGBoost memiliki kemampuan yang baik dalam merepresentasikan hubungan *nonlinear* antara variabel meteorologi dan curah hujan, sehingga menghasilkan prediksi yang lebih mendekati data aktual.

3.6. Hasil Pemodelan Long Short-Term Memory (LSTM)

Model Long Short-Term Memory (LSTM) digunakan untuk mempelajari ketergantungan temporal jangka panjang pada data curah hujan bulanan. Sesuai dengan rancangan model pada Bagian 2.5, penelitian ini menggunakan panjang *look-back* selama 12 bulan sehingga model memanfaatkan informasi curah hujan dan variabel meteorologi selama satu tahun sebelumnya untuk memprediksi curah hujan pada periode berikutnya [13]. Pendekatan ini memungkinkan model mempelajari pola musiman tahunan sekaligus hubungan temporal yang kompleks antardata.

Sebelum proses pelatihan dilakukan, seluruh variabel dinormalisasi untuk meningkatkan stabilitas proses pembelajaran dan mempercepat konvergensi model. Setelah proses pembentukan data sekuens, diperoleh data masukan berukuran $X_{\text{train}} = (228, 12, 6)$ dan $X_{\text{test}} = (60, 12, 6)$, yang menunjukkan bahwa model menggunakan 228 sekuens data pelatihan dengan panjang jendela waktu sebanyak 12 bulan dan enam variabel masukan. Arsitektur model terdiri atas lapisan *Bidirectional LSTM*, *Batch Normalization*, *Dropout*, dan *Dense Layer* dengan total 58.145 parameter yang dapat dilatih. Proses pelatihan dilakukan menggunakan maksimum 200 *epoch* dengan *batch size* sebesar 16. Untuk mencegah terjadinya *overfitting*, diterapkan mekanisme *early stopping*. Berdasarkan hasil pelatihan, proses berhenti secara otomatis pada *epoch* ke-109, sedangkan model terbaik diperoleh pada *epoch* ke-84.

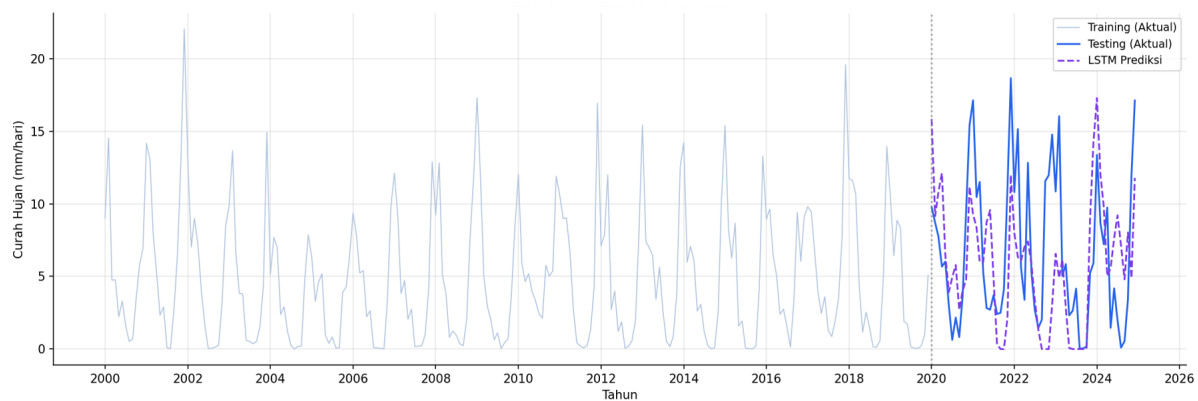
Selama proses pelatihan, nilai *training loss* dan *validation loss* menunjukkan tren penurunan yang cukup signifikan pada tahap awal pelatihan. Penurunan tersebut mengindikasikan bahwa model mampu mempelajari karakteristik data curah hujan. Setelah beberapa *epoch*, kedua nilai *loss* mulai konvergen sehingga mekanisme *early stopping* menghentikan proses pelatihan untuk menghindari penurunan kemampuan generalisasi model. Visualisasi proses pelatihan disajikan pada Gambar 11.



Gambar 11. Training history model LSTM.

Berdasarkan Gambar 11, kurva *training loss* dan *validation loss* mengalami penurunan yang cukup tajam pada awal proses pelatihan, kemudian cenderung stabil setelah sekitar 50 *epoch*. Kondisi tersebut menunjukkan bahwa model telah mencapai titik konvergensi. Meskipun demikian, masih terlihat adanya selisih antara *training loss* dan *validation loss*, yang mengindikasikan bahwa kemampuan generalisasi model belum sepenuhnya optimal. Hal ini menunjukkan masih terdapat peluang untuk meningkatkan performa model, misalnya melalui penambahan jumlah data pelatihan atau penyempurnaan konfigurasi *hyperparameter*.

Perbandingan antara nilai aktual dan hasil prediksi model LSTM pada data pengujian ditunjukkan pada Gambar 12. Secara umum, model telah mampu mengikuti kecenderungan perubahan curah hujan, namun belum dapat merepresentasikan fluktuasi data secara akurat, khususnya pada periode dengan curah hujan yang tinggi. Prediksi yang dihasilkan cenderung



Gambar 12. Perbandingan nilai aktual dan hasil prediksi model LSTM.

lebih halus (*smooth*) dibandingkan data aktual sehingga beberapa puncak curah hujan tidak dapat ditangkap dengan baik. Kondisi tersebut menunjukkan bahwa model masih mengalami keterbatasan dalam mempelajari pola yang kompleks pada data penelitian.

Performa LSTM yang belum optimal diduga dipengaruhi oleh jumlah data pelatihan yang relatif terbatas untuk pendekatan *deep learning*. Model berbasis jaringan saraf umumnya memerlukan jumlah data yang lebih besar agar mampu mempelajari representasi pola temporal secara efektif. Selain itu, kompleksitas arsitektur LSTM yang cukup tinggi juga meningkatkan kebutuhan terhadap data pelatihan yang memadai untuk menghasilkan kemampuan generalisasi yang baik [15]. Oleh karena itu, pada kondisi jumlah observasi yang terbatas, pendekatan *machine learning* seperti XGBoost berpotensi memberikan performa yang lebih baik dibandingkan model *deep learning*.

3.7. Evaluasi dan Perbandingan Model

Evaluasi performa model dilakukan untuk membandingkan tingkat akurasi model SARIMAX, XGBoost, dan LSTM dalam memprediksi curah hujan bulanan di Kota Makassar. Pada penelitian ini digunakan empat metrik evaluasi, yaitu *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), dan koefisien determinasi (R^2). Semakin kecil nilai RMSE, MAE, dan MAPE menunjukkan tingkat kesalahan prediksi yang semakin rendah, sedangkan nilai R^2 yang semakin mendekati satu menunjukkan kemampuan model yang semakin baik dalam menjelaskan variasi data.

Persamaan RMSE diberikan oleh

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}, \quad (15)$$

sedangkan MAE dirumuskan sebagai

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|, \quad (16)$$

MAPE dihitung menggunakan persamaan

$$\text{MAPE} = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|, \quad (17)$$

dan koefisien determinasi (R^2) diberikan oleh

$$R^2 = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y})^2}, \quad (18)$$

dengan y_t menyatakan nilai aktual, $\hat{y}_t$ menyatakan nilai prediksi, $\bar{y}$ merupakan rata-rata data aktual, dan n adalah jumlah observasi. Hasil evaluasi ketiga model disajikan pada **Tabel 4**.

Tabel 4. Hasil evaluasi performa model.

Model	RMSE	MAE	MAPE (%)	R^2
SARIMAX	2.6935	1.8534	39.2614	0.7880
XGBoost	2.5441	1.8815	118.7125	0.8063
LSTM	5.2125	4.1871	223.6393	0.1646

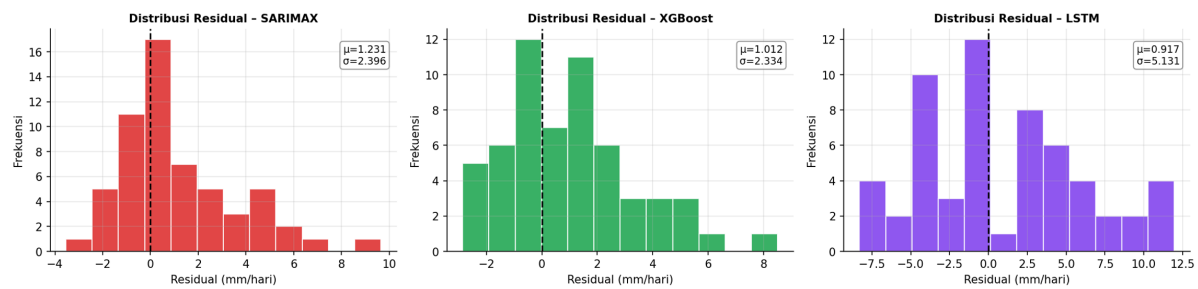
Berdasarkan **Tabel 4**, model XGBoost menghasilkan nilai RMSE terkecil, yaitu sebesar 2.5441, serta nilai koefisien determinasi (R^2) terbesar, yaitu 0.8063. Hasil tersebut menunjukkan bahwa XGBoost memiliki kemampuan terbaik dalam memodelkan variasi curah hujan bulanan dan menghasilkan prediksi yang paling mendekati data aktual. Kemampuan tersebut didukung oleh mekanisme *gradient boosting*, yang efektif dalam menangkap hubungan *nonlinear* antarvariabel meteorologi [21].

Di sisi lain, model SARIMAX menghasilkan nilai MAE terkecil sebesar 1.8534 dan nilai MAPE terendah sebesar 39.2614%. Nilai MAPE yang relatif rendah menunjukkan bahwa model SARIMAX memberikan tingkat kesalahan relatif yang lebih kecil dibandingkan dua model lainnya. Hasil ini mengindikasikan bahwa pendekatan statistik berbasis deret waktu masih sangat efektif dalam merepresentasikan pola musiman curah hujan yang relatif teratur [2, 6]. Meskipun nilai RMSE SARIMAX sedikit lebih besar dibandingkan XGBoost, perbedaannya relatif kecil sehingga kedua model memiliki tingkat akurasi yang sebanding.

Sementara itu, model LSTM menghasilkan performa yang paling rendah dengan nilai RMSE sebesar 5.2125, MAE sebesar 4.1871, MAPE sebesar 223.6393%, dan R^2 sebesar 0.1646. Nilai tersebut menunjukkan bahwa model LSTM belum mampu mempelajari pola curah hujan secara optimal pada data yang digunakan. Kondisi ini diduga disebabkan oleh jumlah data pelatihan yang relatif terbatas untuk pendekatan *deep learning*, sehingga kemampuan model dalam melakukan generalisasi belum maksimal [13].

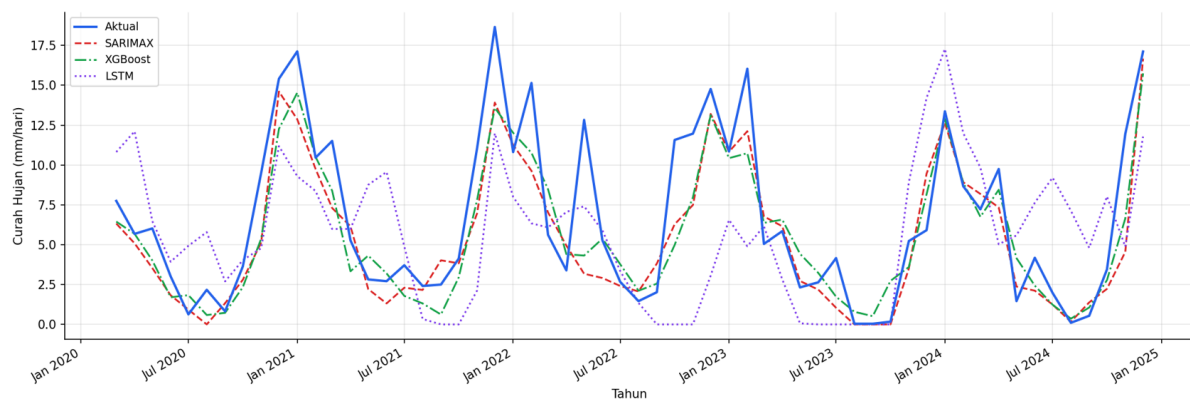
Perlu diperhatikan bahwa nilai MAPE pada model XGBoost dan LSTM relatif tinggi meskipun keduanya menghasilkan nilai RMSE yang berbeda secara signifikan. Kondisi ini disebabkan oleh adanya beberapa observasi curah hujan yang bernilai sangat kecil atau mendekati nol. Pada kondisi tersebut, penyebut pada *pers. (17)* menjadi sangat kecil sehingga menghasilkan persentase kesalahan yang sangat besar. Oleh karena itu, dalam penelitian ini interpretasi performa model lebih tepat didasarkan pada kombinasi metrik RMSE, MAE, dan R^2 , sedangkan nilai MAPE digunakan sebagai informasi pelengkap.

Secara keseluruhan, hasil penelitian menunjukkan bahwa setiap model memiliki karakteristik yang berbeda dalam memodelkan curah hujan bulanan. Model XGBoost memberikan performa terbaik dalam menangkap hubungan *nonlinear* antarvariabel meteorologi sehingga menghasilkan nilai RMSE dan R^2 terbaik. Model SARIMAX menunjukkan kemampuan yang baik dalam merepresentasikan pola musiman dengan tingkat kesalahan relatif yang rendah, sedangkan model LSTM belum memberikan performa yang kompetitif akibat keterbatasan jumlah data pelatihan. Temuan ini mengindikasikan bahwa pada data meteorologi dengan jumlah observasi yang relatif terbatas, pendekatan *machine learning* berbasis *boosting* lebih efektif dibandingkan pendekatan statistik klasik maupun *deep learning*.



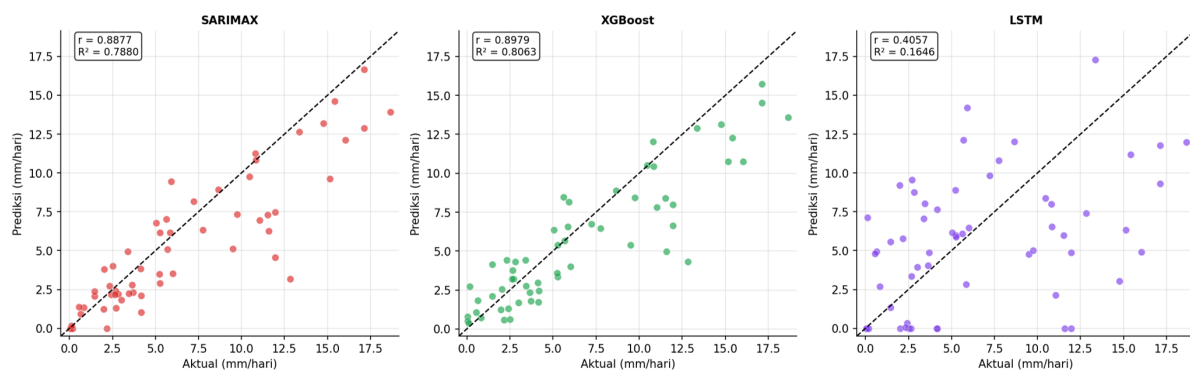
Gambar 13. Distribusi galat (*error*) ketiga model.

Distribusi galat masing-masing model ditunjukkan pada **Gambar 13**. Model XGBoost menghasilkan distribusi galat yang paling sempit dan terpusat di sekitar nol, yang menunjukkan konsistensi prediksi yang tinggi. Model SARIMAX juga menghasilkan distribusi galat yang relatif terpusat, meskipun dengan penyebaran yang sedikit lebih lebar. Sebaliknya, model LSTM memperlihatkan distribusi galat yang paling lebar dan kurang simetris, sehingga mengindikasikan tingkat kesalahan prediksi yang lebih besar.



Gambar 14. Overlay nilai aktual dan hasil prediksi ketiga model.

Perbandingan nilai aktual dan hasil prediksi ketiga model disajikan pada **Gambar 14**. Terlihat bahwa kurva prediksi XGBoost paling mendekati data aktual, terutama pada periode dengan perubahan curah hujan yang cukup besar. Model SARIMAX juga mampu mengikuti pola musiman dengan baik, namun cenderung menghasilkan amplitudo yang lebih rendah pada nilai curah hujan ekstrem. Adapun model LSTM menghasilkan penyimpangan yang lebih besar dibandingkan kedua model lainnya sehingga belum mampu mengikuti pola data secara optimal.



Gambar 15. Scatter plot antara nilai aktual dan hasil prediksi ketiga model.

Hubungan antara nilai aktual dan hasil prediksi masing-masing model ditunjukkan pada **Gambar 15**. Titik-titik hasil prediksi XGBoost tersebar paling dekat dengan garis diagonal ($y = x$), yang mengonfirmasi nilai R^2 tertinggi sebesar 0.8063. Model SARIMAX juga memperlihatkan penyebaran titik yang cukup dekat dengan garis diagonal dengan nilai R^2 sebesar 0.7880. Sebaliknya, titik-titik hasil prediksi LSTM tersebar lebih jauh dari garis diagonal, yang sejalan dengan nilai R^2 sebesar 0.1646 dan menunjukkan kemampuan prediksi yang masih terbatas.

3.8. Hasil Forecast Curah Hujan Tahun 2025

Setelah dilakukan evaluasi terhadap ketiga model, tahap selanjutnya adalah melakukan *forecast* curah hujan bulanan selama 12 bulan pada tahun 2025 menggunakan model SARIMAX, XGBoost, dan LSTM. Tujuan *forecast* ini adalah untuk memberikan gambaran pola curah hujan yang diprediksi oleh masing-masing model sekaligus membandingkan karakteristik hasil prediksi yang dihasilkan. Hasil *forecast* disajikan pada **Tabel 5**.

Tabel 5. *Forecast* curah hujan bulanan tahun 2025.

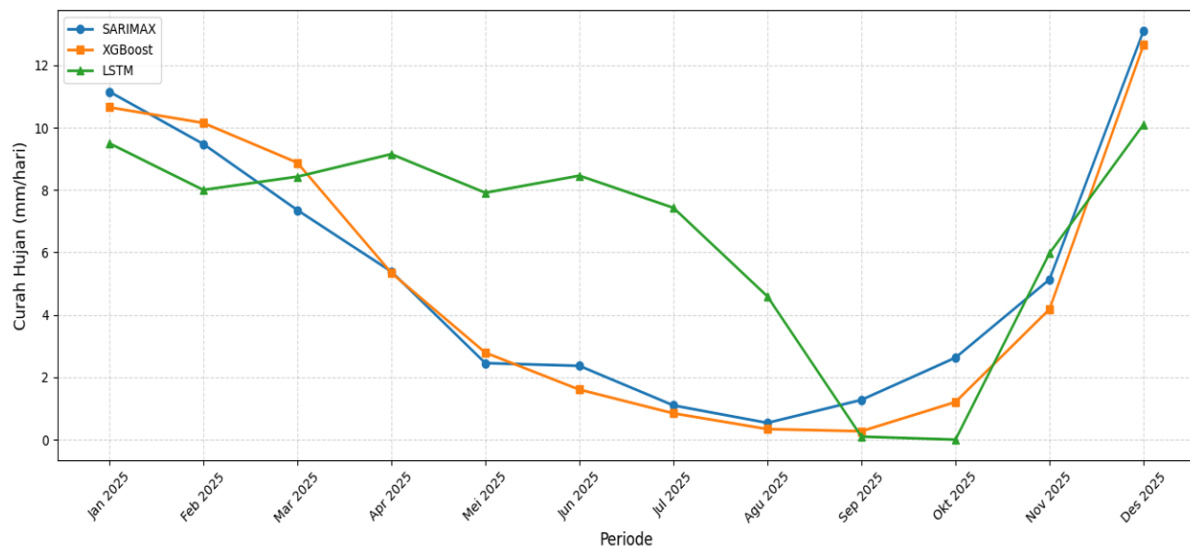
Bulan	SARIMAX	XGBoost	LSTM
Januari	11.1533	10.6488	9.4977
Februari	9.4742	10.1501	8.0022
Maret	7.3570	8.8757	8.4266
April	5.3860	5.3446	9.1490
Mei	2.4519	2.7869	7.9109
Juni	2.3649	1.6031	8.4587
Juli	1.0977	0.8439	7.4297
Agustus	0.5365	0.3371	4.5908
September	1.2737	0.2689	0.0970
Oktober	2.6260	1.2050	0.0000
November	5.1324	4.1797	5.9721
Desember	13.0993	12.6473	10.1014

Berdasarkan **Tabel 5**, model SARIMAX dan XGBoost menghasilkan pola *forecast* yang relatif serupa. Kedua model memprediksi curah hujan yang tinggi pada awal dan akhir tahun, sedangkan penurunan curah hujan terjadi pada pertengahan tahun, khususnya selama periode Juni hingga September. Pola tersebut sesuai dengan karakteristik iklim musiman di Kota Makassar yang dipengaruhi oleh pergantian musim hujan dan musim kemarau [2]. Kesesuaian pola tersebut menunjukkan bahwa kedua model mampu mempertahankan karakteristik musiman yang terdapat pada data historis.

Model SARIMAX menghasilkan pola prediksi yang relatif lebih halus dan stabil karena secara eksplisit memodelkan komponen musiman melalui struktur autoregresif musiman dan *moving average* musiman [6, 17]. Sebaliknya, model XGBoost menghasilkan pola yang sedikit lebih adaptif terhadap perubahan nilai curah hujan antarbulan tanpa menghilangkan pola musiman yang dominan. Karakteristik tersebut menunjukkan kemampuan XGBoost dalam menangkap hubungan *nonlinear* antara variabel meteorologi dan curah hujan, sehingga mampu menghasilkan prediksi yang tetap konsisten dengan pola historis.

Di sisi lain, model LSTM menghasilkan pola *forecast* yang berbeda dibandingkan dua model lainnya. Pada beberapa bulan yang secara historis merupakan periode musim kemarau, seperti Mei hingga Agustus, model LSTM masih memprediksi curah hujan yang relatif tinggi. Sebaliknya, pada bulan September dan Oktober model menghasilkan prediksi yang sangat rendah hingga mencapai nol. Pola tersebut mengindikasikan bahwa model belum mampu mempelajari karakteristik musiman secara optimal. Kondisi ini diduga berkaitan dengan ke-

terbatasan jumlah data pelatihan, sehingga kapasitas model *deep learning* yang relatif kompleks belum dapat dimanfaatkan secara maksimal [13, 15].



Gambar 16. Forecast curah hujan bulanan tahun 2025 menggunakan model SARIMAX, XGBoost, dan LSTM.

Visualisasi hasil *forecast* disajikan pada Gambar 16. Terlihat bahwa kurva hasil prediksi SARIMAX dan XGBoost mengikuti pola musiman yang konsisten dengan data historis, yaitu curah hujan yang tinggi pada awal dan akhir tahun serta menurun pada pertengahan tahun. Sementara itu, kurva hasil prediksi LSTM menunjukkan fluktuasi yang lebih besar dan kurang mampu merepresentasikan pola musim kemarau secara konsisten. Secara keseluruhan, hasil visualisasi memperkuat temuan pada tahap evaluasi model bahwa SARIMAX dan terutama XGBoost memberikan hasil *forecast* yang lebih realistis untuk data curah hujan bulanan di Kota Makassar.

Hasil penelitian ini memberikan kontribusi terhadap pengembangan metode peramalan curah hujan berbasis data meteorologi di Indonesia melalui perbandingan yang komprehensif antara pendekatan statistik klasik, *machine learning*, dan *deep learning*. Temuan penelitian menunjukkan bahwa model berbasis *boosting*, khususnya XGBoost, mampu memberikan performa yang lebih baik dalam memodelkan hubungan *nonlinear* antarvariabel meteorologi dibandingkan model statistik konvensional maupun model *deep learning* ketika jumlah data yang tersedia relatif terbatas. Selain memberikan kontribusi pada pengembangan metodologi peramalan deret waktu, hasil penelitian ini juga memiliki implikasi praktis sebagai informasi pendukung dalam pengambilan keputusan di Kota Makassar, seperti perencanaan pengelolaan sumber daya air, mitigasi banjir, penyusunan pola tanam pertanian, serta perencanaan infrastruktur yang dipengaruhi oleh kondisi curah hujan. Informasi *forecast* curah hujan bulanan juga dapat dimanfaatkan oleh instansi terkait sebagai dasar penyusunan strategi mitigasi risiko bencana hidrometeorologi yang lebih terukur dan berbasis data.

4. Kesimpulan

Berdasarkan hasil penelitian, dapat disimpulkan bahwa model *Seasonal Autoregressive Integrated Moving Average with Exogenous Variables* (SARIMAX), *Extreme Gradient Boosting* (XGBoost), dan *Long Short-Term Memory* (LSTM) dapat diterapkan untuk peramalan curah hujan bulanan di Kota Makassar dengan memanfaatkan variabel meteorologi berupa suhu udara, kelembapan, kecepatan angin, tekanan udara, dan radiasi matahari sebagai variabel prediktor. Hasil evalua-

si menunjukkan bahwa model XGBoost memberikan performa terbaik dengan nilai *Root Mean Square Error* (RMSE) sebesar 2.5441 dan koefisien determinasi (R^2) sebesar 0.8063, yang menunjukkan kemampuan terbaik dalam memodelkan variasi curah hujan bulanan. Sementara itu, model SARIMAX menghasilkan nilai *Mean Absolute Percentage Error* (MAPE) terkecil sebesar 39.2614%, yang mengindikasikan kemampuan yang baik dalam merepresentasikan pola musiman. Adapun model LSTM menunjukkan performa yang lebih rendah dibandingkan kedua model lainnya, yang mengindikasikan bahwa pendekatan *deep learning* memerlukan jumlah data yang lebih besar agar mampu menghasilkan prediksi yang optimal.

Secara keseluruhan, hasil penelitian menunjukkan bahwa pendekatan *machine learning* berbasis *boosting* lebih efektif dalam menangani hubungan *nonlinear* pada data curah hujan bulanan dibandingkan pendekatan statistik klasik maupun *deep learning* ketika jumlah data yang tersedia relatif terbatas. Temuan ini memberikan kontribusi terhadap pengembangan metode peramalan curah hujan berbasis data meteorologi, sekaligus menyediakan informasi yang dapat dimanfaatkan sebagai pendukung pengambilan keputusan dalam pengelolaan sumber daya air, mitigasi bencana hidrometeorologi, serta perencanaan sektor pertanian dan infrastruktur di Kota Makassar. Penelitian selanjutnya disarankan untuk menggunakan data dengan periode pengamatan yang lebih panjang, mengeksplorasi metode *hybrid* atau *ensemble*, serta mempertimbangkan penambahan variabel iklim global, seperti *El Niño–Southern Oscillation* (ENSO) dan *Indian Ocean Dipole* (IOD), guna meningkatkan akurasi model peramalan.

Kontribusi Penulis. Mohammad Zahid: Konseptualisasi, metodologi, perangkat lunak, kurasi data, analisis formal, investigasi, visualisasi, penulisan–penyusunan draf awal, penulisan–peninjauan dan penyuntingan. Rahmawati: Penulisan–peninjauan dan penyuntingan. Andi Seppewali: Perangkat lunak, analisis formal, visualisasi. Bintang Guntur: Validasi, supervisi, metodologi, penulisan–peninjauan dan penyuntingan, administrasi proyek, sumber daya.

Ucapan Terima Kasih. Penulis mengucapkan terima kasih kepada seluruh pihak yang telah memberikan dukungan selama pelaksanaan penelitian dan penyusunan manuskrip ini. Apresiasi juga disampaikan kepada NASA POWER sebagai penyedia data meteorologi yang digunakan dalam penelitian. Selain itu, penulis menyampaikan terima kasih kepada editor dan para reviewer atas masukan serta saran yang sangat bermanfaat dalam penyempurnaan manuskrip ini.

Pembiayaan. Penelitian ini tidak menerima pendanaan eksternal.

Konflik Kepentingan. Para penulis menyatakan bahwa tidak terdapat konflik kepentingan yang berkaitan dengan penelitian maupun publikasi artikel ini.

Ketersediaan Data. Data yang digunakan dalam penelitian ini diperoleh dari NASA POWER dan tersedia secara terbuka untuk keperluan penelitian.

Referensi

- [1] U. Fitriati, G. Rusmayadi, G. M. Hatta, D. Anggrainie, D. E. Adriani, S. Kadir, and Novitasari, “Climate change in north kalimantan province, indonesia,” in *IOP Conference Series: Earth and Environmental Science*. Institute of Physics, 2026, doi:10.1088/1755-1315/1593/1/012050.
- [2] F. Ghaly, A. Susrifalah, and Y. Kurniawati, “Peramalan curah hujan sebagai upaya mitigasi bencana menggunakan seasonal autoregressive integrated moving average,” *Jurnal Matematika dan Statistika serta Aplikasinya*, vol. 13, no. 1, pp. 1–8, 2025, doi:10.24252/msa.v13i1.55289.
- [3] H. A. Yusuf, I. Djakaria, and R. Resmawan, “Penerapan metode double moving average untuk meramalkan hasil produksi tanaman padi di provinsi gorontalo,” *d’CARTESIAN*, vol. 9, no. 2, pp. 92–99, 2020, doi:10.35799/dc.9.2.2020.29033.

- [4] W. A. Salmi, I. Djakaria, and R. Resmawan, "Penerapan metode exponential moving average pada peramalan penggunaan air di pdam kota gorontalo," *Jambura Journal of Probability and Statistics*, vol. 1, no. 2, pp. 69–77, 2020, doi:[10.34312/jjps.v1i2.5919](https://doi.org/10.34312/jjps.v1i2.5919).
- [5] A. M. Hamdani, F. Widhiatmoko, and S. Fitri, "Perbandingan akurasi metode autoregressive integrated moving average dan geometric brownian motion untuk peramalan harga saham indonesia," *Euler: Jurnal Ilmiah Matematika, Sains dan Teknologi*, vol. 13, no. 1, pp. 14–20, Apr 2025, doi:[10.37905/euler.v13i1.30760](https://doi.org/10.37905/euler.v13i1.30760).
- [6] H. Khaulasari and J. R. M. Akbar, "Modelling the effect of calendar variation in the gstarimax for predicting nitrogen monoxide air quality," *Euler: Jurnal Ilmiah Matematika, Sains dan Teknologi*, vol. 13, no. 3, pp. 360–371, Dec 2025, doi:[10.37905/euler.v13i3.33830](https://doi.org/10.37905/euler.v13i3.33830).
- [7] A. T. R. Dani and F. B. Putra, "Time series modeling with intervention analysis to evaluate of covid-19 impact on the stock markets in indonesia and global," *Euler: Jurnal Ilmiah Matematika, Sains dan Teknologi*, vol. 13, no. 1, pp. 113–126, Apr 2025, doi:[10.37905/euler.v13i1.31081](https://doi.org/10.37905/euler.v13i1.31081).
- [8] R. Y. Wawo, D. T. Salaki, H. A. H. Komalig, D. Hatidja, M. S. Paendong, and T. Manurung, "Perbandingan metode triple exponential smoothing additive dan additive parameter damped untuk peramalan indeks harga konsumen," *Euler: Jurnal Ilmiah Matematika, Sains dan Teknologi*, vol. 13, no. 1, pp. 77–83, Apr 2025, doi:[10.37905/euler.v13i1.30928](https://doi.org/10.37905/euler.v13i1.30928).
- [9] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 785–794.
- [10] S. Gnjatao, I. Leščešen, Q. Zhou, and M. Djukanović, "Explainable machine learning for streamflow forecasting: Application to the bosna river basin," *Water*, vol. 18, no. 10, May 2026, doi:[10.3390/w18101226](https://doi.org/10.3390/w18101226).
- [11] M. M. Hameed, A. Masood, A. Hamid, A. Elbeltagi, S. F. M. Razali, and A. Salem, "Forecasting monthly runoff in a glacierized catchment: A comparison of extreme gradient boosting (xgboost) and deep learning models," *PLoS One*, vol. 20, no. 5, May 2025, doi:[10.1371/journal.pone.0321008](https://doi.org/10.1371/journal.pone.0321008).
- [12] R. J. Buhungo, I. K. Hasan, and Nurwan, "Penerapan hybrid metode arfima-ann menggunakan algoritma backpropagation pada peramalan indeks harga saham gabungan," *Euler: Jurnal Ilmiah Matematika, Sains dan Teknologi*, vol. 12, no. 2, pp. 200–205, 2024, doi:[10.37905/euler.v12i2.28474](https://doi.org/10.37905/euler.v12i2.28474).
- [13] D. I. Puteri, "Implementasi long short term memory (lstm) dan bidirectional long short term memory (bilstm) dalam prediksi harga saham syariah," *Euler: Jurnal Ilmiah Matematika, Sains dan Teknologi*, vol. 11, no. 1, pp. 35–43, Jun 2023, doi:[10.34312/euler.v11i1.19791](https://doi.org/10.34312/euler.v11i1.19791).
- [14] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov 1997.
- [15] A. Damayanti and D. Agustina, "Implementasi metode adaptive neuro fuzzy inference system (anfis) dalam prediksi harga saham x," *Euler: Jurnal Ilmiah Matematika, Sains dan Teknologi*, vol. 12, no. 1, pp. 71–76, Jun 2024, doi:[10.37905/euler.v12i1.25278](https://doi.org/10.37905/euler.v12i1.25278).
- [16] M. Megawati, B. Sartono, and S. D. Oktarina, "A study on prediction intervals produced using quantile regression forest with and without variable selection," *Euler: Jurnal Ilmiah Matematika, Sains dan Teknologi*, vol. 13, no. 3, pp. 319–327, Dec 2025, doi:[10.37905/euler.v13i3.34392](https://doi.org/10.37905/euler.v13i3.34392).
- [17] G. T. Wilson, "Time series analysis: Forecasting and control, 5th edition, by george e. p. box, gwilym m. jenkins, gregory c. reinsel and greta m. ljung, 2015," *Journal of Time Series Analysis*, vol. 37, no. 5, pp. 709–711, Sep 2016, doi:[10.1111/jtsa.12194](https://doi.org/10.1111/jtsa.12194).
- [18] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, vol. 74, no. 366, pp. 427–431, Jun 1979.
- [19] G. M. Ljung and G. E. P. Box, "On a measure of lack of fit in time series models," *Biometrika*, vol. 65, no. 2, pp. 297–303, Aug 1978.
- [20] C. Kişmıroğlu and O. Isik, "Temperature prediction using transformer-lstm deep learning models and sarimax from a signal processing perspective," *Applied Sciences*, vol. 15, no. 17, Sep 2025, doi:[10.3390/app15179372](https://doi.org/10.3390/app15179372).
- [21] M. Putra, M. S. Rosid, and D. Handoko, "High-resolution rainfall estimation using ensemble learning techniques and multisensor data integration," *Sensors*, vol. 24, no. 15, Aug 2024, doi:[10.3390/s24155030](https://doi.org/10.3390/s24155030).
- [22] P. Mishra, S. Ray, P. Lal, S. B. Nair, A. Matuka, Y. Tashkandy, and W. Emam, "Climate modeling for south asia: Statistical and deep learning for rainfall and temperature prediction," *Scientific Reports*, vol. 15, no. 1, Dec 2025, doi:[10.1038/s41598-025-22149-1](https://doi.org/10.1038/s41598-025-22149-1).
- [23] M. E. Hafyani, K. E. Himdi, and S. E. E. Adlouni, "Improving monthly precipitation prediction accuracy using machine learning models: A multi-view stacking learning technique," *Frontiers in Water*, vol. 6, 2024, doi:[10.3389/frwa.2024.1378598](https://doi.org/10.3389/frwa.2024.1378598).