

# PEMODELAN DAN PERAMALAN CURAH HUJAN DI BALIKPAPAN MENGUNAKAN METODE ARIMAX

## Studi Kasus Curah Hujan Kota Balikpapan Bulan Januari-Desember 2024

*(Rainfall Modeling and Forecasting in Balikpapan Using the ARIMAX Method: A Case Study of Balikpapan Rainfall from January to December 2024)*

**Rifky Septiansyah<sup>1</sup>, Hashifah Najma Zahra<sup>2</sup>, Ananda Reza Putra Rahmadan<sup>3</sup>,  
Sarah Katerina Simbolon<sup>4</sup>, Surya Puspita Sari<sup>5</sup>, Magdalena Effendi<sup>6</sup>**

Institut Teknologi Kalimantan<sup>1,2,3,4,5,6</sup>

E-mail: [16231005@student.itk.ac.id](mailto:16231005@student.itk.ac.id)

### ABSTRAK

Kota Balikpapan memiliki intensitas curah hujan yang tinggi dan fluktuatif, sehingga rentan terhadap bencana hidrometeorologi seperti banjir dan tanah longsor. Peramalan curah hujan yang akurat menjadi krusial untuk mitigasi risiko. Penelitian ini bertujuan untuk memodelkan dan meramalkan curah hujan harian di Balikpapan menggunakan metode *Autoregressive Integrated Moving Average with Exogenous Variables* (ARIMAX) dengan mempertimbangkan faktor lingkungan. Penelitian ini menggunakan data harian periode Januari-Desember 2024 yang bersumber dari NASA POWER, mencakup variabel curah hujan, suhu rata-rata, kelembaban udara, dan kecepatan angin. Data dianalisis menggunakan pendekatan Box-Jenkins, yang mencakup uji stasioneritas, estimasi parameter, dan uji diagnostik. Hasil penelitian menunjukkan bahwa data tidak stasioner dalam varians dan memerlukan transformasi logaritma. Model terbaik yang diperoleh berdasarkan nilai AIC terendah (582,05) adalah ARIMAX (1, 0, 1). Analisis variabel eksogen mengidentifikasi bahwa kelembaban udara dan kecepatan angin berpengaruh signifikan terhadap curah hujan, sedangkan suhu rata-rata tidak. Uji diagnostik Ljung-Box menunjukkan residual model bersifat *white noise* ( $p$ -value 0,2662). Evaluasi peramalan model menghasilkan nilai RMSE sebesar 9,9617. Model ini terbukti cukup baik dalam menangkap pola umum curah hujan, meskipun memiliki keterbatasan dalam memprediksi lonjakan ekstrem. Temuan ini dapat berkontribusi sebagai dasar ilmiah untuk mendukung sistem peringatan dini dan kebijakan mitigasi bencana di Balikpapan.

**Kata kunci:** Curah hujan, Peramalan, ARIMA, ARIMAX, Balikpapan

### ABSTRACT

*The city of Balikpapan experiences high and fluctuating rainfall intensity, rendering it vulnerable to hydrometeorological disasters such as floods and landslides. Accurate rainfall forecasting is crucial for risk mitigation. This study aims to model and forecast daily rainfall in Balikpapan using the Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) method, considering environmental factors. This study utilizes daily data from the January-December 2024 period sourced from NASA POWER, encompassing the variables of rainfall, average temperature, air humidity, and wind speed. The data was analyzed using the Box-Jenkins approach, which includes stationarity tests, parameter estimation, and diagnostic checks. The results indicated that the data was not stationary in variance, necessitating a logarithmic transformation. The best-fit model, identified by the lowest AIC value (582.05), was ARIMAX (1, 0, 1). Analysis of exogenous variables identified that Air Humidity and Wind Speed significantly influence rainfall, whereas Average Temperature does not. The Ljung-Box diagnostic test confirmed that the model's residuals behave as white noise ( $p$ -value 0.2662). The model's forecasting evaluation yielded an RMSE of 9.9617. The model proved reasonably effective in capturing the general rainfall patterns, despite limitations in predicting extreme spikes. These findings can contribute a scientific basis to support early warning systems and disaster mitigation policies in Balikpapan.*

**Keywords:** Rainfall, Forecasting, ARIMA, ARIMAX, Balikpapan

## PENDAHULUAN

Curah hujan merupakan salah satu elemen iklim yang memiliki peran penting dalam menjaga keseimbangan ekosistem dan keberlanjutan lingkungan hidup. Variasi curah hujan secara langsung mempengaruhi berbagai aspek lingkungan, seperti ketersediaan air tanah, produktivitas pertanian, kestabilan lahan, serta dinamika sistem hidrologi. Ketidakteraturan curah hujan dapat menyebabkan gangguan lingkungan seperti banjir, tanah longsor, hingga kekeringan, yang pada akhirnya berdampak pada ekosistem alami maupun aktivitas sosial-ekonomi masyarakat. Faktor-faktor yang mempengaruhi curah hujan meliputi kelembaban udara, suhu, kecepatan angin, aktivitas Madden Julian Oscillation (MJO), massa udara basah, serta fenomena iklim global seperti El Nino Southern Oscillation (ENSO) (Nisa, Chairun et al., 2021). Oleh karena itu, pemahaman mendalam terhadap pola curah hujan menjadi langkah penting dalam pengelolaan lingkungan yang berkelanjutan dan mitigasi risiko bencana hidrometeorologi.

Kota Balikpapan, yang terletak di pesisir Kalimantan Timur, memiliki intensitas curah hujan yang relatif tinggi dibandingkan wilayah lain di sekitarnya. Berdasarkan data BMKG, rata-rata curah hujan tahunan di Balikpapan mencapai lebih dari 2.500 mm/tahun dengan puncak musim hujan berlangsung antara bulan Desember hingga April. Pola curah hujan yang fluktuatif ini menyebabkan Balikpapan rentan terhadap bencana hidrometeorologi seperti banjir dan tanah longsor pada musim hujan, serta potensi kekeringan ketika curah hujan menurun drastis. Kondisi tersebut berdampak pada kualitas lingkungan, seperti peningkatan erosi tanah, pencemaran air, dan gangguan terhadap keseimbangan ekosistem pesisir. Oleh sebab itu, peramalan curah hujan yang akurat menjadi sangat penting sebagai dasar dalam perencanaan lingkungan, pengelolaan sumber daya air, dan mitigasi bencana di wilayah ini.

Dalam konteks ilmiah, peramalan curah hujan dilakukan dengan pendekatan *time series* yang memanfaatkan data historis untuk memperkirakan nilai di masa depan. Pendekatan ini menjadi salah satu upaya penting dalam manajemen lingkungan berbasis data, khususnya untuk memprediksi dinamika iklim lokal. Model *Autoregressive Integrated Moving Average* (ARIMA) merupakan salah satu metode statistik yang banyak digunakan dalam peramalan deret waktu karena kemampuannya dalam memodelkan pola data historis secara sistematis melalui pendekatan Box-Jenkins. Namun, model ARIMA hanya memperhitungkan satu variabel utama tanpa mempertimbangkan faktor lingkungan lain yang dapat mempengaruhi curah hujan.

Untuk mengatasi keterbatasan tersebut, dikembangkan model *Autoregressive Integrated Moving Average with Exogenous Variables* (ARIMAX) yang memungkinkan dimasukkannya variabel eksogen atau faktor luar yang relevan. Model ini mampu memberikan hasil peramalan yang lebih akurat karena memperhitungkan pengaruh faktor lingkungan seperti suhu udara, kelembapan, tekanan udara, dan kecepatan angin terhadap perubahan curah hujan (Wangdi et al., 2010). Dengan demikian, ARIMAX menjadi pendekatan yang lebih representatif dalam menganalisis sistem iklim dan lingkungan yang kompleks.

Beberapa penelitian sebelumnya telah menerapkan metode ARIMAX dalam konteks peramalan curah hujan dengan spesifikasi variabel eksogen yang beragam. Penelitian oleh R. Amelia et al. (2021) di Kota Pangkalpinang membuktikan bahwa penambahan variabel kecepatan angin maksimum sebagai variabel eksogen mampu meningkatkan kinerja model dibandingkan ARIMA standar, dengan model terbaik yang dihasilkan adalah ARIMAX(0,1,3) untuk curah hujan bulanan. Sementara itu, Chairun Nisa et al. (2021) memperoleh model terbaik dengan nilai *Akaike Information Criterion* (AIC) sebesar 1334,629 di Bali dengan memanfaatkan variabel eksogen iklim global, yaitu Indeks Nino 3.4. Namun, kajian serupa di wilayah Kalimantan, khususnya Kota Balikpapan, masih jarang dilakukan. Berbeda dengan penelitian terdahulu yang melibatkan variabel iklim global atau fokus pada kecepatan angin maksimum saja, penelitian ini bertujuan untuk menerapkan model ARIMAX dengan fokus pada pengaruh *contemporaneous* (seketika) dari kombinasi variabel lingkungan lokal (suhu, kelembaban, dan angin) terhadap curah hujan di Kota Balikpapan. Keputusan untuk tidak menerapkan lag pada variabel eksogen didasarkan pada asumsi bahwa respons atmosfer lokal terhadap pembentukan hujan terjadi secara langsung pada periode waktu yang sama, serta untuk menjaga parsimoni model agar tidak terlalu kompleks. Diharapkan hasil penelitian ini dapat memberikan kontribusi terhadap pengelolaan lingkungan berbasis data,

mendukung kebijakan adaptasi terhadap perubahan iklim, serta membantu mitigasi risiko bencana di kawasan perkotaan Balikpapan.

**METODE**

Penelitian ini dilakukan di Kota Balikpapan, Provinsi Kalimantan Timur, Indonesia, dengan menggunakan data sekunder harian yang bersumber dari *Prediction of Worldwide Energy Resources* (NASA POWER) periode Januari hingga Desember 2024. Penelitian ini menerapkan pendekatan kuantitatif dengan metode analisis deret waktu untuk memodelkan hubungan antara curah hujan dan faktor-faktor eksogen, seperti suhu udara, kelembaban, dan kecepatan angin. Data dibagi secara kronologis dengan proporsi 80% sebagai data training (Januari–November 2024) dan 20% sebagai data testing (Desember 2024), karena proporsi ini memberikan keseimbangan antara kecukupan data untuk estimasi parameter dan keandalan evaluasi kinerja model pada data di luar sampel (*out-of-sample forecasting*) (Hyndman & Athanasopoulos, 2021).

**ARIMA**

Model *Autoregressive Integrated Moving Average* (ARIMA) merupakan metode peramalan deret waktu yang dikembangkan oleh Box dan Jenkins (Box et al., 2015). Model ini secara umum dinotasikan sebagai ARIMA (p, d, q), di mana parameter p mewakili orde *autoregressive* (ketergantungan pada nilai masa lalu), d adalah orde *differencing* (jumlah proses pembedaan untuk menstabilkan rata-rata), dan q adalah orde *moving average* (ketergantungan pada *error* masa lalu). Struktur matematis model ARIMA dinyatakan dalam persamaan berikut:

$$\phi(B)(1 - B)^d y_t = \theta(B)\varepsilon_t \dots\dots\dots (1)$$

Dimana:

- $Y_t$  adalah Nilai variabel pada waktu ke-t
- $B$  adalah operator mundur (*backshift operator*), di mana  $BY_t = Y_{t-1}$
- $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$  adalah polynomial AR (*backshift operator B*)
- $(1 - B)^d$  adalah *operator differencing orde d*,
- $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$  adalah polinomial MA,
- $\varepsilon_t$  adalah *white noise (error term)* dengan rata-rata nol dan varians konstan

Penerapan model ini didasarkan pada asumsi bahwa data deret waktu harus stasioner (rata-rata dan varians konstan). Selain itu, sisaan (residual) model diasumsikan bersifat *white noise*, yaitu berdistribusi normal, independen (tidak ada autokorelasi), dan homoskedastisitas.

Pemilihan model ARIMA yang tepat dilakukan dengan mengikuti prosedur iteratif Box-Jenkins (Box et al., 2015). Pendekatan ini menyediakan kerangka kerja sistematis untuk membangun model deret waktu dengan memastikan bahwa model yang dipilih mampu menangkap pola data secara memadai sebelum digunakan untuk peramalan. Tahapan-tahapan dalam prosedur Box-Jenkins meliputi:

1. Identifikasi Model (*Model Identification*): Tahap awal untuk menentukan nilai parameter d (melalui uji stasioneritas seperti ADF), serta menetapkan orde p dan q sementara berdasarkan pola *cut-off* atau *tail-off* pada plot ACF dan PACF dari data yang telah stasioner.
2. Estimasi Parameter (*Parameter Estimation*): Mengestimasi nilai koefisien model ( $\phi$  dan  $\theta$ ) menggunakan metode *Maximum Likelihood* (ML) atau *Least Squares* (LS) untuk meminimalkan jumlah kuadrat galat.
3. Pemeriksaan Diagnostik (*Diagnostic Checking*): Menganalisis sisaan (residual) model untuk memastikan asumsi *white noise* terpenuhi (tidak ada autokorelasi signifikan berdasarkan uji Ljung-Box dan berdistribusi normal). Jika asumsi tidak terpenuhi, model perlu diidentifikasi ulang.
4. Peramalan (*Forecasting*): Menggunakan model terbaik yang telah lolos uji diagnostik untuk memprediksi nilai deret waktu di masa depan.

**ARIMAX**

Model *Autoregressive Integrated Moving Average with Exogenous Variables* (ARIMAX) merupakan pengembangan dari model ARIMA yang mengintegrasikan satu atau lebih variabel independen eksternal (X) ke dalam pemodelan deret waktu. Penambahan variabel eksogen bertujuan untuk meningkatkan akurasi peramalan dengan menangkap variasi yang dijelaskan oleh faktor-faktor luar yang relevan (Box et al., 2015). Secara matematis, bentuk umum model ARIMAX dinyatakan dalam persamaan berikut:

$$\phi(B)(1 - B)^d y_t = \theta(B)\varepsilon_t + \beta_0 + \sum_{i=1}^k \sum_{j=0}^{m_i} \beta_{i,j} x_{i,t-j} \dots\dots\dots (2)$$

dimana:

- $X_{i,t-j}$  = variabel eksogen ke-i dengan lag j,
- $\beta_{i,j}$  = koefisien regresi untuk variabel xi pada lag j,
- $Y_t$  = variabel dependen (deret waktu utama) pada waktu ke-t
- $\phi_p(B)$  = Komponen *Autoregressive* (AR) orde p
- $(1 - B)^d$  = Komponen *differencing* orde d untuk menstabilkan data
- $\theta_q(B)$  = Komponen *Moving Average* (MA) orde q
- $\beta_0$  = Konstanta atau *intercept* model
- $k$  = Jumlah total variabel eksogen yang digunakan
- $m_i$  = Orde lag maksimum untuk variabel eksogen ke-i
- $\varepsilon_t$  = Sisaan (*error term*) pada waktu ke-t

Dengan menyertakan variabel eksternal, ARIMAX mampu menangkap dampak variabel lingkungan terhadap sasaran (curah hujan). Secara teoritis, pemilihan variabel eksogen dalam penelitian ini didasarkan pada mekanisme fisik atmosfer. Kelembaban udara memiliki kaitan erat dengan curah hujan karena tingginya kadar uap air di atmosfer merupakan syarat utama terjadinya proses kondensasi yang menghasilkan presipitasi. Sementara itu, suhu udara mempengaruhi laju evaporasi dan konveksi termal yang memicu pembentukan awan hujan. Kecepatan angin juga memegang peranan krusial dalam mentransportasikan massa udara basah dan mempengaruhi pembentukan awan konvektif. Pendekatan ini didukung oleh Wangdi et al. (2010), yang menyatakan bahwa integrasi faktor-faktor meteorologi lokal ke dalam model peramalan dapat menangkap variabilitas curah hujan yang tidak dapat dijelaskan hanya oleh data historis curah hujan itu sendiri.

**JENIS DAN SUMBER DATA**

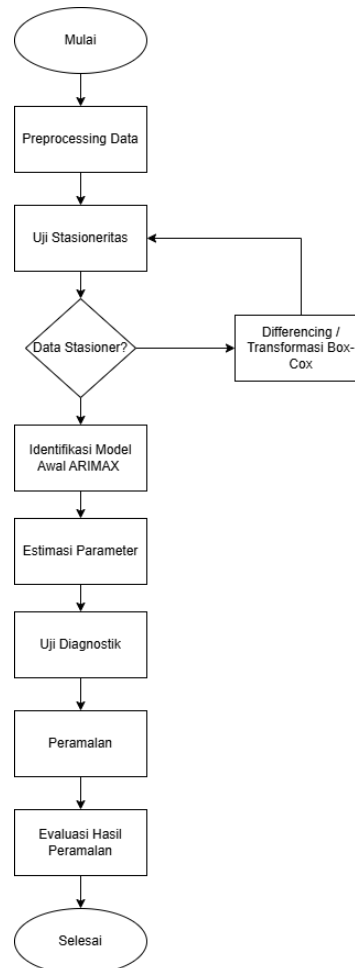
Data yang digunakan dalam penelitian ini merupakan data sekunder dengan resolusi waktu harian yang bersumber dari *Prediction of Worldwide Energy Resources* (NASA POWER). Penggunaan data harian dipilih untuk menangkap volatilitas dan fluktuasi curah hujan jangka pendek yang menjadi karakteristik iklim di Kota Balikpapan. Dalam pemodelan ARIMAX ini, variabel eksogen (suhu, kelembaban, dan kecepatan angin) diposisikan pada waktu ke-t (*current time*), atau tanpa penerapan lag. Keputusan ini didasarkan pada alasan bahwa dalam skala mikro-meteorologi, respons atmosfer lokal terhadap pembentukan presipitasi (hujan) diasumsikan terjadi secara langsung (*contemporaneous*) pada hari yang sama. Variabel yang digunakan dirinci pada Tabel 1 berikut:

**Tabel 1.** Jenis dan Variabel Data.

No.	Variabel	Jenis	Keterangan	Sumber
1.	Curah hujan (mm/hari)	Y	Jumlah curah hujan harian	NASA POWER
2.	Suhu udara rata-rata (°C)	X1	Rata-rata suhu udara harian	NASA POWER
3.	Kelembaban udara rata-rata (%)	X2	Tingkat kelembaban harian	NASA POWER
5.	Kecepatan angin rata-rata (m/s)	X3	Laju angin rata-rata harian	NASA POWER

## TAHAP ANALISIS DATA

Tahapan analisis data dalam penelitian ini dilakukan secara sistematis untuk memperoleh model peramalan curah hujan yang optimal menggunakan metode ARIMAX (*Autoregressive Integrated Moving Average with Exogenous Variables*). Adapun langkah-langkahnya dijelaskan dalam gambar 1 berikut ini.



**Gambar 1.** Diagram Alir Tahapan Analisis Data

### ***Pengolahan Awal Data (Preprocessing)***

Tahap awal dilakukan dengan memeriksa data untuk mendeteksi adanya nilai hilang (*missing values*) dan pencilan (*outliers*). Penanganan nilai hilang dilakukan menggunakan metode interpolasi linier. Metode ini dipilih karena keunggulannya dalam mengestimasi nilai yang hilang dengan mempertahankan tren linier antara titik data sebelum dan sesudahnya, sehingga kontinuitas deret waktu tetap terjaga tanpa mengurangi variabilitas data secara drastis (seperti yang terjadi pada metode imputasi rata-rata). Selanjutnya, seluruh variabel dipastikan tetap dalam format harian guna menyesuaikan dengan tujuan utama penelitian ini, yaitu memodelkan dan meramalkan dinamika fluktuasi curah hujan harian di Kota Balikpapan secara presisi.

### ***Uji Stasioneritas Data***

Uji stasioneritas dilakukan untuk memastikan data deret waktu memiliki rata-rata dan variansi yang konstan sehingga model menghasilkan estimasi yang stabil dan valid (Box et al., 2015). Stasioneritas pada rata-rata diuji menggunakan Augmented Dickey-Fuller (ADF Test), di mana hipotesis nol menyatakan bahwa data mengandung *unit root* atau tidak stasioner (Dickey & Fuller, 1979). Sementara itu, stasioneritas variansi diuji dan diperbaiki menggunakan Transformasi Box-

Cox untuk menstabilkan variansi data (Box & Cox, 1964). Apabila data belum stasioner, maka dilakukan differencing hingga memenuhi kriteria stasioneritas (Wei, 2006).

### **Identifikasi Model Awal ARIMAX ( $p, d, q$ )**

Tahap identifikasi model bertujuan untuk menentukan nilai awal parameter  $p$ ,  $d$ , dan  $q$  pada model ARIMAX. Penentuan orde  $p$  dan  $q$  dilakukan melalui analisis pola ACF dan PACF dari data variabel Curah Hujan yang telah dibuat stasioner, bukan dari residual model (Box et al., 2015). Pola ACF yang terpotong pada lag ke- $q$  mengindikasikan model MA( $q$ ), sedangkan pola PACF yang terpotong pada lag ke- $p$  mengindikasikan model AR( $p$ ); sementara parameter  $d$  ditentukan dari jumlah *differencing* yang diperlukan (Wei, 2006). Model awal selanjutnya dievaluasi melalui analisis residual, di mana residual yang bersifat acak (*white noise*) menunjukkan bahwa model telah menangkap struktur data dengan baik (Gujarati & Porter, 2009).

### **Estimasi Parameter dan Perbandingan Model ARIMAX**

Setelah model awal ARIMAX teridentifikasi, dilakukan estimasi parameter menggunakan metode *Maximum Likelihood* (ML), yaitu dengan mencari nilai parameter yang memaksimalkan fungsi likelihood berdasarkan asumsi bahwa residual model berdistribusi normal dengan rata-rata nol dan variansi konstan (Box et al., 2015). Pada pendekatan ini, fungsi likelihood dibangun dari distribusi probabilitas residual, kemudian dioptimalkan secara numerik untuk memperoleh estimasi parameter *autoregressive*, *moving average*, serta koefisien variabel eksogen yang paling mungkin menghasilkan data observasi (Hamilton, 1994). Metode ML dipilih karena bersifat efisien dan menghasilkan estimator yang konsisten serta asimtotik normal pada model deret waktu (Wei, 2006). Selanjutnya, beberapa kandidat model dibandingkan menggunakan *Akaike Information Criterion* (AIC) dan *Bayesian Information Criterion* (BIC), yang mempertimbangkan keseimbangan antara kecocokan model dan kompleksitas parameter, di mana model dengan nilai AIC dan BIC terendah dipilih sebagai model terbaik (Akaike, 1974; Schwarz, 1978).

### **Uji Diagnostik**

Uji diagnostik dilakukan untuk memastikan bahwa residual model memenuhi asumsi statistik. Ljung–Box Test digunakan untuk menguji ada tidaknya autokorelasi residual secara simultan hingga lag tertentu, dengan hipotesis nol bahwa residual bersifat *white noise* (Ljung & Box, 1978). Sementara itu, Jarque–Bera Test digunakan untuk menguji normalitas residual berdasarkan nilai skewness dan kurtosis, dengan hipotesis nol bahwa residual berdistribusi normal (Jarque & Bera, 1987). Model dinyatakan layak apabila hipotesis nol pada kedua pengujian tidak ditolak, yang menunjukkan bahwa residual tidak berkorelasi dan berdistribusi normal (Box et al., 2015).

Hipotesis:

- Ljung–Box:  $H_0$  residual tidak berautokorelasi;  $H_1$  residual berautokorelasi.
- Jarque–Bera:  $H_0$  residual berdistribusi normal;  $H_1$  residual tidak normal.

### **Peramalan**

Model ARIMAX yang telah memenuhi kriteria diagnostik kemudian digunakan untuk melakukan peramalan terhadap variabel dependen (misalnya curah hujan) dengan mempertimbangkan pengaruh variabel independen seperti suhu rata-rata, kelembaban udara, dan kecepatan angin. Hasil peramalan selanjutnya dibandingkan dengan data aktual untuk melihat tingkat kecocokan dan reliabilitas model.

### **Evaluasi Hasil Peramalan**

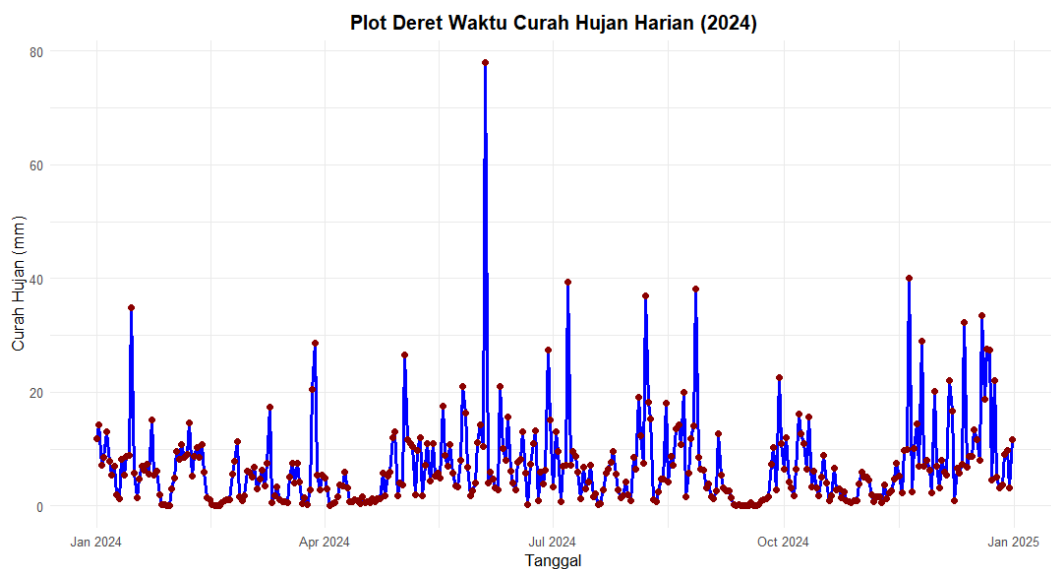
Kinerja model peramalan dievaluasi menggunakan *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), dan *Mean Absolute Percentage Error* (MAPE). MAE mengukur rata-rata kesalahan absolut dan memberikan bobot yang sama pada setiap kesalahan sehingga mudah

diinterpretasikan dan relatif tidak sensitif terhadap nilai ekstrim, sedangkan RMSE menghitung akar dari rata-rata kuadrat kesalahan sehingga memberikan penalti yang lebih besar terhadap kesalahan berskala besar dan lebih sensitif terhadap *outliers* (Hyndman & Athanasopoulos, 2021). Sementara itu, MAPE mengukur kesalahan dalam bentuk persentase terhadap nilai aktual sehingga bersifat bebas skala dan memudahkan perbandingan antar model, namun memiliki keterbatasan ketika nilai aktual mendekati nol (Makridakis et al., 2018). Oleh karena itu, penggunaan ketiga ukuran kesalahan ini secara bersamaan memberikan gambaran yang lebih komprehensif mengenai akurasi dan karakteristik kesalahan model peramalan.

## HASIL DAN PEMBAHASAN

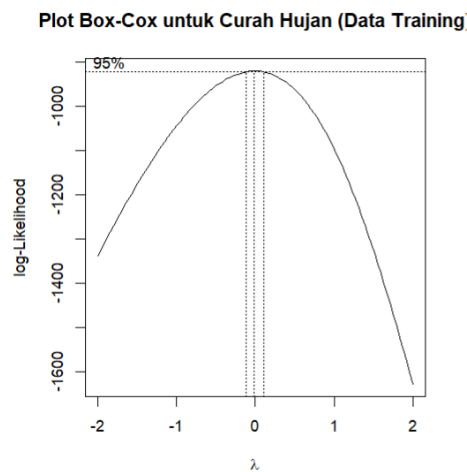
### Pengujian Stasioneritas Data

Stasioneritas data merupakan salah satu asumsi penting yang harus dipenuhi dalam analisis deret waktu parametrik. Pengujian stasioneritas dapat dilakukan melalui beberapa pendekatan, antara lain dengan mengamati plot deret waktu, melakukan uji Augmented Dickey–Fuller (ADF), serta menganalisis plot *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF) untuk menilai kestasioneran rata-rata data. Selain itu, pemeriksaan stasioneritas varians dapat dilakukan menggunakan plot transformasi Box–Cox. Gambar 2 menampilkan plot deret waktu variabel curah hujan harian di Kota Balikpapan selama periode 1 Januari hingga 31 Desember 2024.



**Gambar 2.** Plot Deret Waktu Curah Hujan Balikpapan

Berdasarkan gambar 2, dapat disimpulkan bahwa data tersebut tidak stasioner. Sebuah data deret waktu dianggap stasioner jika sifat statistiknya, terutama rata-rata dan varians, konstan dari waktu ke waktu. Meskipun data ini tampaknya stasioner dalam rata-rata (karena tidak menunjukkan tren kenaikan atau penurunan yang jelas), plot ini secara jelas melanggar asumsi stasioneritas dalam varians (heteroskedastisitas). Terlihat adanya periode-periode dengan fluktuasi yang tenang di mana data berkumpul di dekat nol (misalnya sekitar September–Oktober 2024), yang menunjukkan varians yang sangat kecil. Untuk memastikan keakuratan uji stasioneritas maka dilakukan uji Box-Cox, serta uji ADF agar hasil yang didapatkan lebih dapat dipercaya. Gambar 3 adalah plot uji Box-Cox yang dilakukan.



**Gambar 3.** Plot Box-Cox

Dari gambar 3 dapat diketahui bahwa  $\lambda = -0,02$  maka dapat disimpulkan bahwa data curah hujan tidak stasioner dalam varians. Karena lamda tidak sama dengan 1 maka perlu dilakukan transformasi agar lamda nya menjadi sama dengan 1. Setelah dilakukan transformasi dengan menggunakan transformasi logaritma  $+1$  lamda yang dihasilkan sebesar 0,788 mendekati 1 sehingga dapat disimpulkan bahwa varians data curah hujan telah menjadi relatif stabil. Nilai  $\lambda$  yang mendekati 1 menunjukkan bahwa transformasi yang dilakukan sudah cukup efektif dalam menstabilkan varians data, meskipun tidak harus tepat bernilai 1 secara absolut (Montgomery et al., 2016). Dengan terpenuhinya asumsi stasioneritas varians, analisis selanjutnya dapat dilanjutkan dengan pengujian stasioneritas rata-rata menggunakan uji Augmented Dickey-Fuller (ADF).

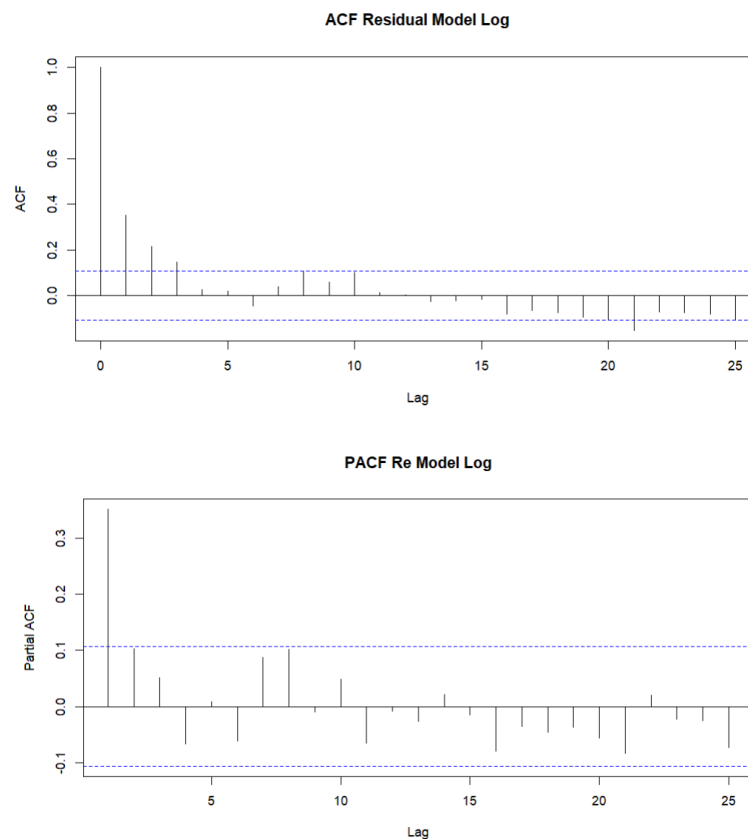
**Tabel 2.** Uji ADF.

Variable	P-Value	Keputusan
Curah Hujan	0.01	Signifikan
Suhu Rata-Rata	0.03	Signifikan
Kelembaban Udara	0.01	Signifikan
Kecepatan Angin	0.01	Signifikan

Berdasarkan tabel 2 diatas, dapat dilihat bahwa semua  $p$ -value pada uji adf kurang 0,05 artinya tolak  $H_0$  sehingga dapat disimpulkan bahwa semua variabel yang diuji stasioner terhadap rata rata. Dengan demikian, curah hujan, suhu rata rata, kelembaban udara, dan kecepatan angin di Kota Balikpapan pada tahun 2024 menunjukkan rata rata yang stabil pada periode tahun 2024.

### Permodelan ARIMAX

Setelah memastikan bahwa data yang digunakan telah memenuhi asumsi stasioneritas selanjutnya dilakukan pemodelan ARIMAX. Pemodelan dilakukan dengan membentuk model regresi berganda dengan data curah hujan sebagai variabel Y, dan suhu rata rata, kelembaban udara, dan kecepatan angin sebagai variabel x. Pembentukan model ARIMAX dapat dilakukan dengan melihat plot ACF dan PACF dari residual model regresi yang telah dibentuk. Gambar 4 merupakan plot ACF dan PACF yang dihasilkan dengan menggunakan data training dari residual model regresi.



**Gambar 4.** Plot ACF dan PACF Residual Model

Pada plot ACF diatas menjelaskan model AR atau orde q, berdasarkan hasil plot acf diatas dapat dilihat bahwa terjadi *cut off* setelah lag 3 maka orde q yang memungkinkan untuk digunakan adalah 1,2,3. Selanjutnya pada plot PACF digunakan untuk menjelaskan model MA atau orde p, dari hasil plot pacf diatas dapat disimpulkan bahwa terjadi *cut off* setelah lag 1, maka orde p yang memungkinkan untuk digunakan adalah 1. Maka model ARIMAX sementara yang dapat digunakan antara lain ARIMAX (1, 0 ,1), ARIMAX (1, 0, 2), dan ARIMAX (1, 0, 3).

### Estimasi Parameter dan Perbandingan Model ARIMAX

Tahapan selanjutnya adalah melakukan estimasi parameter untuk semua model yang masuk kedalam kandidat untuk dilakukan pengujian dan membandingkan model mana yang terbaik berdasarkan nilai AIC dan signifikansinya. Tabel dibawah ini merupakan hasil dari pengujian signifikansi yang dilakukan.

**Tabel 3.** Estimasi Parameter Model ARIMAX.

Model	Parameter	P-value	Keputusan	AIC
ARIMAX (1, 0, 1)	AR(1)	0.000	Signifikan	582.05
	MA(1)	0.009	Signifikan	
ARIMAX (1, 0, 2)	AR(1)	0.000	Signifikan	584.01
	MA(1)	0.059	Tidak Signifikan	
	MA(2)	0.825	Tidak Signifikan	

ARIMAX (1, 0, 3)	AR(1)	0.106	Tidak Signifikan	584.57
	MA(1)	0.953	Tidak Signifikan	
	MA(2)	0.265	Tidak Signifikan	
	MA(3)	0.144	Tidak Signifikan	

Berdasarkan hasil estimasi parameter pada Tabel 3, diketahui bahwa tidak semua model ARIMAX yang diuji memiliki parameter yang signifikan secara statistik. Model ARIMAX (1,0,1) menunjukkan bahwa seluruh parameter utamanya, yaitu AR(1) dan MA(1), signifikan pada tingkat signifikansi 5%. Sementara itu, pada model ARIMAX (1,0,2) dan ARIMAX (1,0,3), sebagian besar parameter *Moving Average* tidak signifikan, sehingga kedua model tersebut kurang layak digunakan untuk peramalan. Selain signifikansi parameter, pemilihan model terbaik juga mempertimbangkan nilai *Akaike Information Criterion* (AIC), di mana model dengan nilai AIC terkecil dianggap paling optimal. Berdasarkan kriteria tersebut, model ARIMAX (1,0,1) memiliki nilai AIC paling kecil, yaitu sebesar 582,05, sehingga dipilih sebagai model terbaik untuk peramalan curah hujan. Tahap selanjutnya adalah melakukan pengujian signifikansi parameter dengan menyertakan variabel eksogen, yaitu suhu rata-rata, kelembapan udara, dan kecepatan angin, untuk mengetahui pengaruh faktor-faktor tersebut terhadap curah hujan.

**Tabel 4.** Uji Signifikan Variabel Eksogen.

Model	Parameter	P-value	Keputusan
ARIMAX (1, 0, 1)	AR(1)	0.000	Signifikan
	MA(1)	0.009	Signifikan
	Suhu Rata-Rata	0.153	Tidak Signifikan
	Kelembapan Udara	0.000	Signifikan
	Kecepatan Angin	0.000	Signifikan

Berdasarkan hasil uji signifikansi parameter dengan variabel eksogen pada model ARIMAX (1, 0, 1) variabel Kelembapan Udara, dan Kecepatan Angin signifikan terhadap curah hujan di kota Balikpapan sedangkan pada variabel Suhu Udara Rata-Rata tidak signifikan. Dengan demikian karena variabel suhu udara rata rata tidak signifikan maka pada saat melakukan peramalan dengan model ARIMAX tidak perlu menyertakan variabel suhu udara. Selanjutnya dilakukan uji diagnostik untuk melakukan pengecekan autokorelasi menggunakan Ljung Box dan normalitas residual dengan menggunakan uji Jarque-Bera pada residual data untuk model terbaik yaitu ARIMAX (1, 0, 1).

**Tabel 5.** Hasil Uji Diagnostik.

Model	White Noise	Hasil	Distribusi Normal	Hasil
ARIMAX (1, 0, 1)	0.2662	Tidak ada Autokorelasi	0.0000	Tidak berdistribusi normal

Berdasarkan hasil uji diagnostik yang disajikan pada Tabel 5, model ARIMAX(1,0,1) menunjukkan nilai *p-value* uji Ljung-Box sebesar 0,2662, yang lebih besar dari taraf signifikansi 0,05. Hal ini menunjukkan bahwa tidak terdapat autokorelasi yang signifikan pada sisaan model, sehingga dapat disimpulkan bahwa sisaan bersifat *white noise* atau acak. Kondisi ini menandakan bahwa

model sudah mampu menangkap seluruh pola sistematis yang ada dalam data, dan tidak ada informasi penting yang tertinggal di dalam sisaan.

Sementara itu, hasil uji Jarque–Bera menunjukkan nilai *p-value* sebesar 0,0000 (<0,05) yang mengindikasikan bahwa sisaan model tidak berdistribusi normal. Namun, ketidaknormalan sisaan tidak serta-merta menggugurkan kelayakan model ARIMAX, terutama ketika tujuan utama analisis adalah peramalan. Dalam pemodelan ARIMA/ARIMAX, asumsi yang paling penting adalah bahwa sisaan bersifat *white noise*, yaitu tidak mengandung autokorelasi dan memiliki varians yang konstan, karena hal tersebut menunjukkan bahwa model telah mampu menangkap struktur utama data (Hyndman & Athanasopoulos, 2021). Normalitas sisaan lebih berperan dalam inferensi statistik dan pembentukan interval prediksi, tetapi bukan syarat mutlak untuk menghasilkan peramalan titik yang valid (Box et al., 2015). Oleh karena itu, model ARIMAX tetap layak digunakan selama sisaan bersifat acak dan tidak berkorelasi.

### Peramalan ARIMAX (1, 0 ,1)

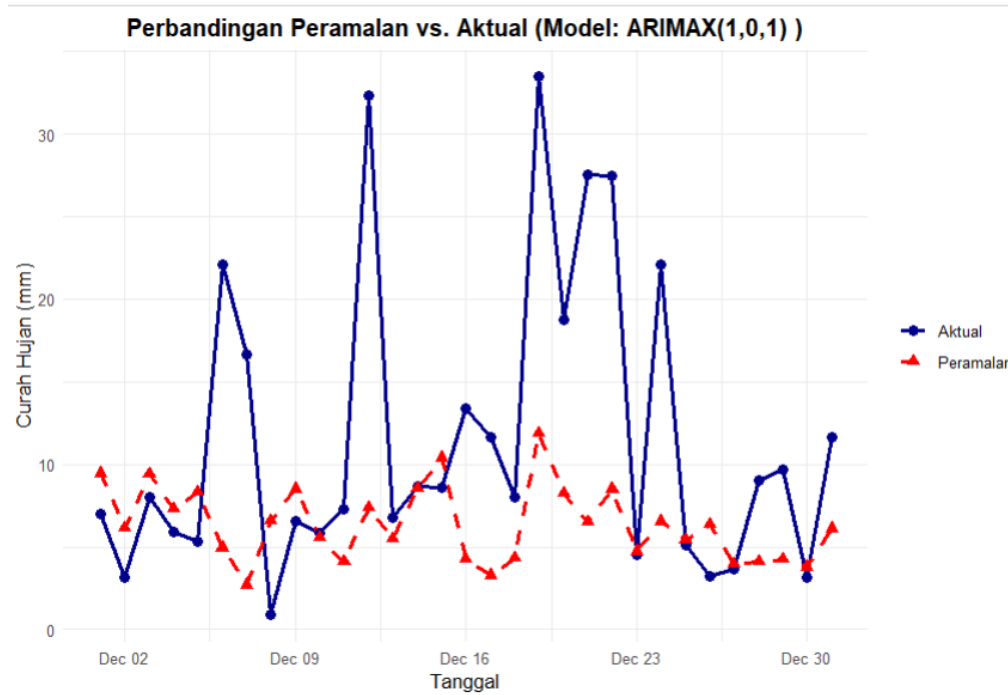
Pada tahap ini akan dilakukan peramalan dengan menggunakan data testing yang telah dibagi yaitu data curah hujan bulan desember 2024 sebanyak 31 titik. Peramalan dilakukan dengan menggunakan model ARIMAX (1, 0, 1). Berikut dibawah ini adalah hasil peramalan dengan data aktual dan data hasil peramalan.

**Tabel 6.** Hasil Peramalan dan Evaluasi.

Tanggal	Aktual	Peramalan	RMSE
1 Desember 2024	7.02	9.43	
2 Desember 2024	3.15	6.15	
3 Desember 2024	7.99	9.42	
4 Desember 2024	5.89	7.32	
5 Desember 2024	5.34	8.32	
...			9.9617
28 Desember 2024	9.07	4.12	
29 Desember 2024	9.68	4.26	
30 Desember 2024	3.13	3.79	
31 Desember 2024	11.66	6.09	

Nilai *Root Mean Square Error* (RMSE) sebesar 9,9617 menunjukkan tingkat kesalahan rata-rata antara nilai curah hujan aktual dan hasil peramalan yang dihasilkan oleh model ARIMAX(1,0,1). RMSE mengukur seberapa besar penyimpangan atau deviasi antara hasil prediksi dan data sebenarnya, dengan satuan yang sama seperti variabel yang diamati — dalam hal ini milimeter (mm) curah hujan. Semakin kecil nilai RMSE, maka semakin baik kemampuan model dalam menghasilkan prediksi yang mendekati data aktual.

Nilai RMSE sebesar 9,9617 berarti bahwa secara rata-rata, hasil peramalan memiliki selisih sekitar  $\pm 9,96$  mm dari nilai curah hujan sebenarnya. Meskipun terdapat beberapa selisih yang cukup besar pada titik-titik tertentu (terutama saat terjadi lonjakan curah hujan), nilai ini masih menunjukkan bahwa model memiliki tingkat akurasi yang moderat, terutama dalam menggambarkan pola umum perubahan curah hujan. Untuk meningkatkan ketepatan peramalan, dapat dilakukan optimasi parameter model atau penambahan variabel eksogen lain yang relevan dengan kondisi curah hujan.



**Gambar 5.** Plot Perbandingan Peramalan vs Aktual Curah Hujan di Balikpapan

Gambar 5 menunjukkan hasil perbandingan antara nilai aktual dan nilai peramalan curah hujan menggunakan model ARIMAX(1,0,1). Berdasarkan grafik tersebut, terlihat bahwa pola pergerakan data hasil peramalan (garis merah putus-putus) mengikuti arah tren data aktual (garis biru), meskipun dengan tingkat fluktuasi yang lebih rendah. Hal ini menunjukkan bahwa model ARIMAX(1,0,1) mampu menangkap pola umum perubahan curah hujan, terutama pada periode-periode meningkat atau menurunnya curah hujan. Namun demikian, terlihat bahwa nilai peramalan cenderung lebih halus dan tidak sepenuhnya mampu menangkap lonjakan ekstrem yang terjadi pada data aktual, seperti pada puncak curah hujan di pertengahan Desember. Secara keseluruhan, model ini masih dapat dikatakan cukup baik dalam memberikan estimasi arah dan kecenderungan curah hujan, meskipun terdapat perbedaan tingkat variasi antara data aktual dan hasil peramalan.

## IMPLIKASI KEBIJAKAN

Temuan penelitian ini bahwa model ARIMAX (1, 0, 1) dengan variabel eksogen Kelembaban Udara dan Kecepatan Angin adalah model terbaik untuk peramalan curah hujan di Balikpapan memiliki implikasi kebijakan yang spesifik dan dapat ditindaklanjuti. Hasil ini mengkonfirmasi bahwa penambahan variabel eksogen yang relevan sangat penting untuk meningkatkan akurasi peramalan. Pentingnya variabel iklim ini sejalan dengan penelitian lain di Indonesia; misalnya, studi di Makassar (Fitriani, 2014) juga berhasil mengidentifikasi kelembaban udara sebagai variabel input (eksogen) yang signifikan untuk memodelkan curah hujan. Temuan ini juga mendukung penelitian di kota terdekat, Samarinda, yang menyimpulkan bahwa model ARIMA univariat (tanpa variabel eksogen) memiliki keterbatasan dan merekomendasikan penelitian selanjutnya untuk mempertimbangkan

variabel lain seperti kelembaban udara menggunakan metode ARIMAX atau fungsi transfer (Syawal et al., 2022).

Secara terukur, Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) dan Badan Penanggulangan Bencana Daerah (BPBD) Kota Balikpapan diidentifikasi sebagai pemangku kepentingan utama yang dapat menerapkan temuan ini. Direkomendasikan agar kedua instansi ini berkolaborasi untuk menguji dan mengintegrasikan model ARIMAX (1, 0, 1) spesifik ini ke dalam sistem peringatan dini bencana hidrometeorologi. Dengan memasukkan data kelembaban dan kecepatan angin lokal yang telah dikonfirmasi sebagai faktor yang mempengaruhi curah hujan peramalan operasional jangka pendek untuk banjir dan tanah longsor dapat menjadi lebih akurat dan relevan bagi wilayah Balikpapan.

Meskipun model ini menunjukkan akurasi yang moderat (RMSE 9,9617) dan mampu menangkap pola umum, temuannya juga mengidentifikasi keterbatasan dalam memprediksi lonjakan ekstrem. Oleh karena itu, implikasi kebijakan jangka panjang bagi Pemerintah Kota Balikpapan adalah perlunya pengembangan model berkelanjutan. Mengingat data curah hujan di Indonesia cenderung membentuk pola musiman, penelitian di masa depan harus mengeksplorasi model SARIMAX (*Seasonal* ARIMAX), yang secara eksplisit memasukkan komponen musiman. Selain itu, untuk meningkatkan ketepatan, dapat dipertimbangkan penggunaan metode *ensemble averaging* menggabungkan beberapa model ARIMAX yang terbukti efektif dalam memprediksi curah hujan di Sintang (Rahayu et al., 2022).

## KESIMPULAN

Penelitian ini bertujuan untuk memodelkan dan meramalkan curah hujan harian di Balikpapan selama tahun 2024 menggunakan metode ARIMAX dengan mempertimbangkan variabel lingkungan. Hasil penelitian menunjukkan bahwa setelah data distasionerkan dalam varians melalui transformasi logaritma, model terbaik yang diperoleh adalah ARIMAX (1, 0, 1). Model ini memiliki nilai AIC terendah yaitu sebesar 582,05 dengan parameter AR(1) dan MA(1) yang signifikan. Analisis variabel eksogen mengidentifikasi bahwa Kelembaban udara dan kecepatan angin berpengaruh signifikan terhadap curah hujan, sedangkan suhu rata-rata tidak berpengaruh signifikan. Uji diagnostik menunjukkan residual model telah memenuhi asumsi *white noise* atau tidak ada autokorelasi ( $p$ -value Ljung-Box > 0.05), meskipun tidak terdistribusi normal, yang dianggap tidak menghalangi kelayakan model untuk peramalan. Evaluasi peramalan pada data pengujian menghasilkan nilai RMSE sebesar 9,9617, yang menunjukkan bahwa model cukup baik dalam menangkap pola umum pergerakan data aktual, meskipun belum sepenuhnya mampu memprediksi lonjakan curah hujan ekstrem.

## DAFTAR PUSTAKA

- Akaike, H. (1974). *A new look at the statistical model identification*. IEEE Transactions on Automatic Control, 19(6), 716–723.
- Amelia, R., Dalimunthe, D. Y., Kustiawan, E., & Sulistiana, I. (2021). ARIMAX model for rainfall forecasting in Pangkalpinang, Indonesia. *IOP Conference Series: Earth and Environmental Science*, 926(1), 012034. <https://doi.org/10.1088/1755-1315/926/1/012034>
- Amri, M., Farid, A., & Mulyana, T. (2023). *Perbandingan ARIMA dan ARIMAX untuk peramalan curah hujan di Indonesia*. *Square: Journal of Mathematics and Applied Research*, 5(1), 45–56. <https://journal.walisongo.ac.id/index.php/square/article/download/17059/5820>
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). John Wiley & Sons.
- Dickey, D. A., & Fuller, W. A. (1979). *Distribution of the estimators for autoregressive time series with a unit root*. Journal of the American Statistical Association, 74(366), 427–431.

- Fitriani. (2014). *Peramalan Curah Hujan di Kota Makassar Menggunakan Model ARIMAX*. Skripsi, Universitas Islam Negeri (UIN) Alauddin Makassar.
- Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics* (5th ed.). McGraw-Hill.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts.
- Islam, F., Rahman, S. M., & Alam, S. (2022). *A hybrid GEP-ARIMAX model for rainfall prediction*. *Water Resources Management*, 36(5), 1783–1798. <https://doi.org/10.1007/s11269-022-03219-9>
- Ljung, G. M., & Box, G. E. P. (1978). *On a measure of lack of fit in time series models*. *Biometrika*, 65(2), 297–303.
- Jarque, C. M., & Bera, A. K. (1987). *A test for normality of observations and regression residuals*. *International Statistical Review*, 55(2), 163–172.
- Mahmudi, M. (2024). *Modeling Rainfall in Jakarta with Hybrid ARIMAX-ANN Model*. *Proceedings of Indonesian Meteorological Research*, 12(2), 44–51.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). *Statistical and Machine Learning forecasting methods: Concerns and ways forward*. *PLOS ONE*, 13(3), e0194889.
- Misshuari, S. (2023). *Application of ARIMA Method for Rainfall Forecasting in Asahan Region*. *Journal of Environmental and Applied Science Research*, 7(2), 22–30.
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2016). *Introduction To Time Series Analysis and Forecasting* (2nd ed.). Hoboken, NJ: John Wiley & Sons.
- Nisa, C., Sumarjaya, I. W., & Srinadi, I. G. A. M. (2021). *Penggunaan model ARIMAX untuk meramalkan data curah hujan bulanan di Bali*. *E-Jurnal Matematika*, 10(4), 186–191. <https://doi.org/10.24843/MTK.2021.v10.i04.p341>.
- Pandit, P., Kumar, S., & Mehta, R. (2023). *Hybrid ARIMAX-LSTM model for nonlinear rainfall forecasting*. *Scientific Reports*, 13(1), 8424. <https://doi.org/10.1038/s41598-023-35456-0>
- Permata, R. P. (2024). *Daily Rainfall Forecasting with ARIMA Exogenous Variables and Support Vector Regression*. *Journal of Applied Meteorological Science*, 15(3), 122–130.
- Rahayu, S., Martha, S., & Rizki, S. W. (2022). *Prediksi Curah Hujan dengan Metode Ensemble Averaging*. *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)*, 11(4), 633-640.
- Schwarz, G. (1978). *Estimating the dimension of a model*. *Annals of Statistics*, 6(2), 461–464.
- Sholihah, M. (2024). *Penerapan model ARIMAX dalam peramalan curah hujan di Kecamatan Batukliang Utara*. Skripsi, Universitas Mataram. <https://eprints.unram.ac.id/46979>
- Syawal, A. F., Wahyuningsih, S., & Siringoringo, M. (2022). *Peramalan Curah Hujan di Kota Samarinda Menggunakan Autoregressive Integrated Moving Average (ARIMA)*. *Jurnal EKSPONENSIAL*, 13(2), 153-159.
- Wangdi, K., Singhasivanon, P., Silawan, T., Lawpoolsri, S., White, N. J., & Kaewkungwal, J. (2010). *Development of Temporal Modelling for Forecasting and Prediction of Malaria Infections Using Time-Series and ARIMAX Analyses: A Case Study in Endemic District of Bhutan*. *Malaria Journal*.
- Wei, W. W. S. (2006). *Time Series Analysis: Univariate and Multivariate Methods* (2nd ed.). Pearson.