

ANALISIS PERBANDINGAN MODEL SARIMAX, RANDOM FOREST, DAN HYBRID SARIMAX–RF PADA PERAMALAN CURAH HUJAN BULANAN DI KOTA MAJENE TAHUN 2025

M. Rifandi¹, Mohammad Zahid^{2*}, Bintang Guntur³,

^{1,2,3} Program Studi Matematika, Fakultas Matematika Dan Ilmu Pengetahuan Alam, Universitas Sulawesi Barat

*e-mail: Zahidmatematika22@gmail.com

DOI: 10.14710/j.gauss.15.2.368-378

Article Info:

Received: 2026-04-25

Accepted: 2026-09-18

Available Online: 2026-09-30

Keywords:

Rainfall forecasting; SARIMAX; Random Forest; *Hybrid* model; time series.

Abstract: This study compares the performance of three forecasting models: SARIMAX, *Random Forest* (RF), and a *Hybrid* SARIMAX–RF approach in predicting monthly rainfall in Majene City for the year 2025. The dataset consists of five years of monthly rainfall and relative humidity data (2020–2024), sourced from BMKG and national meteorological records. The SARIMAX model captures seasonal trends and includes humidity as an exogenous variable, while the RF model addresses nonlinear relationships using lag-based predictors. A *Hybrid* model is then developed by combining SARIMAX outputs with RF-learned residual corrections. The forecasting performance is evaluated using standard error metrics such as MAE, RMSE, and MAPE on 2024 test data before forecasting 2025. Results show that the *Hybrid* model significantly outperforms the individual models, achieving the lowest MAPE (23.31%) compared to SARIMAX (42.21%) and RF (41.23%), and successfully capturing both seasonal patterns and irregular variations. The study demonstrates the importance of model integration for improving rainfall prediction accuracy, especially in tropical coastal regions with complex weather dynamics. Future studies are encouraged to explore additional *Hybrid* architectures and incorporate other climatic variables to further enhance predictive performance.

1. PENDAHULUAN

Curah hujan merupakan komponen penting dalam sistem hidrologi yang memengaruhi ketersediaan sumber daya air, pertanian, pengendalian banjir, dan perencanaan infrastruktur. Kota Majene sebagai wilayah pesisir memiliki variabilitas curah hujan bulanan yang dipengaruhi oleh dinamika iklim global dan lokal, termasuk *El Niño–Southern Oscillation* (ENSO) dan *Indian Ocean Dipole* (IOD) (Dotse dkk., 2024). Penelitian di Indonesia juga menunjukkan bahwa curah hujan memiliki karakteristik musiman yang kompleks sehingga memerlukan pendekatan prediktif yang mampu menangkap pola temporal secara efektif (Ruslan dan Pratomo, 2022; Widodo, Rahman dan Sari, 2022). Ketidakpastian tersebut dapat menghambat perencanaan pertanian, pengelolaan sumber daya air, dan mitigasi bencana.

Berbagai metode telah digunakan untuk meningkatkan akurasi peramalan curah hujan, mulai dari pendekatan statistik hingga *machine learning*. *Seasonal Autoregressive Integrated Moving Average with exogenous variables* (SARIMAX) mampu memodelkan pola musiman dan pengaruh variabel eksogen pada data deret waktu (Pande, Singh dan Mulla, 2024), sedangkan *Random Forest* (RF) mampu menangkap hubungan nonlinier dan interaksi kompleks pada data meteorologi (Siqueira dkk., 2023). Penerapan *machine learning* untuk peramalan curah hujan di Indonesia juga menunjukkan hasil yang menjanjikan, antara lain melalui RF dan *Neural Network* pada beberapa wilayah (Indaryono dan Setiawan, 2024; Rahmayani dan Kurniawan, 2024).

Keterbatasan masing-masing pendekatan mendorong pengembangan model *hybrid* yang menggabungkan kemampuan model statistik dalam menangkap pola linier dan

musiman dengan kemampuan *machine learning* dalam memodelkan pola nonlinier (Alizadeh dan Farajzadeh, 2018; Oliveira dkk., 2025). Sejumlah penelitian menunjukkan bahwa kombinasi model statistik dan *machine learning*, seperti ARIMA–*Neural Network*, dapat menghasilkan akurasi lebih baik dibandingkan model tunggal (Sari, Wibowo dan Hartono, 2021; Hartono dan Wijayanti, 2023). Penelitian lain juga melaporkan peningkatan kinerja melalui kombinasi SARIMAX–LSTM, Wavelet–SARIMAX–LSSVM, dan SARIMAX–XGBoost pada peramalan hidrometeorologi (Shah dkk., 2024; Gurjar dan Ghosh, 2025).

Meskipun demikian, penelitian yang secara khusus membandingkan SARIMAX, RF, dan *Hybrid* SARIMAX–RF untuk peramalan curah hujan bulanan di wilayah pesisir Indonesia, khususnya Kota Majene, masih terbatas. Karakteristik iklim lokal, interaksi laut–daratan, dan kondisi topografi dapat memengaruhi pola curah hujan serta kinerja model prediksi (Singh dan Choudhary, 2025). Kondisi tersebut menunjukkan perlunya evaluasi komparatif terhadap ketiga pendekatan pada karakteristik wilayah yang spesifik.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis kinerja model *Seasonal Autoregressive Integrated Moving Average with Exogenous Variables* (SARIMAX), *Random Forest* (RF), dan *Hybrid* SARIMAX–RF dalam memprediksi curah hujan bulanan di Kota Majene untuk periode tahun 2025. Selain itu, penelitian ini bertujuan untuk membandingkan tingkat akurasi ketiga model tersebut berdasarkan metrik evaluasi yang relevan sehingga dapat diketahui model yang memberikan kinerja prediksi terbaik dalam peramalan curah hujan bulanan di Kota Majene.

Kebaruan (*novelty*) penelitian ini terletak pada penerapan dan perbandingan langsung ketiga pendekatan SARIMAX, RF, dan SARIMAX–RF untuk kasus spesifik curah hujan bulanan di Kota Majene, dengan mempertimbangkan kondisi iklim lokal dan variabel eksogen yang relevan. Hasil penelitian diharapkan dapat memberikan kontribusi pada pengembangan metode peramalan curah hujan yang lebih akurat, yang pada akhirnya mendukung pengambilan keputusan di sektor pertanian, pengelolaan sumber daya air, dan mitigasi bencana di wilayah pesisir Indonesia.

2. TINJAUAN PUSTAKA

Peramalan curah hujan bulanan bertujuan memprediksi jumlah presipitasi pada periode tertentu menggunakan data historis dan variabel eksternal yang berpengaruh (Costa, Menezes Filho dan Canales, 2023). Data curah hujan bulanan biasanya memiliki pola musiman dan dipengaruhi faktor iklim global seperti ENSO dan IOD (Dotse dkk., 2024). Metode peramalan yang efektif harus mampu menangkap komponen musiman (seasonal), tren linier, dan pola non-linier.

Salah satu metode yang mampu menangkap komponen musiman dan pengaruh variabel eksogen adalah model SARIMAX. *Seasonal AutoRegressive Integrated Moving Average with exogenous variables* (SARIMAX) merupakan pengembangan model ARIMA yang mengakomodasi komponen musiman dan variabel eksogen. Bentuk umum model SARIMAX $(p, d, q)(P, D, Q)_s$ dinyatakan pada Persamaan (1):

$$\Phi_p(L^s)\phi_p(L)(1-L)^d(1-L^s)^D y_t = \Theta_q(L^s)\theta_q(L)\varepsilon_t + \beta X_t \quad (1)$$

Pada Persamaan (1), y_t merupakan variabel respons, X_t merupakan variabel eksogen, dan ε_t merupakan residual. Parameter p, d, q masing-masing menunjukkan orde AR, *differencing*, dan MA nonmusiman, sedangkan P, D, Q menunjukkan orde AR, *differencing*, dan MA musiman. Parameter s menunjukkan periode musiman, dengan $s = 12$ untuk data bulanan. Parameter β merupakan koefisien regresi variabel eksogen.

Polinomial AR nonmusiman dan musiman masing-masing dinyatakan pada Persamaan (2) dan (3):

$$\phi p(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p \quad (2)$$

$$\Phi P(L^s) = 1 - \Phi_1 L^s - \Phi_2 L^{2s} - \dots - \Phi_P L^{Ps} \quad (3)$$

Sementara itu, polinomial MA nonmusiman dan musiman dinyatakan pada Persamaan (4) dan (5):

$$\theta_q(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q \quad (4)$$

$$\Theta_Q(L^s) = 1 + \theta_1 L^s + \theta_2 L^{2s} + \dots + \theta_Q L^{Qs} \quad (5)$$

SARIMAX efektif untuk pola linier musiman, namun kurang optimal dalam menangkap hubungan non-linier yang kompleks (Oliveira dkk., 2025). Penelitian di berbagai kota Indonesia seperti Makassar, Bandung, dan Bengkulu menunjukkan bahwa model SARIMAX efektif untuk menangkap pola musiman curah hujan tropis, dengan kelembapan dan variabel meteorologi lainnya sebagai faktor eksogen yang signifikan (Anwar dan Putri, 2022; Hakiqi, Susanto dan Pratiwi, 2023).

Selain pendekatan statistik tersebut, penelitian ini juga menggunakan pendekatan berbasis machine learning untuk menangkap hubungan nonlinier antarvariabel. *Random Forest* adalah algoritma *ensemble learning* yang membangun banyak *decision tree* secara acak, lalu menggabungkan hasil prediksinya melalui voting (klasifikasi) atau rata-rata (regresi).

Persamaan prediksi *Random Forest* dinyatakan pada Persamaan (6):

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (6)$$

Keterangan:

B = jumlah pohon keputusan (*decision trees*)

$T_b(x)$ = prediksi dari pohon ke- b untuk input x

$\hat{y}$ = rata-rata prediksi semua pohon

RF unggul untuk data non-linier dan interaksi variabel kompleks, tetapi tidak menangani musiman secara eksplisit (Siqueira dkk., 2023). Penerapan *Random Forest* untuk klasifikasi dan prediksi curah hujan di Indonesia telah menunjukkan hasil yang baik, terutama dalam menangani ketidakseimbangan data curah hujan dan variabilitas iklim regional (Akbar dan Sanjaya, 2023; Nugroho dan Sari, 2024).

Berdasarkan keunggulan dan kekurangan tersebut, kedua model dapat dikombinasikan agar saling melengkapi dalam menangkap pola linier maupun nonlinier pada data curah hujan. Pendekatan *Hybrid SARIMAX–RF* menggabungkan model *Seasonal Autoregressive Integrated Moving Average with Exogenous Variables* (SARIMAX) dan *Random Forest* (RF) melalui pemodelan residual. Model SARIMAX digunakan untuk menangkap pola linier dan musiman pada deret waktu curah hujan. Selanjutnya, residual dari model SARIMAX digunakan sebagai variabel target pada *Random Forest* untuk mempelajari pola yang belum mampu dijelaskan oleh model SARIMAX. Prediksi akhir diperoleh dengan menggabungkan prediksi SARIMAX dan prediksi residual *Random Forest*.

Prediksi akhir model *Hybrid SARIMAX–RF* dinyatakan pada Persamaan (7) sebagai berikut:

$$\hat{y}_t^{Hybrid} = \hat{y}_t^{SARIMAX} + \hat{e}_t^{RF} \quad (7)$$

dengan $\hat{y}_t^{Hybrid}$ merupakan nilai curah hujan pada waktu t , $\hat{y}_t^{SARIMAX}$ merupakan hasil prediksi model SARIMAX, dan $\hat{e}_t^{RF}$ merupakan prediksi residual yang dihasilkan oleh *Random Forest*.

Residual model SARIMAX dihitung sebagai selisih antara nilai aktual dan nilai prediksi SARIMAX, sebagaimana dinyatakan pada Persamaan (8):

$$e_t = y_t - \hat{y}_t^{SARIMAX} \quad (8)$$

dengan e_t merupakan residual pada waktu t , y_t merupakan nilai aktual curah hujan, dan $\hat{y}_t^{SARIMAX}$ merupakan nilai prediksi yang dihasilkan oleh model SARIMAX. Residual tersebut kemudian digunakan sebagai target yang dipelajari oleh *Random Forest*.

Prediksi residual menggunakan *Random Forest* dinyatakan pada Persamaan (9) sebagai berikut:

$$\hat{e}_t^{RF} = f_{RF}(Z_t) \quad (9)$$

dengan $\hat{e}_t^{RF}$ merupakan prediksi residual menggunakan Random Forest, $f_{RF}(\cdot)$ merupakan fungsi prediksi Random Forest, dan Z_t merupakan himpunan variabel prediktor yang digunakan dalam pemodelan residual pada waktu t . Dengan demikian, model *Hybrid SARIMAX–RF* mengombinasikan kemampuan SARIMAX dalam menangkap pola linier dan musiman pada deret waktu dengan kemampuan *Random Forest* dalam memodelkan pola nonlinier yang masih terdapat pada residual. Penggabungan kedua pendekatan tersebut memungkinkan informasi yang belum dijelaskan oleh model SARIMAX untuk dimanfaatkan melalui pemodelan residual menggunakan Random Forest, sehingga diharapkan dapat meningkatkan kemampuan model dalam menghasilkan prediksi curah hujan yang lebih akurat (Alizadeh dan Farajzadeh, 2018; Gurjar dan Ghosh, 2025).

Selain menjelaskan struktur ketiga model tersebut, penting pula untuk meninjau bagaimana kinerja masing-masing model dinilai. Keakuratan hasil peramalan menjadi salah satu kriteria utama dalam evaluasi dan pemilihan model prediktif. Keakuratan biasanya diukur menggunakan metrik statistik, seperti:

Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (10)$$

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (11)$$

Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (12)$$

Tabel 1. Kriteria Evaluasi MAPE

MAPE	Keterangan
< 10%	Sangat Baik
10%-20%	Baik
20%-50%	Cukup
>50%	Buruk

Model dengan nilai RMSE, MAE, atau MAPE yang lebih rendah dianggap memiliki kinerja yang lebih baik. Dalam banyak penelitian, model hibrida seperti SARIMAX–RF menunjukkan hasil yang lebih akurat dibandingkan model tunggal, karena mampu menangkap baik pola linier maupun non-linier secara simultan (Shabbir dan Chaturvedi, 2022; Singh dan Choudhary, 2025).

3. METODE PENELITIAN

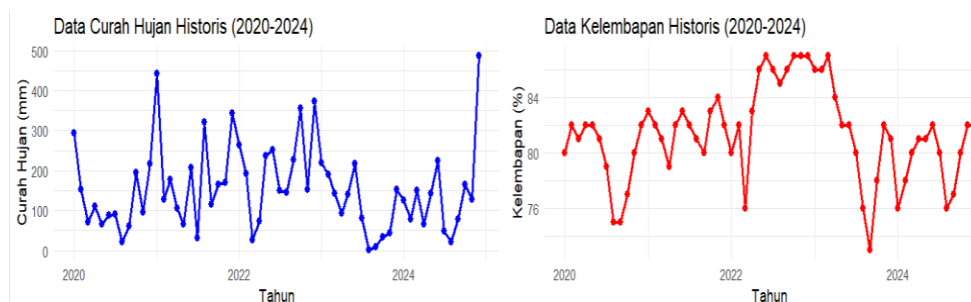
Penelitian ini menggunakan data sekunder curah hujan dan kelembapan relatif bulanan Kota Majene periode Januari 2020–Desember 2024 yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Data terdiri atas 60 observasi, dengan curah hujan sebagai variabel dependen dan kelembapan relatif sebagai variabel eksogen. Data dibagi secara kronologis menjadi data *training* sebesar 80% (Januari 2020–Desember 2023) dan data *testing* sebesar 20% (Januari–Desember 2024).

Analisis dilakukan melalui tahapan identifikasi karakteristik data, pengujian stasioneritas, pemodelan SARIMAX, pemodelan residual menggunakan *Random Forest* (RF), pembentukan *Hybrid SARIMAX–RF*, dan evaluasi kinerja. Pemilihan *hyperparameter* RF dilakukan melalui *GridSearchCV* dengan skema *time series cross-validation* (*TimeSeriesSplit*) pada data *training* (2020–2023) untuk menjaga urutan kronologis data dan mencegah kebocoran informasi dari periode mendatang. Karakteristik data dianalisis melalui statistik deskriptif dan visualisasi deret waktu untuk mengidentifikasi pola tren dan musiman. Uji *Augmented Dickey-Fuller* (ADF) digunakan untuk menguji stasioneritas dan menentukan kebutuhan *differencing*, sedangkan ACF dan PACF digunakan untuk mengidentifikasi orde model SARIMAX. Kandidat model kemudian dibandingkan berdasarkan *Akaike Information Criterion* (AIC) dan diagnostik residual.

Model SARIMAX terpilih digunakan untuk menghasilkan residual yang dihitung sebagai selisih antara nilai aktual dan prediksi. Residual tersebut menjadi target RF untuk memodelkan pola nonlinier yang belum ditangkap SARIMAX. Prediksi akhir *Hybrid SARIMAX–RF* diperoleh melalui penjumlahan prediksi SARIMAX dan prediksi residual RF. Kinerja ketiga model dievaluasi pada data *testing* menggunakan *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), dan *Mean Absolute Percentage Error* (MAPE), dengan nilai metrik yang lebih rendah menunjukkan kinerja prediksi yang lebih baik. Formulasi matematis yang digunakan meliputi model SARIMAX pada Persamaan (1), prediksi RF pada Persamaan (6), prediksi akhir *Hybrid* pada Persamaan (7), residual SARIMAX pada Persamaan (8), dan prediksi residual RF pada Persamaan (9). Selanjutnya, RMSE, MAE, dan MAPE masing-masing dinyatakan pada Persamaan (10), (11), dan (12).

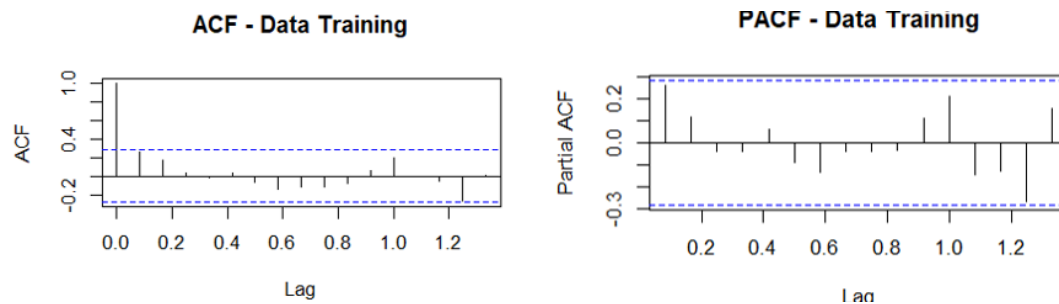
4. HASIL DAN PEMBAHASAN

Data penelitian terdiri atas curah hujan bulanan dan kelembapan relatif bulanan Kota Majene selama periode Januari 2020–Desember 2024, dengan total 60 observasi. Kedua variabel disusun dalam bentuk deret waktu bulanan (*monthly time series*) dan digunakan sebagai dasar analisis peramalan. Untuk mengetahui karakteristik dan pola temporal data, dilakukan eksplorasi awal melalui statistik deskriptif dan visualisasi deret waktu. Hasil visualisasi curah hujan dan kelembapan relatif ditampilkan pada Gambar 1.



Gambar 1. Plot deret waktu curah hujan dan kelembapan relatif bulanan Kota Majene

Gambar 1 menunjukkan pola deret waktu curah hujan dan kelembapan relatif bulanan di Kota Majene selama periode pengamatan. Curah hujan menunjukkan pola musiman tahunan dengan puncak umumnya terjadi pada Desember–Maret, sedangkan kelembapan relatif memiliki fluktuasi lebih rendah dan cenderung meningkat pada musim hujan. Pola tersebut menunjukkan karakteristik musiman dan hubungan temporal antara kedua variabel yang perlu dipertimbangkan dalam pemodelan.



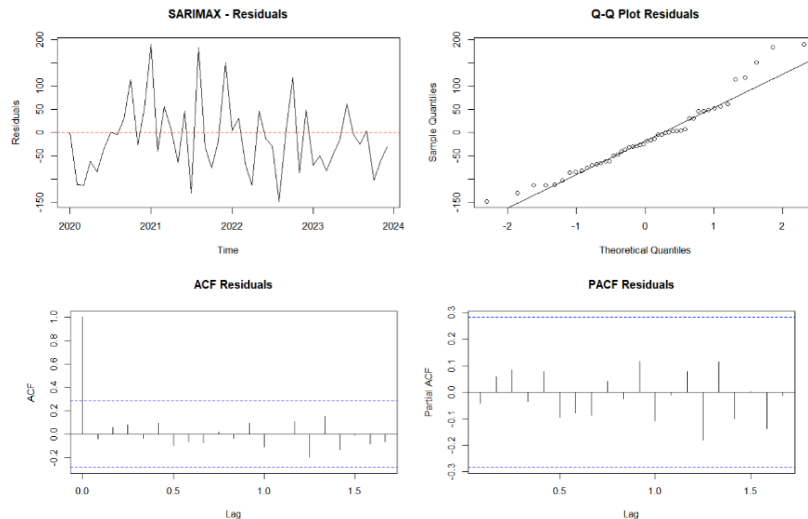
Gambar 2. Plot ACF dan PACF curah hujan bulanan Kota Majene

Identifikasi pola temporal selanjutnya dilakukan melalui plot ACF dan PACF data *training* pada Gambar 2. ACF menunjukkan autokorelasi signifikan pada lag 1 yang mengindikasikan ketergantungan jangka pendek, sedangkan autokorelasi pada lag 12 relatif lebih lemah tetapi tetap positif dan konsisten dengan pola musiman tahunan. PACF menunjukkan signifikansi pada beberapa lag awal sehingga mendukung penggunaan komponen nonmusiman AR/MA. Berdasarkan struktur tersebut, *auto.arima()* memilih $SARIMAX(0,1,1)(1,0,0)_{12}$ dengan AIC sebesar 555.60, lebih rendah daripada model tanpa komponen musiman dengan AIC sebesar 575.72. Hasil ini mengonfirmasi relevansi periode musiman 12 bulan dalam pemodelan.

Sebelum pembentukan model, uji ADF dilakukan untuk menguji stasioneritas data. Hasil pengujian menunjukkan bahwa curah hujan *training* belum stasioner dengan *p-value* sebesar $0.2975 > 0.05$. Differencing satu kali menghasilkan data stasioner dengan *p-value* sebesar $0.01 < 0.05$. Selanjutnya, penggunaan kelembapan relatif sebagai variabel eksogen menghasilkan model $SARIMAX(0,1,1)(1,0,0)_{12}$ dengan AIC sebesar 555.60, yaitu 15.82 poin lebih rendah dibandingkan model tanpa variabel eksogen yang memiliki AIC sebesar 571.42. Hasil tersebut menunjukkan bahwa kelembapan relatif meningkatkan kesesuaian model berdasarkan kriteria AIC.

Kelayakan model kemudian dievaluasi melalui diagnostik residual berdasarkan Persamaan (8). Gambar 3 menunjukkan residual yang berfluktuasi relatif acak terhadap waktu, Q-Q plot yang mendekati distribusi normal dengan sedikit deviasi, serta ACF dan PACF residual yang tidak menunjukkan autokorelasi signifikan. Hasil tersebut menunjukkan bahwa residual tidak mengandung pola sistematis yang berarti sehingga model SARIMAX layak digunakan sebagai dasar pemodelan selanjutnya.

Ketiadaan autokorelasi tersebut termasuk pada lag musiman (lag 12), yang mengindikasikan bahwa komponen musiman periode 12 bulan telah tertangkap sepenuhnya oleh struktur model $SARIMAX(0,1,1)(1,0,0)_{12}$, sehingga residual bersifat white noise tanpa pola musiman yang tersisa.



Gambar 3. Evaluasi Diagnostik Residual Model SARIMAX

Tahapan selanjutnya yaitu validasi kinerja prediktif model, dengan tujuan untuk mengukur akurasi SARIMAX pada data di luar sampel, menggunakan metrik evaluasi RMSE, MAE, dan MAPE (Persamaan 10–12) pada periode pengujian 12 bulan (Januari–Desember 2024), sebagaimana ditunjukkan pada Tabel 2 berikut:

Tabel 2. Hasil Evaluasi Kinerja Model SARIMAX

Metrik Evaluasi	Nilai SARIMAX
MAE	68.59 mm
MSE	12,280.54
RMSE	110.82 mm
MAPE	42.21 %

Model $SARIMAX(0,1,1)(1,0,0)_{12}$ dengan AIC sebesar 555.60. Berdasarkan Tabel 2, model tersebut menghasilkan MAE sebesar 68.59 mm, RMSE sebesar 110.82 mm, dan MAPE sebesar 42.21% pada data pengujian. Nilai MAPE tersebut menunjukkan bahwa akurasi model masih terbatas dalam memprediksi curah hujan. Selanjutnya, pengaruh variabel eksogen dianalisis melalui koefisien regresi dan uji signifikansi parameter. Kelembapan memiliki koefisien sebesar 16.42 dengan $p\text{-value} < 0.001$, yang menunjukkan hubungan positif dan signifikan terhadap curah hujan. Dengan demikian, peningkatan kelembapan sebesar 1% berkaitan dengan peningkatan curah hujan sekitar 16.42 mm dalam model, dengan asumsi variabel lain tetap. Temuan ini mendukung penggunaan kelembapan sebagai variabel eksogen dalam pemodelan curah hujan di wilayah dengan karakteristik iklim tropis (Debataraja, Wijayanti, dan Martha, 2021).

Sebagai pendekatan pembandingan, penelitian ini menggunakan *Random Forest* (RF) untuk memodelkan hubungan nonlinier dan interaksi kompleks yang dapat sulit ditangkap oleh model statistik. Data deret waktu ditransformasikan ke dalam format *supervised learning*, dengan curah hujan bulan ke- t sebagai target serta curah hujan lag 1–3 dan kelembapan bulan ke- t sebagai fitur input. Data tahun 2020–2023 digunakan sebagai data latih, sedangkan tahun 2024 digunakan sebagai data uji. Struktur fitur tersebut memungkinkan RF memanfaatkan informasi historis curah hujan dan kondisi kelembapan untuk memprediksi curah hujan bulan berikutnya.

Model RF dibangun menggunakan *RandomForestRegressor* dan dioptimalkan melalui *GridSearchCV*. Pencarian dilakukan pada $n\text{-estimators} = 100, 200, \text{ dan } 300$; $max_depth = 4, 6, \text{ dan } 8$; serta $min_samples_split = 2 \text{ dan } 4$. Validasi silang dalam proses *tuning* ini

menggunakan `TimeSeriesSplit` sebanyak k fold yang hanya memanfaatkan data *training* (2020–2023) secara berurutan, sehingga data *testing* (2024) tidak terlibat dalam proses seleksi parameter. Kombinasi terbaik diperoleh pada $n_estimators = 300$ dan $max_depth = 6$. Model hasil tuning kemudian digunakan untuk memprediksi curah hujan Januari–Desember 2024. Kinerja RF selanjutnya dievaluasi menggunakan MAE, RMSE, dan MAPE sebagaimana disajikan pada Tabel 3, sehingga hasilnya dapat dibandingkan secara langsung dengan model SARIMAX.

Tabel 3. Hasil Evaluasi Kinerja Model RF

Metrik Evaluasi	Nilai RF
MAE	46.68 mm
MSE	4309.13
RMSE	65.64 mm
MAPE	41.23 %

Random Forest menunjukkan kinerja sedikit lebih baik dibandingkan SARIMAX, dengan MAPE masing-masing sebesar 41.23% dan 42.21%, atau perbaikan sebesar 0.98 poin persentase. Perbedaan tersebut menunjukkan bahwa kedua model tunggal memiliki kemampuan prediksi yang relatif serupa sehingga diperlukan pendekatan yang dapat menggabungkan keunggulan keduanya.

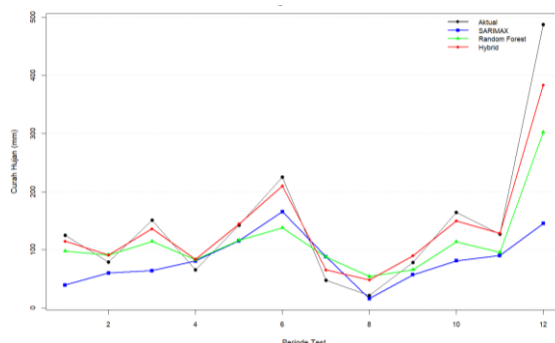
Model *Hybrid* SARIMAX–RF dikembangkan untuk mengatasi keterbatasan tersebut melalui dua tahap pemodelan berurutan. Tahap pertama menggunakan $SARIMAX(0,1,1)(1,0,0)_{12}$ dengan kelembapan sebagai variabel eksogen untuk menangkap komponen linier dan pola musiman 12 bulanan. Model ini menghasilkan prediksi dan residual yang merepresentasikan komponen yang belum dijelaskan oleh SARIMAX. Tahap kedua menggunakan *Random Forest* dengan $n_estimators = 300$ dan $max_depth = 6$ untuk memodelkan residual tersebut berdasarkan lag curah hujan dan kelembapan. Prediksi akhir *Hybrid* diperoleh dari penjumlahan prediksi SARIMAX dan prediksi residual RF. Dengan struktur tersebut, SARIMAX menangkap pola linier dan musiman, sedangkan RF mempelajari pola nonlinier pada residual. Kinerja model *Hybrid* selanjutnya dievaluasi menggunakan MAE, RMSE, dan MAPE pada data pengujian yang sama untuk mengukur peningkatan akurasi dibandingkan kedua model tunggal.

Tabel 4. Hasil Evaluasi Kinerja Model Hybrid

Metrik Evaluasi	Nilai <i>Hybrid</i> (SARIMAX+RF)
MAE	20.7 mm
MSE	1108.3
RMSE	33.29 mm
MAPE	23.31 %

Berdasarkan Tabel 4, dibandingkan dengan SARIMAX, model *Hybrid* menghasilkan penurunan RMSE sebesar 77.53 mm (70.0%) dan MAPE sebesar 18.90 poin persentase (44.8%). Dibandingkan dengan RF, model *Hybrid* menurunkan RMSE sebesar 32.35 mm (49.3%) dan MAPE sebesar 17.92 poin persentase (43.5%). Hasil tersebut menunjukkan bahwa model *Hybrid* SARIMAX–RF memiliki kinerja peramalan yang lebih baik dibandingkan kedua model tunggal. Temuan ini sejalan dengan penelitian Sari, Wibowo dan Hartono (2021) pada peramalan curah hujan di Malang yang menunjukkan bahwa *Hybrid* ARIMA–Neural Network memberikan performa lebih baik dibandingkan model tunggal dalam konteks peramalan hidrometeorologi di Indonesia.

Untuk memperkuat hasil tersebut, kinerja ketiga model selanjutnya dibandingkan secara keseluruhan melalui visualisasi prediksi terhadap nilai aktual dan tabel ringkasan metrik evaluasi. Perbandingan ini digunakan untuk memastikan konsistensi peringkat akurasi dan mengidentifikasi model yang paling sesuai untuk peramalan curah hujan Kota Majene.



Gambar 4. Perbandingan Prediksi vs Aktual

Gambar 4 menunjukkan perbandingan nilai prediksi dan data aktual curah hujan untuk setiap model. Terlihat bahwa model *Hybrid* SARIMAX–RF menghasilkan garis prediksi yang paling mendekati pola data aktual dibandingkan SARIMAX maupun Random Forest, sesuai dengan hasil evaluasi yang menempatkannya pada peringkat pertama dengan nilai error terendah (MAE, RMSE, dan MAPE).

Tabel 5. Perbandingan Kinerja Ketiga Model

Model	MAE (mm)	RMSE (mm)	MAPE (%)	Ranking
SARIMAX	68.59	110.82	42.21	3
Random Forest	46.68	65.64	41.23	2
<i>Hybrid</i> SARIMAX-RF	20.7	33.29	23.31	1

Hasil menunjukkan bahwa model *Hybrid* SARIMAX-RF memberikan kinerja terbaik pada semua metrik evaluasi, diikuti oleh Random Forest, dan SARIMAX menempati posisi ketiga. Tahapan selanjutnya yaitu penerapan model hybrid untuk peramalan masa depan, dengan tujuan untuk memberikan estimasi curah hujan yang dapat dimanfaatkan untuk perencanaan jangka pendek, menggunakan model *Hybrid* SARIMAX–RF terbaik yang telah divalidasi untuk memprediksi curah hujan periode Januari 2025 sebesar 142.7 mm, Februari 2025 sebesar 127.3 mm dan Maret 2025 sebesar 89.6 mm.

5. KESIMPULAN

Penelitian ini membandingkan kinerja model SARIMAX, *Random Forest* (RF), dan *Hybrid* SARIMAX–RF dalam memprediksi curah hujan bulanan di Kota Majene tahun 2025. Hasil evaluasi menunjukkan bahwa *Hybrid* SARIMAX–RF menghasilkan akurasi terbaik dengan MAPE sebesar 23.31%, lebih rendah 44.8% dibandingkan SARIMAX (42.21%) dan 43.5% dibandingkan RF (41.23%). Model *Hybrid* juga menurunkan RMSE sebesar 70% dibandingkan SARIMAX dan 49.3% dibandingkan RF. Keunggulan tersebut menunjukkan bahwa SARIMAX mampu menangkap pola musiman dan hubungan linier, sedangkan RF memodelkan pola nonlinier pada residual, sehingga kombinasi keduanya meningkatkan kemampuan model dalam menangkap fluktuasi curah hujan di Kota Majene.

Hasil peramalan menggunakan model *Hybrid* untuk Januari–Maret 2025 menghasilkan curah hujan masing-masing sebesar 142.7 mm, 127.3 mm, dan 89.6 mm, yang menunjukkan kecenderungan penurunan curah hujan. Informasi tersebut dapat mendukung pengambilan

keputusan pada sektor pertanian melalui penyesuaian jadwal tanam dan pemilihan varietas yang lebih toleran terhadap kondisi kering. Pada pengelolaan sumber daya air, hasil peramalan dapat membantu perencanaan alokasi cadangan air dan irigasi untuk mengantisipasi penurunan pasokan. Dalam mitigasi bencana, hasil prediksi dapat mendukung pemantauan potensi banjir dan kekeringan berdasarkan perubahan curah hujan. Dengan demikian, model *Hybrid* SARIMAX–RF tidak hanya memberikan kinerja prediktif yang lebih baik, tetapi juga memiliki potensi penerapan dalam mendukung pengelolaan pertanian, sumber daya air, dan mitigasi risiko hidrometeorologi di Kota Majene.

Saran: Penelitian selanjutnya dapat menguji metode *Hybrid* lain, seperti integrasi SARIMAX atau ARIMA dengan jaringan saraf dalam (LSTM), serta memasukkan variabel iklim tambahan seperti suhu permukaan laut, tekanan udara, dan kecepatan angin. Uji lintas wilayah juga direkomendasikan untuk menilai kemampuan generalisasi model pada daerah tropis lainnya.

DAFTAR PUSTAKA

- Alizadeh, F. & Farajzadeh, J. 2018. A *Hybrid* linear–nonlinear approach to predict the monthly rainfall over the Urmia Lake watershed using wavelet–SARIMAX–LSSVM conjugated model. *Journal of Hydroinformatics*, [online] 20(1), pp.246–261.
- Anwar, S. & Putri, R.A. 2022. Penerapan Model SARIMA untuk Peramalan Curah Hujan Bulanan di Kota Bengkulu. *AUTOMATA*, [online] 3(2), pp.45–56. Available at: <https://journal.uui.ac.id/AUTOMATA/article/view/24319>.
- Akbar, H., & Sanjaya, W. K. (2023). Kajian performa metode class weight random forest pada klasifikasi imbalance data kelas curah hujan. *Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi*, 3(1), 42–49. <https://doi.org/10.20885/snati.v3i1.30>
- Costa, G.E.M., Menezes Filho, F.C.M. & Canales, F.A. 2023. Assessment of time series models for mean discharge modeling and forecasting in a sub-basin of the Paranaíba River, Brazil. *Hydrology*, [online] 10(11), p.208. Available at: <https://www.mdpi.com/2306-5338/10/11/208>.
- Debataraja, N.N., Wijayanti, K. & Martha, S. 2021. Perbandingan Model ARIMAX dan Fungsi Transfer pada Peramalan Curah Hujan. *Jurnal Matematika*, [online] 20(1), pp.15–24. Available at: <https://onsearch.id/Record/IOS3952.356/Details>.
- Dotse, S.Q., Larbi, I., Limantol, A.M. & De Silva, L.C. 2024. A review of the application of *Hybrid* machine learning models to improve rainfall prediction. *Modeling Earth Systems and Environment*. [online] Available at: <https://link.springer.com/article/10.1007/s40808-023-01835-x>.
- Gurjar, A. & Ghosh, A. 2025. Forecasting inpatient admissions in district hospitals: a *Hybrid* model approach. *Journal of Intelligent Information Systems*. [online] Available at: <https://link.springer.com/article/10.1007/s10844-024-00895-4>.
- Hakiqi, R.F., Susanto, B. & Pratiwi, H. 2023. Peramalan Curah Hujan di Kota Bandung dengan Metode SARIMA (Seasonal Autoregressive Integrated Moving Average). *Inferensi*, [online] 6(2), pp.89–98. Available at: <https://iptek.its.ac.id/index.php/inferensi/article/view/19119>.
- Hartono, D. & Wijayanti, S. 2023. *Implementasi Machine Learning untuk Prediksi Curah Hujan di Kota Bandung*. Available at: <https://repository.telkomuniversity.ac.id/home/catalog/id/195466/slug/>.
- Indaryono, P. & Setiawan, A. 2024. Analisa Perbandingan Algoritma *Random Forest* dan Naïve Bayes untuk Klasifikasi Curah Hujan Berdasarkan Iklim di Indonesia. *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, [online] 9(1), pp.234–242.

- Available at:
<https://jurnal.stkipppgritulungagung.ac.id/index.php/jipi/article/view/4421> .
- Nugroho, B.A. & Sari, D.P. 2024. Penerapan Algoritma *Random Forest* untuk Memprediksi Curah Hujan pada Masa Mendatang di Daerah Berpotensi Banjir. *Building of Informatics, Technology and Science (BITS)*, [online] 6(2), pp.1823–1834. Available at: <https://ejurnal.seminar-id.com/index.php/bits/article/view/5593> .
- Oliveira, E.C.L., Carvalho, E.C., Jesus, E.S., dkk. 2025. A statistical and machine learning approach for monthly precipitation forecasting in an Amazon city. *Frontiers in Earth Science*. [online] Available at: https://www.academia.edu/download/123218467/feart_1_1589753.pdf .
- Pande, C.B., Singh, S.K. & Mulla, S. 2024. Time series forecasting of monthly rainfall using seasonal auto regressive integrated moving average with exogenous variables (SARIMAX) model. *Water Resources Management*, 38(6), pp.1825–1846. [online] Available at: <https://doi.org/10.1007/s11269-024-03756-5>
- Rahmayani, L. & Kurniawan, F. 2024. Penerapan Algoritma *Random Forest* dalam Prediksi Curah Hujan untuk Mendukung Analisis Cuaca. *Building of Informatics, Technology and Science (BITS)*, [online] 6(4), pp.5412–5423. Available at: <https://ejurnal.seminar-id.com/index.php/bits/article/view/6404> .
- Ruslan, A. & Pratomo, S.W. 2022. *Peramalan Curah Hujan dengan Pendekatan Seasonal Autoregressive Integrated Moving Average (SARIMA) di Kota Ambon*. Available at: <https://www.neliti.com/publications/277587/peramalan-curah-hujan-dengan-pendekatan-seasonal-autoregressive-integrated-movin> .
- Sari, K.M., Wibowo, A. & Hartono, R. 2021. Perbandingan Akurasi Peramalan Curah Hujan dengan menggunakan ARIMA, *Hybrid* ARIMA-NN, dan FFNN di Kabupaten Malang. In: *Seminar Nasional Official Statistics*. [online] pp.345–356. Available at: <https://prosiding.stis.ac.id/index.php/semnasoffstat/article/view/853> .
- Shabbir, H. & Chaturvedi, A. 2022. Day-ahead *Hybrid* forecasting of global horizontal irradiance using machine learning (*Random Forest* algorithm) and time-series model (SARIMAX). In: *IEEE Conference Proceedings*. [online] Available at: <https://ieeexplore.ieee.org/abstract/document/9758333/> .
- Shah, V., Patel, N., Shah, D., Swain, D., dkk. 2024. Forecasting maximum temperature trends with SARIMAX: A case study from Ahmedabad, India. *Sustainability*, [online] 16(16), p.7183. Available at: <https://www.mdpi.com/2071-1050/16/16/7183>.
- Singh, H. & Choudhary, M.P. 2025. Rainfall prediction in the context of climate change in Thar Desert India using machine learning algorithms. *Theoretical and Applied Climatology*. [online] Available at: <https://link.springer.com/article/10.1007/s00704-025-05592-y> .
- Siqueira, H. V., Toledo, J.F., Biuk, L.H., dkk. 2023. Climate indices impact in monthly streamflow series forecasting. *IEEE Access*. [online] Available at: <https://ieeexplore.ieee.org/iel7/6287639/6514899/10019598.pdf> .
- Widodo, P.S., Rahman, A. & Sari, I.P. 2022. Peramalan Curah Hujan di Kota Makassar dengan Menggunakan Metode SARIMAX. *Statistika*, [online] 10(2), pp.67–78. Available at: <https://journals.unisba.ac.id/index.php/statistika/article/view/990> .