Jurnal EKSPONENSIAL Volume 17, Nomor 1, April 2026
p-ISSN 2085-7829
e-ISSN 2798-3455

Perbandingan Metode Imputasi Data Hilang untuk Peramalan Curah Hujan dengan ARIMA di Palangka Raya

Comparative Study of Missing Data Imputation Methods for ARIMA Rainfall Forecasting in Palangka Raya

Khusnia Nurul Khikmah[1a)]
[1] Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Palangka Raya, Indonesia

[a)] *Corresponding author: khusnia.nurulkhikmah@mipa.upr.ac.id*

ABSTRACT
Rainfall is one of the fundamental factors influencing various aspects of a region, including the city of Palangka Raya. Therefore, statistical tools are needed to forecast rainfall for mitigation and planning related to its effects. The objective of this study is to evaluate the performance of the ARIMA model in forecasting rainfall and to analyze the impact of three methods of missing data imputation on forecast accuracy. Therefore, this study proposes an evaluation of the autoregressive integrated moving average (ARIMA) model's performance for rainfall forecasting and the handling of missing data through three imputation methods: mean imputation, forward filling, and backward filling. The data used in this study consists of historical rainfall data from 2013 to 2022, obtained from the official website of the Central Statistics Agency of Palangka Raya City. The forecasting results indicate that rainfall data modelled using ARIMA and treated for missing data via mean imputation demonstrated the best forecasting accuracy in the validation test, as rainfall data in Palangka Raya City is fluctuating without a consistent seasonal pattern. Thus, the historical average value is more representative than the preceding or subsequent values, which could reinforce local bias. This is demonstrated by the smallest RMSE, MAE, and MAPE values, which are 124,17%, 103,57%, and 40,32%, respectively.

Keywords: *arima, rainfall, missing data imputation, palangka raya, forecasting*

1. Pendahuluan

Peramalan curah hujan merupakan salah satu aspek fundamental dalam perencanaan, pengelolaan, dan mitigasi terhadap kondisi suatu wilayah. Namun, dalam fakta lapangannya data curah hujan sering mengandung data hilang yang merupakan akibat dari faktor lingkungan, kesalahan transmisi, maupun gangguan alat pencatat (AlSalehy & Bailey, 2025; Zhang & Thorburn, 2022). Pada dataset curah hujan Kota Palangka Raya periode 2013–2022 yang diperoleh dari BPS, ditemukan sebanyak 0,83% data hilang (*missing*) dari total 120 pengamatan bulanan. Akibat dari data tidak lengkap ini dapat mengurangi akurasi peramalan suatu model, seperti *autoregressive integrated moving average* (ARIMA). Model ARIMA secara umum memerlukan data deret waktu kontinu untuk dapat menghasilkan permalan yang optimal (Khikmah et al., 2023; Ospina et al., 2023). Oleh karena itu, pemilihan metode imputasi data hilang yang tepat menjadi sangat penting dalam proses analisis sebelum pemodelan data.

Penelitian ini mengajukan studi komparasi tiga metode penanganan data hilang dengan *mean imputation*, *forward filling*, dan *backward filling*. Metode *mean imputation* ini menggantikan data hilang dengan nilai rata-rata data yang ada, dimana menurut penelitian sebelumnya pada (Khikmah, 2024) menghasilkan akurasi peramalan yang baik dan mudah dalam aplikasinya. Metode *forward filling* dimana mengisi data hilang dengan nilai sebelumnya, dimana menurut penelitian sebelumnya pada (Appaia & Palraj, 2023) cocok digunakan untuk data dengan tren stabil dan mempertahankan *outlier*. Metode terakhir dengan *backward filling* dimana mengisi data hilang dengan nilai selanjutnya, yang menurut penelitian sebelumnya pada (Enders, 2022; Wang et al., 2023) baik digunakan ketika data hilang tersebut relevan terhadap nilai selanjutnya. Ketiga metode penanganan data hilang ini diajukan pada penelitian ini dengan pertimbangan performa dan efektivitas metode yang bergantung pada data sehingga diperlukan penelitian yang komprehensif untuk membandingkan kinerja penanganan data hilang tersebut yang dimodelkan dengan ARIMA.

Kebaruan penelitian ini adalah belum adanya studi komparasi sistematis ketiga metode imputasi data hilang (*mean imputation, forward filling*, dan *backward filling*) pada data curah hujan di wilayah tropis basah seperti Kalimantan Tengah. Selain itu, penelitian ini juga mengintegrasikan penanganan data hilang dengan pemodelan ARIMA untuk lokasi spesifik Kota Palangka Raya yang menggunakan data historis riil selama sepuluh tahun (2013–2022) dengan pola missing alami, sehingga memberikan kontribusi praktis bagi pengambil kebijakan di bidang mitigasi bencana dan perencanaan tata guna air.

Kota Palangka Raya dipilih pada penelitian ini, dengan pertimbangan kondisi wilayahnya. Dimana masuk dalam kategori iklim tropis basah yang memiliki curah hujan yang rentan terhadap ketidaklengkapan

p-ISSN 2085-7829
e-ISSN 2798-3455

data akibat intensitas curah hujan dan infrastruktur. Penelitian sebelumnya pada (Risnayah, 2023) menunjukkan bahwa curah hujan memiliki deviasi ramalan yang signifikan pada kesalahan imputasi data, sehingga penelitian ini diajukan secara komprehensif bertujuan untuk membandingkan metode imputasi penanganan data hilang dalam meramalkan curah hujan untuk spesifik kota Palangka Raya.

2. Metodologi

A. Penanganan Data Hilang

a. *Mean Imputation*

Mean imputation atau pengisian dengan nilai rata-rata ini adalah salah satu metode penanganan data hilang untuk data deret waktu dimana nilai yang hilang diisi dengan nilai rata-rata dari data yang tersedia (Khikmah, 2024). Secara umum metode ini mudah diimplementasikan, namun memiliki kelemahan terhadap variabilitas data. Jika x_t adalah nilai curah hujan bulanan yang tercatat dan n adalah jumlah total pengamatan yang tersedia, maka data hilang $\hat{x}$ dapat dihitung melalui persamaan berikut.

$$\hat{x} = \frac{1}{n}\sum_{t=1}^{n} x_t \tag{1}$$

b. *Forward Filling*

Forward filling atau *last observation carried forward* (LOCF) adalah salah satu metode penanganan data hilang untuk data deret waktu dimana nilai yang hilang diisi dengan nilai terakhir yang diamati sebelum titik yang hilang. Metode ini sering digunakan dalam deret waktu dengan tujuan mempertahankan tren sebelumnya, namun akan dapat menghasilkan bias jika data berfluktuasi tinggi. Menurut penelitian sebelumnya pada (Appaia & Palraj, 2023) menunjukkan hasil bahwa metode *forward filling* ini cocok untuk data dengan kategori *missing completely at random* atau data hilang acak yang tidak bergantung pada nilai data itu sendiri.

c. *Backward Filling*

Backward filling atau *next observation carried backward* (NOCB) adalah salah satu metode penanganan data hilang untuk deret waktu dimana nilai yang hilang diisi dengan nilai berikutnya yang tersedia. Pertimbangan yang dilakukan adalah ketika nilai terbaru relevan dibanding nilai sebelumnya. Secara umum metode ini dapat mengabaikan perubahan struktural dalam data sehingga menyebabkan distorsi jika data memiliki pola tak-stasioner (Enders, 2022).

B. Autoregressive Integrated Moving Average

Autoregressive integrated moving average (ARIMA) adalah model statistika yang digunakan untuk peramalan data deret waktu non-stasioner dan berpola temporal (Tandon et al., 2022). Model ARIMA terdiri dari tiga komponen utama, yaitu *autoregressive* (AR) yang diidentifikasi sebagai orde p, *differencing* atau pembeda (I) sebagai orde d, dan *moving average* (MA) sebagai orde q (Alabdulrazzaq et al., 2021). Secara umum, model ARIMA dinotasikan sebagai $ARIMA(p,d,q)$. Metode ini bertujuan untuk memodelkan ketergantungan nilai saat ini berdasarkan nilai masa lalunya, sehingga keakuratan ramalannya didasarkan penuh pada identifikasi orde dan validasi asumsinya (Kobiela et al., 2022). Jika, koefisien komponen AR dengan orde p disimbolkan sebagai ϕ_p, koefisien komponen MA dengan orde q disimbolkan sebagai θ_q, B adalah operator *backshift*, e_t adalah residual ke-t, dan Y_t adalah data deret waktu ke-t, maka secara matematis model $ARIMA(p,d,q)$ dapat ditulis sebagai berikut (Khikmah et al., 2023).

$$\phi_p(1-B)^d Y_t = \mu + \theta_q(B)e_t \tag{2}$$

Pendugaan parameter dari model $ARIMA(p,d,q)$ dapat dilakukan dengan metode *maximum likelihood estimation* (MLE) dengan memaksimalkan fungsi kemungkinan (*likelihood function*) dari data terhadap parameter model atau melalui *conditional least squares* (CLS) dengan meminimalkan jumlah kuadrat residual (*sum of squared errors*). Jika n adalah jumlah observasi, σ^2 adalah varians residual, dan e_t adalah residual yang dihitung berdasarkan model, maka fungsi *log-likelihood* untuk model ARIMA dengan asumsi residual berdistribusi normal dinyatakan sebagai berikut.

$$\log\ L(\theta,\phi,\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{n} e_t^2 \tag{3}$$

Nilai dari parameter ϕ dan θ di atas dapat diduga dengan memaksimalkan fungsi *log-likelihood* atau secara ekuivalen meminimalkan jumlah kuadrat residual. Secara matematis, pendekatan CLS sendiri dirumuskan sebagai berikut.

p-ISSN 2085-7829
e-ISSN 2798-3455

$$\hat{\phi}, \hat{\theta} = \arg \min_{\phi,\theta} \sum_{t=1}^{n} e_t^2 \tag{4}$$

C. Ukuran Keakuratan Ramalan

Evaluasi performa peramalan data deret waktu yang diajukan dalam penelitian ini *root mean square error* (RMSE), *mean absolute error* (MAE), dan *mean absolute percentage error* (MAPE). Matrik evaluasi pertama adalah RMSE, dimana matrik ini mengukur rata-rata kuadrat selisih antara nilai prediksi dengan nilai aktualnya. Secara umum jika y_i adalah nilai actual, $\hat{y}_i$ adalah nilai prediksi, dan memiliki jumlah observasi sebesar n maka secara matematis nilai RMSEnya sbagai berikut.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

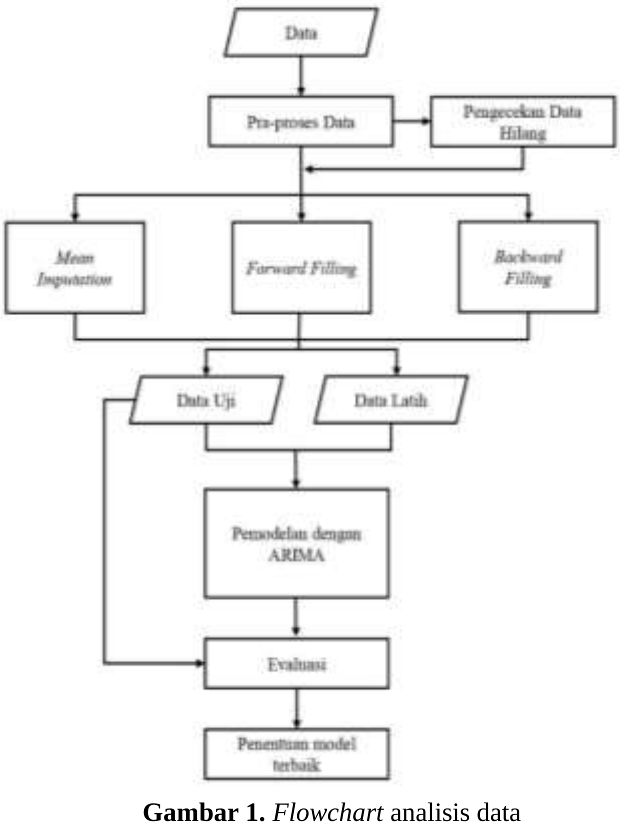
Matrik evaluasi kedua adalah MAE, matrik ini berdasarkan penelitian sebelumnya pada (Hodson, 2022) lebih *robust* terhadap *outlier* dimana nilainya dihitung berdasarkan rata-rata absolut selisih nilai prediksi dan aktualnya. Jika y_i adalah nilai aktual, $\hat{y}_i$ adalah nilai prediksi, dan memiliki jumlah observasi sebesar n maka nilai MAE secara matematis adalah sebagai berikut.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{6}$$

Matrik evaluasi terakhir adalah MAPE, matriks ini mengukur persentase kesalahan relatif terhadap nilai aktualnya. Jika y_i adalah nilai actual, $\hat{y}_i$ adalah nilai prediksi, dan memiliki jumlah observasi sebesar n maka, nilai MAPE secara matematis dirumuskan sebagai berikut (Khikmah et al., 2025).

$$MAPE = 100\% \times \frac{1}{n}\sum_{i=1}^{n}\left|\frac{(y_i - \hat{y}_i)}{y_i}\right| \tag{7}$$

D. Langkah Analisis



Gambar 1. *Flowchart* analisis data

p-ISSN 2085-7829
e-ISSN 2798-3455

Tahapan analisis yang dilakukan oleh penelitian ini adalah dengan pra-pemrosesan data dan pengecekan data hilang, dimana jika terdapat data yang hilang akan dilakukan penanganan data hilang dengan metode *mean imputation* atau pengisian data hilang dengan nilai rata-rata, *forward filling* atau pengisian dengan nilai sebelumnya, dan *backward filling* atau pengisian dengan nilai setelahnya. Data yang telah dilakukan penanganan data hilang selanjutnya dilakukan pembagian menjadi data latih sebesar 80% dari data dan data uji sebesar 20% sisanya. Selanjutnya dilakukan pemodelan ARIMA dengan menggunakan data latih yang setelahnya dievaluasi untuk menentukan performa model terbaik dengan data uji. Terakhir, adalah penentuan model terbaik berdasarkan matriks evaluasi yang diajukan.

E. Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diambil dari website resmi Badan Pusat Statistik (BPS) Kota Palangka Raya terkait curah hujan di Kota Palangka Raya dari Januari 2013 hingga Desember 2022. Dimana statistik deskriptif data ini tersaji pada Tabel 1.

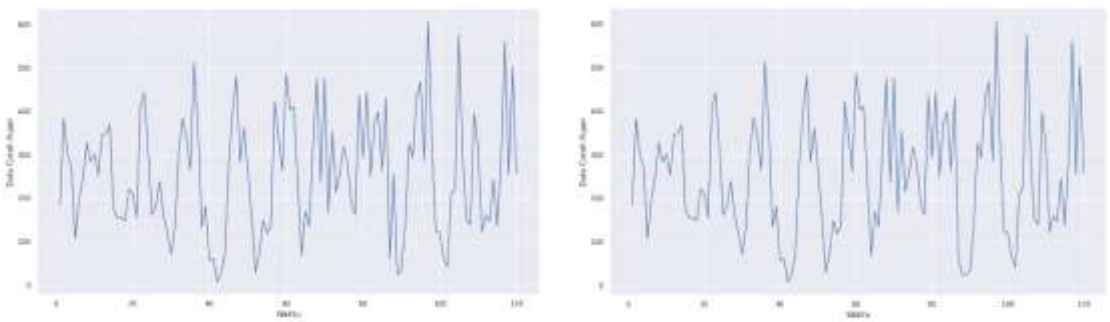
Tabel 1. Statistik Deskriptif Curah Hujan

Ukuran	**Nilai (mm)**
Min.	7,5
Q_1	148,8
Median	253,4
Rata-rata	254,8
Q_3	347,0
Maks.	604,7
NA’s	1

Berdasarkan Tabel 1, data curah hujan di Kota Palangka Raya periode Januari 2013 hingga Desember 2022 memiliki nilai rata-rata sebesar 254,8 mm dengan rentang data dari minimum 7,5 mm hingga maksimum 604,7 mm, serta nilai median (253,4 mm) yang mendekati rata-rata menunjukkan sebaran data cenderung simetris. Selain itu, hanya ditemukan 1 data hilang (NA's=1) dari total 120 pengamatan bulanan, sehingga persentase data hilang sangat kecil atau sebesar 0,83%.

3. Hasil dan Pembahasan

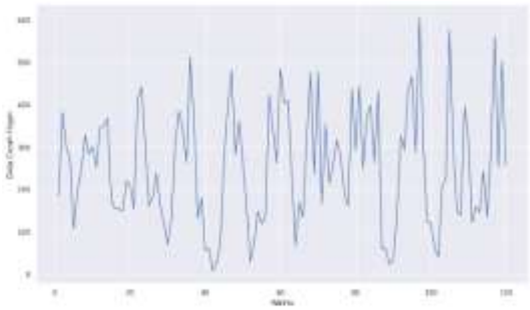
Analisis yang dilakukan pada penelitian ini dimulai dengan pra-pemrosesan data yang telah dikumpulkan, dimana ditemukan data hilang. Penelitian ini mengajukan tiga pendekatan penanganan data hilang, yaitu *mean imputation*, *forward filling*, dan *backward filling*. Hasil penanganan data hilang ini memiliki visualisasi yang tersaji pada Gambar 2.



(a) *mean imputation* (b) *forward filling*

DOI 10.30872/eksponensial.v17i1.1611

 p-ISSN 2085-7829
e-ISSN 2798-3455



(c) *backward filling*

Gambar 2. Plot curah hujan

Berdasarkan Gambar 2, tiga metode imputasi menghasilkan pola deret waktu yang relatif serupa secara visual, dengan fluktuasi curah hujan yang tinggi sepanjang tahun dan puncak curah hujan yang terlihat pada bulan-bulan tertentu, mencerminkan karakteristik iklim tropis basah Kota Palangka Raya yang memiliki variabilitas antar waktu yang cukup besar. Perbedaan utama terlihat pada metode forward filling (b) dan backward filling (c) yang cenderung mempertahankan nilai dari periode sebelumnya atau sesudahnya sehingga menghasilkan pola yang lebih tersambung tanpa lonjakan tajam pada titik data hilang, sementara metode mean imputation (a) mengisi data hilang dengan nilai rata-rata yang relatif stabil sehingga menghasilkan pola yang lebih halus di sekitar titik imputasi.

Gambar 2 data yang telah dilakukan penanganan data hilang memiliki statistik deskriptif yang berbeda dengan sebelum penanganan data hilang Tabel 1, sehingga statistik deskriptif data baru yang digunakan dalam penelitian ini disajikan pada Tabel 2.

Tabel 2. Statistik Deskriptif Curah Hujan Dengan Penanganan Data Hilang

Ukuran	Penanganan data hilang curah hujan (mm)		
	Mean imputation	***Forward filling***	***Backward Filling***
Min.	7,5	7,5	7,5
Q_1	149,1	148,2	148,2
Median	254,1	252,7	252,7
Rata-rata	254,8	252,9	253,2
Q_3	346,5	346,5	346,5
Maks.	604,7	604,7	604,7

Data hasil penanganan data hilang ini selanjutnya dilakukan analisis dengan ARIMA dengan langkah pertama adalah pembagian data, dimana digunakan 80% sebagai data latih dan 20% sebagai data uji. Masing-masing data latih dari hasil penanganan data hilang dilakukan pengujian kestasioneran data dengan uji *augmented dickey fuller* (ADF). Dimana hipotesis awal pengujian ini H_0 adalah data tidak stasioner dan hipotesis alternatifnya H_1 adalah data stasioner, dengan daerah penolakannya adalah $p - value > \alpha = 0.05$ akan gagal tolak H_0. Hasil uji ADF ini tersaji pada Tabel 3.

Tabel 3. Hasil Uji ADF Data Latih Curah Hujan

Penanganan Data Hilang	Pembeda	Nilai $p - value$	Keputusan
Mean Imputation	0	0,01	Tolak H_0
Forward Filling	0	0,01	Tolak H_0
Backward Filling	0	0,01	Tolak H_0

Berdasarkan Tabel 3 ketiga data tersebut memenuhi asumsi kestasioneran data sehingga, bisa dilakukan analisis lanjut dengan menentukan kandidat model ARIMA berdasarkan nilai dari plot *autocorrelation function* (ACF), *partial autocorrelation function* (PACF), dan *extended autocorrelation function* (EACF). Hasil dari kandidat model masing-masing data hasil penanganan data hilang ini dirangkum pada Tabel 4.

Tabel 4. Kandidat Model ARIMA

Penanganan Data Hilang	Kandidat Model	Nilai AIC
Mean Imputation	$ARIMA(0,0,2)$	1184.26
	$ARIMA(1,0,0)$	1187.29
	$ARIMA(1,0,2)$	1186.01
	$\boldsymbol{ARIMA(3,0,1)}$	**1176.12**
	$ARIMA(2,0,2)$	1187.97
Forward Filling	$ARIMA(0,0,2)$	1184.26
	$ARIMA(1,0,0)$	1185.26
	$ARIMA(1,0,2)$	1185.43
	$ARIMA(3,0,2)$	1164.28
	$\boldsymbol{ARIMA(2,0,2)}$	**1163.91**
Backward Filling	$ARIMA(0,0,2)$	1183.54
	$ARIMA(1,0,0)$	1184.94
	$ARIMA(1,0,2)$	1184.82
	$ARIMA(3,0,1)$	1174.21
	$\boldsymbol{ARIMA(2,0,2)}$	**1163.25**

Model terbaik yang diperoleh oleh masing-masing penanganan data hilang ini dipilih berdasarkan kandidat model dengan nilai terkecil, sehingga model terbaik untuk penanganan data hilang dengan *mean imputation* adalah $ARIMA(3,0,1)$ dan dengan *forward* dan *backward filling* adalah $ARIMA(2,0,2)$. Selanjutnya, ketiga model terbaik ini dilakukan pendugaan parameter dan diagnostik model. Diagnostik model yang diajukan dilakukan pengujian formal hipotesis terhadap sisaannya, yaitu uji normalitas dengan uji Kolmogorov-Smirnov, uji tengah sisaan dengan uji t, dan uji autokorelasi dengan uji Ljung-Box. Uji normalitas yang dilakukan akan gagal menolak H_0 atau sisaan menyebar normal jika $p-value > \alpha$. Untuk uji tengah sisaan akan gagal menolak H_0 atau nilai tengah sisaan sama dengan nol jika $p-value > \alpha$. Terakhir untuk uji autokorelasi akan gagal menolak H_0 atau tidak terdapat gejala autokorelasi jika $p-value > \alpha$. Dimana dengan menggunakan taraf signifikansi ($\alpha = 5\%$) hasil pengujian hipotesis dan keputusan ini tersaji pada Tabel 5.

Tabel 5. Pendugaan Parameter Dan Diagnostik Model ARIMA

Penanganan Data Hilang	Pendugaan parameter		Diagnostik model					
			Uji normalitas		Uji tengah sisaan		Uji autokorelasi	
	Nilai dugaan	p-val	p-val	keputusan	p-val	keputusan	p-val	keputusan
Mean Imputation	$\phi_1 = 1.0082$ $\phi_2 = -0.0075$ $\phi_3 = -0.3639$ $\theta_1 = 0.6985$	2.22×10^{-16}* 9.59×10^{-1} 2.11×10^{-4}* 2.22×10^{-16}*	1.33×10^{-15}	Tolak H_0	9.8478×10^{-1}	Gagal Tolak H_0	3.226×10^{-1}	Gagal Tolak H_0
Forward Filling	$\phi_1 = 1.7183$ $\phi_2 = -0.9914$ $\theta_1 = 1.6098$ $\theta_2 = -0.8914$	2.22×10^{-16}* 2.22×10^{-16}* 2.22×10^{-16}* 4.42×10^{-8}*	1.33×10^{-15}	Tolak H_0	9.4088×10^{-1}	Gagal Tolak H_0	8.6998×10^{-1}	Gagal Tolak H_0
Backward Filling	$\phi_1 = 1.7182$ $\phi_2 = -0.9911$ $\theta_1 = 1.6093$ $\theta_2 = -0.8897$	2.22×10^{-16}* 2.22×10^{-16}* 2.22×10^{-16}* 4.73×10^{-9}*	1.33×10^{-15}	Tolak H_0	9.4175×10^{-1}	Gagal Tolak H_0	8.5925×10^{-1}	Gagal Tolak H_0

Ket: *signifikan terhadap $\alpha = 5\%$.

p-ISSN 2085-7829
e-ISSN 2798-3455

Hasil diagnostik model pada Tabel 5 menunjukkan masing-masing model dari hasil penanganan data tidak seimbang memenuhi asumsi nilai tengah sisaannya sama dengan nol dan tidak terdapat gejala autokorelasi. Namun, perlu diketahui bahwa asumsi normalitas residual di atas tidak terpenuhi pada ketiga model, yang ditunjukkan oleh nilai $p-value$ uji normalitas dimana $p-value<\alpha$, sehingga H_0 yang menyatakan residual berdistribusi normal ditolak. Implikasi dari tidak terpenuhinya asumsi normalitas adalah bahwa interval kepercayaan dan uji signifikansi parameter yang ditunjukkan oleh nilai $p-value$ pada pendugaan parameter bersifat tidak eksak atau hanya bersifat asimtotik, sehingga interpretasi signifikansi parameter harus dilakukan dengan hati-hati. Meskipun demikian, estimator parameter ARIMA dengan metode *maximum likelihood estimation* (MLE) tetap konsisten dan efisien secara asimtotik, serta hasil ramalan titik tetap tidak bias dan dapat diandalkan selama asumsi nilai tengah sisaan nol dan tidak adanya autokorelasi terpenuhi. Dengan kata lain, ketidaknormalan residual berdampak pada inferensi statistik, namun tidak secara signifikan mengganggu akurasi nilai ramalan itu sendiri. Sehingga, model yang diperoleh layak dilanjutkan untuk digunakan untuk peramalan. Secara matematis dengan menggunakan operator *backshift* model ARIMA dengan nilai dugaan pada Tabel 5 memiliki persamaan tersaji pada tabel 6 berikut.

Tabel 6. Persamaan Model ARIMA

Penanganan Data Hilang	**Model**
Mean Imputation	$Y_t=\mu+1.0082Y_{t-1}-0.0075Y_{t-2}-0.3639Y_{t-3}+e_t-0.6985e_{t-1}$
Forward Filling	$Y_t=\mu+1.7183Y_{t-1}-0.9914Y_{t-2}+e_t-1.6098e_{t-1}+0.8914e_{t-2}$
Backward Filling	$Y_t=\mu+1.7182Y_{t-1}-0.9911Y_{t-2}+e_t-1.6093e_{t-1}+0.8897e_{t-2}$

Model ARIMA pada Tabel 6 selanjutnya digunakan untuk meramalkan data baik pada data latih maupun untuk validasi dengan data uji. Peramalan dengan model ARIMA ini selanjutnya diukur kebaikannya dengan menggunakan ukuran kebaikan RMSE, MAE, dan MAPE. Akurasi hasil peramalan ini dirangkum pada Tabel 7.

Tabel 7. Akurasi Hasil Peramalan Model ARIMA

Penanganan Data Hilang		**Akurasi Ramalan**		
		RMSE	MAE	MAPE
Mean Imputation	Data Latih	104,59	86,45	36,33
	Data Uji	**124,17**	**103,57**	**40,32**
Forward Filling	Data Latih	97,12	80,19	35,29
	Data Uji	139,97	108,17	41,21
Backward Filling	Data Latih	96.82	79,97	35,02
	Data Uji	139,97	108,17	41,21

Akurasi hasil ramalan berdasarkan Tabel 7 menunjukkan model ARIMA dengan penanganan data hilang *mean imputation* memiliki hasil kekuatan ramalan yang baik ketika dilakukan validasi dengan data uji dibandingkan yang dimodelkan ARIMA dengan penanganan data hilang *forward* dan *backward filling*. Dimana nilai ini ditunjukkan nilai MAPE ramalannya merupakan yang terkecil, yaitu sebesar 40.32% atau akurasi ramalannya sebesar 59.68% dimana menurut penelitian sebelumnya pada (Blasco et al., 2013) nilai ini masuk dalam kategori cukup. Selain itu, temuan lainnya perlu diketahui bahwa nilai RMSE, MAE, dan MAPE pada data uji untuk metode *forward filling* dan *backward filling* menunjukkan angka yang persis sama, yaitu RMSE sebesar 139,97, MAE sebesar 108,17, dan MAPE sebesar 41,21%. Kesamaan ini secara teknis karena pada metode *forward filling* dan *backward filling*, memiliki pola imputasi data hilang bersifat simetris terhadap waktu, terutama jika data hilang bersifat tunggal atau tidak berurutan. Ketika data hilang diisi dengan nilai sebelumnya atau *forward* atau nilai sesudahnya atau *backward*, efek lokal dari imputasi tersebut dapat menghasilkan struktur residual yang hampir identik pada tahap peramalan, khususnya jika data hilang tersebar secara sporadis/ acak dan tidak membentuk pola yang panjang.

Hasil analisis yang telah dilakukan oleh penelitian ini diharapkan mampu memberikan wawasan penting terhadap masalah data hilang data data deret waktu da khususnya permasalahan curah hujan. Penjelasan lebih lanjut penelitian ini mengangkat masalah iklim terkait curah hujan di Kota Palangka Raya berdasarkan sudut pandang statistika, yang mana memberikan gambaran terkait peramalan data iklim yang memiliki masalah data hilang. Hal ini mengindikasikan bahwa integrasi statistika dalam penyelesaian masalah sehari-hari memiliki kekuatan yang baik dan mampu dikembangkan lebih lanjut

4. Kesimpulan

Berdasarkan hasil analisis yang dilakukan untuk mengevaluasi kinerja penanganan data hilang dengan model ARIMA dalam peramalan curah hujan di Palangka Raya melalui tiga Teknik imputasi, yaitu *mean imputation*, *forward filling*, dan *backward filling*. Hasil menunjukkan bahwa data yang dimodelkan dengan ARIMA dan dilakukan penanganan data hilang dengan *mean imputation* menghasilkan peramalan paling akurat ketika dilakukan validasi dengan data uji. Hasil ini didasarkan pada nilai RMSE, MAE, dan MAPEnya yang terkecil dibandingkan imputasi data hilang yang lain, yaitu sebesar 124.17, 103.57, dan 40.32%. Penelitian lanjutan yang mampu dikembangkan dengan menyoroti penyempurnaan lebih lanjut adalah dengan membandingkan penanganan data hilang lainnya seperti *interpolation imputation* hingga *multiple imputation* maupun dalam eksplorasi metode pendekatan peramalannya melalui metode *ensemble* yang memungkinkan peningkatan akurasi ramalan.

5. Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Program Studi Matematika, Universitas Palangka Raya, atas dukungan yang diberikan. Penulis juga menyampaikan terima kasih kepada Badan Pusat Statistik (BPS) Kota Palangka Raya atas penyediaan data curah hujan yang digunakan dalam penelitian ini.

6. Daftar Pustaka

Alabdulrazzaq, H., Alenezi, M. N., Rawajfih, Y., Alghannam, B. A., Al-Hassan, A. A., & Al-Anzi, F. S. (2021). On the accuracy of ARIMA based prediction of COVID-19 spread. *Results in Physics*, *27*, 104509.

AlSalehy, A. S., & Bailey, M. (2025). Improving Time Series Data Quality: Identifying Outliers and Handling Missing Values in a Multilocation Gas and Weather Dataset. *Smart Cities*, *8*(3), 82.

Appaia, L., & Palraj, S. (2023). On replacement of outliers and missing values in time series. *EQA-International Journal of Environmental Quality*, *53*, 1–10.

Blasco, B. C., Moreno, J. J. M., Pol, A. P., & Abad, A. S. (2013). Using the R-MAPE index as a resistant measure of forecast accuracy. *Psicothema*, *25*(4), 500–506.

Enders, C. K. (2022). *Applied missing data analysis*. Guilford Publications.

Hodson, T. O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development Discussions*, *2022*, 1–10.

Khikmah, K. N. (2024). *Kajian Kinerja Multiple Long Short-Term Memory untuk Analisis Data Deret Waktu Takstasioner* [Master Thesis]. IPB University.

Khikmah, K. N., Sadik, K., & Indahwati, I. (2023). Transfer Function And Arima Model For Forecasting Bi Rate In Indonesia. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *17*(3), 1359–1366.

Khikmah, K. N., Sadik, K., & Notodiputro, K. A. (2025). Simulation and Empirical Studies of Long Short-Term Memory Performance to Deal with Limited Data. *Jurnal Online Informatika*, *10*(1), 216–226.

Kobiela, D., Krefta, D., Król, W., & Weichbroth, P. (2022). ARIMA vs LSTM on NASDAQ stock exchange data. *Procedia Computer Science*, *207*, 3836–3845.

Ospina, R., Gondim, J. A. M., Leiva, V., & Castro, C. (2023). An overview of forecast analysis with ARIMA models during the COVID-19 pandemic: Methodology and case study in Brazil. *Mathematics*, *11*(14), 3069.

Risnayah, S. (2023). Penerapan Imputasi Locf Dan Cross Mean Dalam Pengisian Data Kosong Pada Curah Hujan Harian Arg. *Megasains*, *14*(2), 23–31.

Tandon, H., Ranjan, P., Chakraborty, T., & Suhag, V. (2022). Coronavirus (COVID-19): ARIMA-based Time-series Analysis to Forecast near Future and the Effect of School Reopening in India. *Journal of Health Management*, 09720634221109087.

Wang, X., Zhang, H., Wang, P., Zhang, Y., Wang, B., Zhou, Z., & Wang, Y. (2023). An observed value consistent diffusion model for imputing missing values in multivariate time series. *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2409–2418.

Zhang, Y., & Thorburn, P. J. (2022). Handling missing data in near real-time environmental monitoring: A system and a review of selected methods. *Future Generation Computer Systems*, *128*, 63–72.