

PERAMALAN KUALITAS UDARA DI KOTA BALIKPAPAN BERDASARKAN INDIKATOR NILAI PM2.5 MENGGUNAKAN METODE AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA)

(Air Quality Forecasting In Balikpapan City Based On PM2.5 Indicator Values Using The Autoregressive Integrated Moving Average (ARIMA) Method)

Judah Toba Butar Butar¹, Anastasya², Riswanty Margareth Malau³, Galavinozky Roigabe Gumilang Rajaguguk⁴, Surya Puspita Sari⁵, Magdalena Effendi⁶

Institut Teknologi Kalimantan^{1, 2, 3, 4, 5, 6}

E-mail: 16231028@student.itk.ac.id

ABSTRAK

Kualitas udara merupakan indikator penting dalam menilai kondisi lingkungan karena berpengaruh langsung terhadap kesehatan manusia. Peningkatan aktivitas industri, transportasi, serta pembangunan infrastruktur di Kota Balikpapan seiring dengan pengembangan Ibu Kota Nusantara (IKN) berpotensi meningkatkan konsentrasi polutan udara, khususnya partikulat halus PM2.5. Penelitian ini bertujuan untuk menganalisis pola dan meramalkan konsentrasi PM2.5 di Kota Balikpapan. Penelitian ini menggunakan metode Autoregressive Integrated Moving Average (ARIMA) untuk memodelkan data deret waktu PM2.5 harian. Data yang digunakan mencakup periode Januari hingga September 2025 dengan total 273 observasi, yang dibagi menjadi 80% data training yaitu sebanyak 218 observasi dan 20% data testing sebanyak 55 observasi. Metode ARIMA dipilih karena kemampuannya dalam menangkap pola fluktuatif pada data deret waktu. Hasil penelitian menunjukkan bahwa model ARIMA(2,0,1) merupakan model terbaik untuk peramalan konsentrasi PM2.5 di Kota Balikpapan berdasarkan kriteria pemilihan model dan evaluasi kinerja peramalan. Model ini mampu merepresentasikan pola data historis dengan baik dan memberikan hasil peramalan yang cukup akurat pada data pengujian. Kesimpulan dari penelitian ini menunjukkan bahwa model ARIMA(2,0,1) dapat digunakan sebagai alat peramalan kualitas udara, khususnya konsentrasi PM2.5 di Kota Balikpapan, serta berpotensi mendukung pengambilan kebijakan dalam pengendalian pencemaran udara di wilayah tersebut.

Kata kunci: Kualitas udara, PM2.5, ARIMA, Peramalan, Balikpapan.

ABSTRACT

Air quality is an important indicator in assessing environmental conditions because it directly affects human health. The increase in industrial activities, transportation, and infrastructure development in Balikpapan City along with the development of the Nusantara Capital (IKN) has the potential to increase the concentration of air pollutants, particularly fine particulate matter PM2.5. This study aims to analyze patterns and forecast PM2.5 concentrations in Balikpapan City. This study uses the Autoregressive Integrated Moving Average (ARIMA) method to model daily PM2.5 time series data. The data used covers the period from January to September 2025 with a total of 273 observations, divided into 80% training data, which is 218 observations, and 20% testing data, which is 55 observations. The ARIMA method was chosen because of its ability to capture fluctuating patterns in time series data. The research results indicate that the ARIMA(2,0,1) model is the best model for forecasting PM2.5 concentration in Balikpapan City based on model selection criteria and forecast performance evaluation. This model is able to represent historical data patterns well and provides fairly accurate forecasting results on the test data. The conclusion of this study shows that the ARIMA(2,0,1) model can be used as an air quality forecasting tool, particularly for PM2.5 concentration in Balikpapan City, and has the potential to support policy-making in controlling air pollution in the area.

Keywords: Air quality, PM2.5, ARIMA, Forecasting, Balikpapan.

PENDAHULUAN

Kualitas udara merupakan indikator penting dalam menilai kondisi lingkungan karena berpengaruh langsung terhadap kesehatan manusia, ekosistem, dan kualitas hidup. Menurut Rita et al. (2016), pencemaran udara disebabkan oleh polutan berbentuk partikel dan gas, dengan partikulat halus berukuran kurang dari $2,5 \mu\text{m}$ (PM_{2.5}) menjadi perhatian utama. PM_{2.5} mampu bertahan lama di udara dan menembus sistem pernapasan, sehingga meningkatkan risiko penyakit pernapasan dan kardiovaskular. World Health Organization (WHO) menetapkan ambang batas aman konsentrasi tahunan PM_{2.5} sebesar $5 \mu\text{g}/\text{m}^3$ sebagai acuan global dalam penilaian kualitas udara (WHO, 2021).

Kota Balikpapan, sebagai pusat industri dan transportasi di Kalimantan Timur, mengalami peningkatan aktivitas ekonomi seiring pembangunan Ibu Kota Nusantara (IKN). Kondisi tersebut berpotensi meningkatkan konsentrasi PM_{2.5}, yang jika tidak dikendalikan akan berdampak pada kesehatan masyarakat dan kenyamanan lingkungan. Oleh karena itu, diperlukan analisis yang mampu menggambarkan kondisi kualitas udara saat ini serta memproyeksikannya di masa mendatang sebagai dasar pengambilan kebijakan lingkungan.

Berbagai studi sebelumnya menunjukkan bahwa metode deret waktu efektif dalam menganalisis dan meramalkan kualitas udara. Aladağ et al. (2021) membuktikan bahwa model ARIMA mampu memberikan hasil prediksi yang akurat terhadap konsentrasi PM_{2.5}. Penelitian serupa oleh Kurniawan (2024) dan Putra (2024) juga mengonfirmasi efektivitas ARIMA dan model hybrid dalam memprediksi polutan udara di wilayah perkotaan Indonesia. Penelitian mengenai peramalan kualitas udara umumnya terpusat di kota-kota besar seperti Jakarta dan Surabaya, sehingga kajian serupa di wilayah Kalimantan, khususnya Kota Balikpapan, belum banyak dikembangkan.

Berdasarkan hal tersebut, penelitian ini bertujuan untuk menganalisis pola historis dan melakukan peramalan kualitas udara di Kota Balikpapan berdasarkan nilai PM_{2.5} menggunakan metode Autoregressive Integrated Moving Average (ARIMA). Hasil penelitian diharapkan dapat menjadi dasar bagi pemerintah daerah dalam pengendalian dan mitigasi pencemaran udara guna mewujudkan lingkungan yang sehat dan berkelanjutan.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode deret waktu (*time series analysis*) untuk memodelkan dan meramalkan konsentrasi polutan udara PM_{2.5} di Kota Balikpapan. Lokasi penelitian dipilih karena Balikpapan merupakan salah satu kota dengan aktivitas industri dan transportasi yang meningkat seiring pembangunan Ibu Kota Nusantara (IKN).

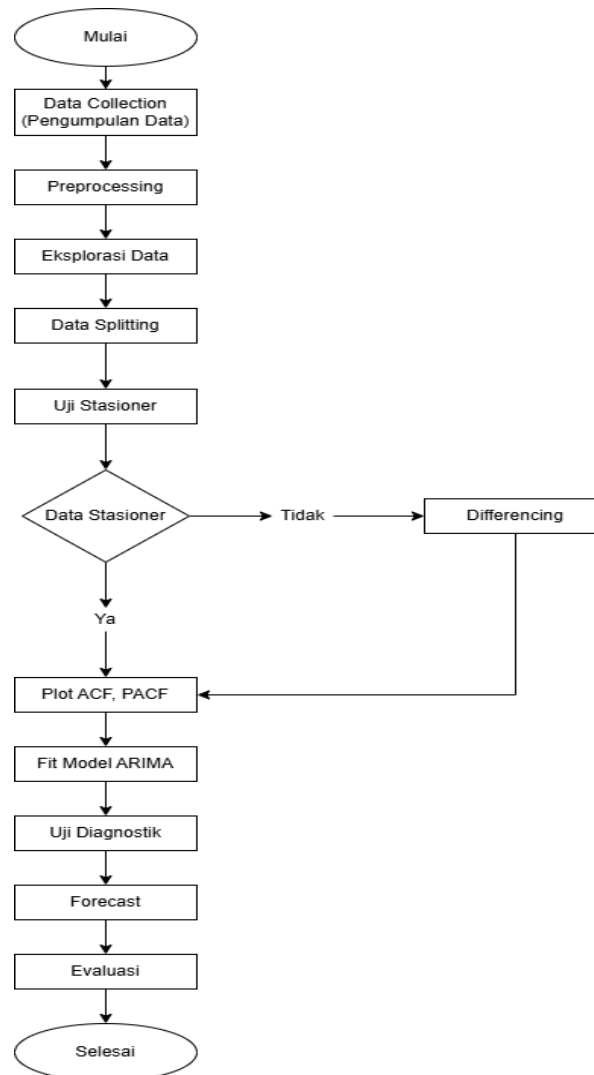
Sumber Data

Penelitian ini menggunakan metode kuantitatif dengan memanfaatkan data sekunder yang diperoleh dari World Air Quality Index (WAQI) yaitu sebuah proyek global yang bertujuan untuk menyediakan informasi kualitas udara secara real-time (langsung) dari berbagai kota di seluruh dunia, termasuk Kota Balikpapan. Data yang digunakan dalam penelitian ini merupakan data harian kualitas udara dengan fokus pada polutan partikulat halus (PM_{2.5}) yang diukur oleh stasiun pemantauan udara Balikpapan. Periode data yang digunakan mencakup rentang waktu 1 Januari 2025 hingga 30 September 2025, dengan total 272 data pengamatan harian. Data harian yang diambil merupakan median dari keseluruhan nilai PM_{2.5} yang tercatat di setiap waktu observasinya.

Tahapan Penelitian

Langkah-langkah penelitian dalam peramalan kualitas udara di Kota Balikpapan menggunakan metode Autoregressive Integrated Moving Average (ARIMA) dilakukan secara sistematis

berdasarkan pendekatan analisis deret waktu modern yang banyak digunakan dalam studi lingkungan dan kualitas udara (Hyndman & Athanasopoulos, 2021; Shumway & Stoffer, 2017). dilakukan secara sistematis melalui beberapa tahap berikut:



Gambar 1. Diagram Alir Penelitian

1. *Data Collection (Pengumpulan Data)*

Data yang digunakan dalam penelitian ini berupa data konsentrasi PM2.5 harian di Kota Balikpapan yang diperoleh dari World Air Quality Index (WAQI). Data tersebut merupakan data deret waktu (time series), yaitu data yang diamati secara berurutan dalam interval waktu yang sama, sehingga sesuai untuk dimodelkan menggunakan pendekatan ARIMA (Hyndman & Athanasopoulos, 2021).

2. *Preprocessing*

Tahap preprocessing bertujuan untuk memastikan kualitas data sebelum analisis lebih lanjut. Proses ini meliputi pemeriksaan dan penanganan missing value, penghapusan data duplikat, serta identifikasi dan penanganan outlier. Tahapan ini penting karena kualitas data sangat memengaruhi performa model ARIMA, khususnya dalam peramalan jangka pendek (Shumway & Stoffer, 2017).

3. *Eksplorasi Data*

Eksplorasi data dilakukan untuk memahami karakteristik dan pola data PM2.5 di Kota Balikpapan. Analisis ini mencakup perhitungan statistik deskriptif serta visualisasi time series plot untuk mengidentifikasi tren, pola musiman, dan fluktuasi data. Tahap ini memberikan gambaran

awal mengenai struktur data dan membantu dalam menentukan metode analisis lanjutan (Hyndman & Athanasopoulos, 2021).

4. Data Splitting

Data dibagi menjadi dua bagian, yaitu data training dan data testing. Data training digunakan untuk membangun model ARIMA, sedangkan data testing digunakan untuk mengevaluasi kemampuan model dalam memprediksi data di luar sampel. Pendekatan ini umum digunakan dalam analisis deret waktu untuk mengukur kemampuan generalisasi model (Brockwell & Davis, 2016).

5. Uji Stasioner

Sebelum pemodelan ARIMA dilakukan, data diuji kestasionerannya menggunakan Augmented Dickey-Fuller (ADF) Test. Uji ini digunakan untuk menentukan apakah data memiliki akar unit (unit root) atau tidak. Data dinyatakan stasioner apabila nilai p-value lebih kecil dari tingkat signifikansi 0,05. Jika data belum stasioner, maka dilakukan proses differencing hingga data memenuhi asumsi kestasioneritas (Hyndman & Athanasopoulos, 2021; Shumway & Stoffer, 2017).

6. Identifikasi Model

Setelah data dinyatakan stasioner, dilakukan identifikasi model ARIMA dengan menganalisis Autocorrelation Function (ACF) dan Partial Autocorrelation Function (PACF). Plot ACF dan PACF digunakan untuk menentukan nilai parameter p (autoregressive) dan q (moving average), sedangkan parameter d ditentukan berdasarkan jumlah proses differencing yang diperlukan. Pendekatan ini merupakan prosedur standar dalam pemodelan ARIMA modern (Hyndman & Athanasopoulos, 2021).

7. Fit Model ARIMA

Model ARIMA dengan kombinasi parameter terbaik (p, d, q) dibangun menggunakan data training. Estimasi parameter dilakukan dengan metode Maximum Likelihood Estimation (MLE). Pemilihan model terbaik ditentukan berdasarkan nilai Akaike Information Criterion (AIC) dan Bayesian Information Criterion (BIC), di mana model dengan nilai AIC dan BIC terkecil dipilih sebagai model yang paling optimal (Brockwell & Davis, 2016; Hyndman & Athanasopoulos, 2021).

Model ARIMA terbentuk berdasarkan 3 komponen yaitu autoregressive yang memiliki orde $p \neq 0$ dan $q = 0$, integrated yang memiliki orde d ditentukan berdasarkan banyaknya proses differencing dilakukan, dan moving average yang memiliki orde $q \neq 0$ dan $p = 0$. Apabila orde $p \neq 0$ dan $q \neq 0$, maka disebut model ARMA(p,q). Adapun model-model tersebut dapat dijelaskan lebih rinci sebagai berikut:

a. Model AR (Autoregressive)

Secara umum, model *autoregressive* dengan orde p dapat dirumuskan sebagai berikut:

$$Z_t = \mu + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \varepsilon_t \tag{1}$$

dengan:

μ = Nilai konstan

ϕ_k = Parameter autoregressive ke- k ; untuk $k = 1, 2, \dots, p$

ε_t = Nilai residual pada saat t

b. Model MA (Moving Average)

Secara umum, model moving average dengan orde q dapat dirumuskan sebagai berikut:

$$Z_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \tag{2}$$

dengan:

θ_j = Parameter moving average ke- j ; $j = 1, 2, \dots, q$

c. Model ARMA (Autoregressive Moving Average)

Model ini merupakan model gabungan dari model AR (p) dan model MA (q), yang dapat dirumuskan sebagai berikut:

$$Z_t = \mu + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \varepsilon_t - \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (3)$$

Model ARMA diperoleh dengan menggeneralisasikan model ARMA dan menambahkan gagasan integrasi di dalamnya.

8. Uji Diagnostik

Uji diagnostik dilakukan untuk memastikan bahwa model yang diperoleh telah memenuhi asumsi kelayakan. Residual model harus bersifat acak (white noise) dan tidak menunjukkan adanya autokorelasi. Pengujian ini dilakukan menggunakan Ljung-Box test serta analisis visual terhadap plot residual dan ACF residual. Jika residual masih menunjukkan pola tertentu, maka model perlu disesuaikan kembali (Hyndman & Athanasopoulos, 2021).

9. Forecasting

Setelah model ARIMA dinyatakan valid, dilakukan peramalan konsentrasi PM2.5 untuk periode waktu berikutnya. Hasil peramalan divisualisasikan dalam bentuk grafik untuk membandingkan nilai aktual dan nilai prediksi, sehingga pola peramalan dapat dianalisis secara lebih jelas (Shumway & Stoffer, 2017).

10. Evaluasi Model

Tahap terakhir adalah melakukan peramalan nilai PM2.5 untuk periode tertentu berdasarkan model ARIMA terbaik yang telah diperoleh. Akurasi hasil peramalan diukur menggunakan beberapa ukuran kesalahan, yaitu Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE). Model dengan nilai kesalahan terkecil dianggap sebagai model paling akurat untuk memprediksi konsentrasi PM2.5 di masa mendatang (Brockwell & Davis, 2016).

HASIL DAN PEMBAHASAN

Preprocessing Data

Pada tahap ini dilakukan proses pembersihan dan penyiapan data agar layak dianalisis. Pemeriksaan awal menunjukkan bahwa tidak terdapat nilai hilang (missing values) pada dataset, sehingga tidak dilakukan penghapusan data. Nilai PM2.5 yang bernilai 0 diganti dengan nilai rata-rata pada bulan yang sama agar tidak mengganggu hasil analisis. Proses deteksi outlier menggunakan metode *Interquartile Range (IQR)* menghasilkan batas bawah sebesar $2 \mu\text{g}/\text{m}^3$ dan batas atas sebesar $66 \mu\text{g}/\text{m}^3$. Berdasarkan hasil tersebut, tidak ditemukan nilai PM2.5 yang berada di luar rentang IQR sehingga tidak ada data yang dihapus. Setelah seluruh tahap preprocessing dilakukan, jumlah data tetap 269 observasi, dan dataset telah siap digunakan untuk analisis statistika deskriptif.

Eksplorasi Data

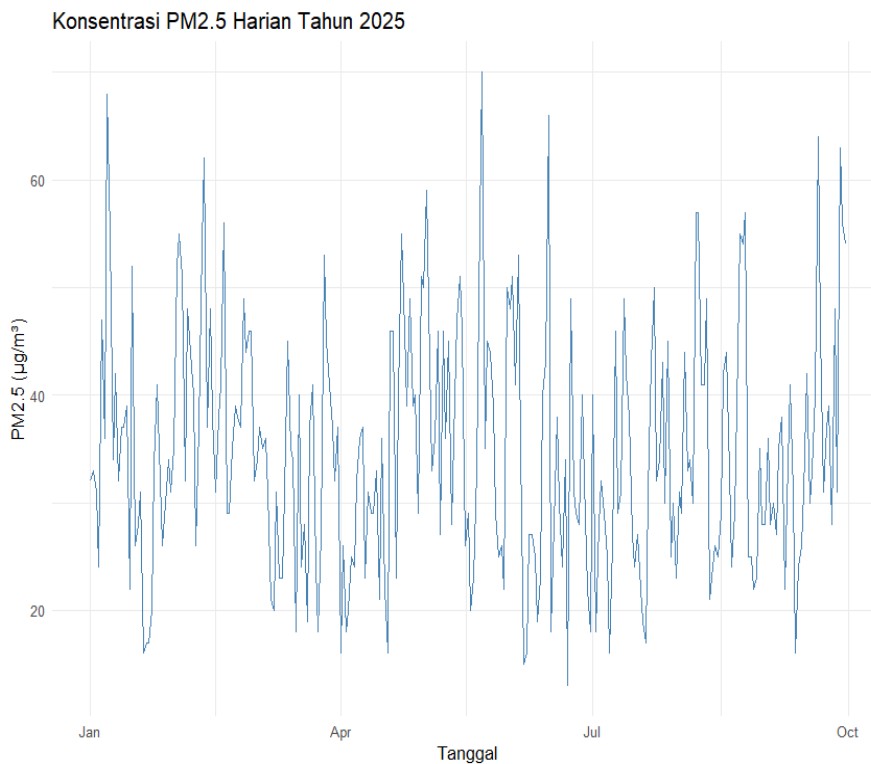
Berdasarkan hasil analisis statistika deskriptif (**Tabel 1.**) terhadap data konsentrasi PM2.5 harian periode 1 Januari - 30 September tahun 2025, diperoleh informasi bahwa nilai PM2.5 memiliki rentang antara $13 \mu\text{g}/\text{m}^3$ hingga $70 \mu\text{g}/\text{m}^3$. Nilai rata-rata (mean) sebesar $34,84 \mu\text{g}/\text{m}^3$, sedangkan median bernilai $33 \mu\text{g}/\text{m}^3$, menunjukkan bahwa sebaran data relatif simetris dengan sedikit kecenderungan ke arah nilai yang lebih tinggi. Nilai kuartil pertama (Q1) adalah $27 \mu\text{g}/\text{m}^3$, dan

kuartil ketiga (Q3) adalah $42 \mu\text{g}/\text{m}^3$, sehingga sebagian besar nilai PM2.5 berada di antara rentang tersebut.

Tabel 1. Statistika Deskriptif

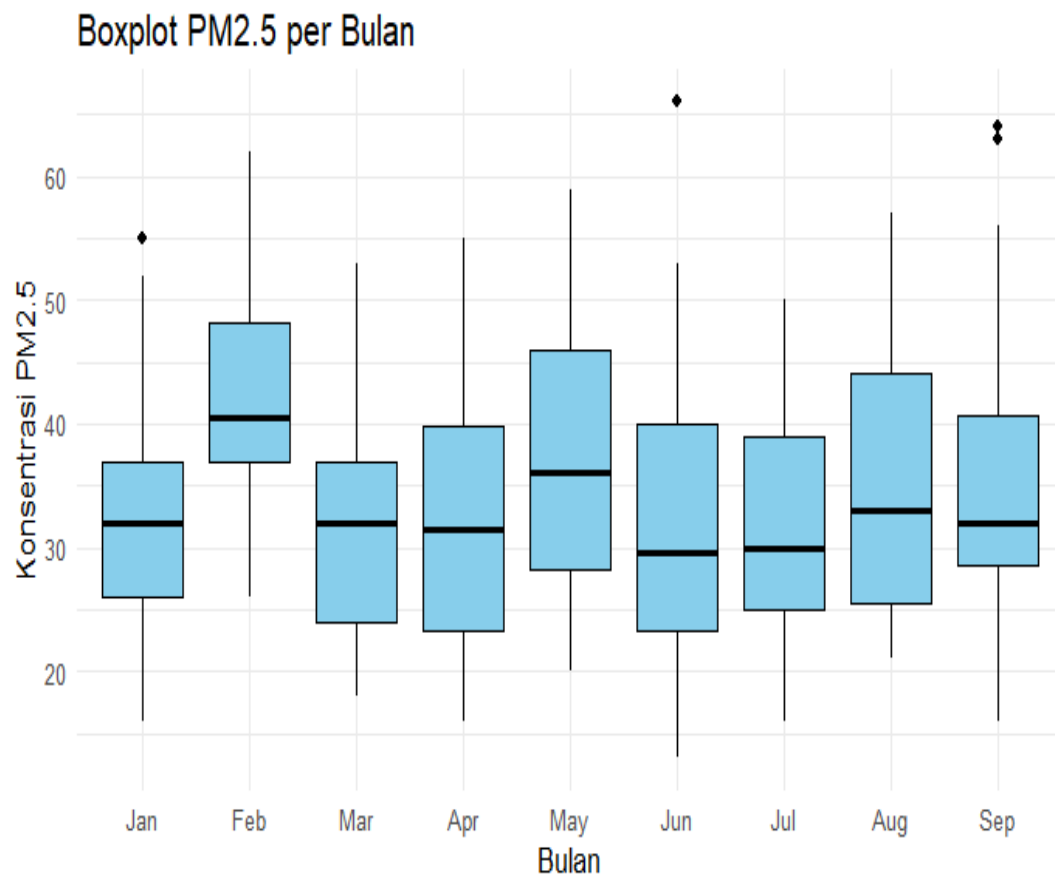
	Nilai
Min	13.00
Q1	27.00
Median	33.00
Mean	34.84
Q3	42.00
Max	70.00

Grafik di bawah (**Gambar 2.**) menunjukkan pola fluktuasi konsentrasi PM2.5 harian selama periode Januari hingga September 2025. Berdasarkan grafik tersebut dapat diamati bahwa pola dari data data penelitian ini adalah musiman bukan tren. Secara umum, data memperlihatkan perubahan nilai PM2.5 yang cukup dinamis dari hari ke hari, dengan kisaran antara sekitar 15 hingga $70 \mu\text{g}/\text{m}^3$. Terlihat bahwa konsentrasi PM2.5 mengalami beberapa puncak (spike) pada bulan-bulan tertentu, khususnya di sekitar Januari, Mei, dan Agustus, yang mengindikasikan adanya kejadian peningkatan polusi udara sesaat. Sebaliknya, periode Maret hingga April tampak relatif lebih stabil dengan nilai PM2.5 yang cenderung lebih rendah.



Gambar 2. Grafik Time Series Konsentrasi PM2.5

Pada Boxplot (**Gambar 3.**) menunjukkan sebaran konsentrasi PM2.5 per bulan selama 1 Januari - 30 September tahun 2025. Terlihat bahwa bulan Februari dan Mei memiliki median tertinggi, menandakan kualitas udara yang lebih buruk pada periode tersebut. Sementara itu, bulan Januari, Maret, dan Juli menunjukkan median lebih rendah, mengindikasikan udara relatif lebih bersih. Beberapa outlier muncul terutama pada bulan Januari, Mei, dan September, yang menandakan adanya peningkatan ekstrem PM2.5 pada hari-hari tertentu. Variasi yang cukup besar pada bulan Mei dan Agustus menunjukkan fluktuasi harian yang tinggi, menandakan adanya pola musiman dalam konsentrasi PM2.5. Temuan ini mengindikasikan bahwa kualitas udara berubah seiring waktu dan perlu diperhatikan pada tahap pemodelan runtun waktu selanjutnya.



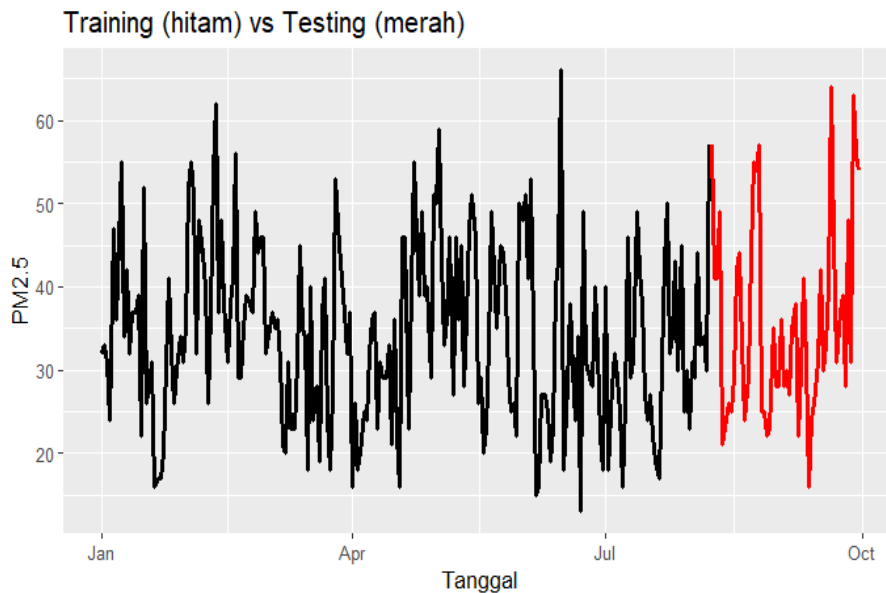
Gambar 3. Boxplot Bulanan

Data Splitting

Berdasarkan Grafik Data Splitting (**Gambar 4.**), data PM2.5 harian yang digunakan dalam penelitian ini dipisahkan menjadi data training dan data testing menggunakan pendekatan proporsi data, yaitu 80% data sebagai training dan 20% data sebagai testing. Pendekatan pembagian data dengan proporsi tersebut umum digunakan dalam analisis deret waktu untuk membangun model peramalan yang stabil sekaligus menyediakan data yang cukup untuk evaluasi kinerja model (out-of-sample evaluation) (Hyndman & Athanasopoulos, 2021; Shumway & Stoffer, 2017). Seluruh data yang dianalisis mencakup periode Januari hingga September 2025 dengan total 273 observasi harian. Dari keseluruhan data tersebut, 218 observasi digunakan sebagai data training, sedangkan 55 observasi digunakan sebagai data testing.

Pemisahan data dilakukan secara berurutan berdasarkan waktu (time-based split) tanpa melakukan pengacakan, sehingga urutan kronologis data tetap terjaga. Pendekatan ini sesuai dengan karakteristik data deret waktu, di mana penggunaan data masa depan untuk membangun model harus dihindari agar tidak menimbulkan look-ahead bias (Hyndman & Athanasopoulos, 2021). Data training mencakup sebagian besar periode Januari hingga September 2025 dan menunjukkan fluktuasi konsentrasi PM2.5 yang cukup dinamis dengan kisaran nilai antara 15 hingga 65 $\mu\text{g}/\text{m}^3$. Beberapa puncak konsentrasi terlihat pada periode awal hingga pertengahan tahun yang mengindikasikan adanya peningkatan polusi udara pada hari-hari tertentu. Sementara itu, data testing yang mencakup 20% data terakhir dari keseluruhan periode pengamatan juga memperlihatkan pola fluktuatif yang serupa. Hal ini menunjukkan bahwa data testing bersifat representatif dan dapat digunakan untuk mengevaluasi kemampuan model ARIMA dalam melakukan peramalan terhadap data di luar sampel pelatihan (Shumway & Stoffer, 2017).

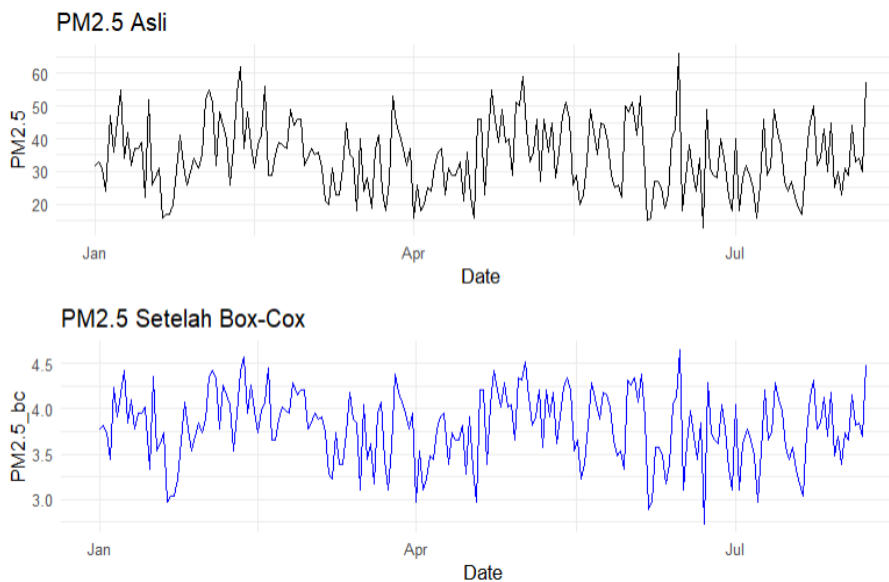
Pendekatan pembagian data ini memastikan bahwa variasi pola data pada periode pengamatan telah tercakup dengan baik, serta memberikan dasar yang konsisten dan memadai untuk melakukan pemodelan deret waktu dan evaluasi performa prediksi secara lebih akurat.



Gambar 4. Grafik Data Splitting

Uji Stasioner

Uji stasioneritas dilakukan menggunakan Augmented Dickey-Fuller Test (ADF Test) terhadap data training PM2.5. Hasil pengujian menunjukkan nilai statistik uji sebesar $-5,5067$ dengan *lag order* 5 dan nilai *p-value* sebesar 0,01. Karena *p-value* lebih kecil dari tingkat signifikansi 0,05, maka hipotesis nol yang menyatakan bahwa deret waktu bersifat non-stasioner ditolak. Dengan demikian, data PM2.5 pada subset training dapat disimpulkan telah bersifat stasioner, sehingga layak digunakan untuk tahap pemodelan deret waktu selanjutnya tanpa memerlukan proses *differencing* tambahan. Stasioneritas ini memastikan bahwa model dapat menghasilkan estimasi parameter yang stabil dan dapat diinterpretasikan dengan baik.

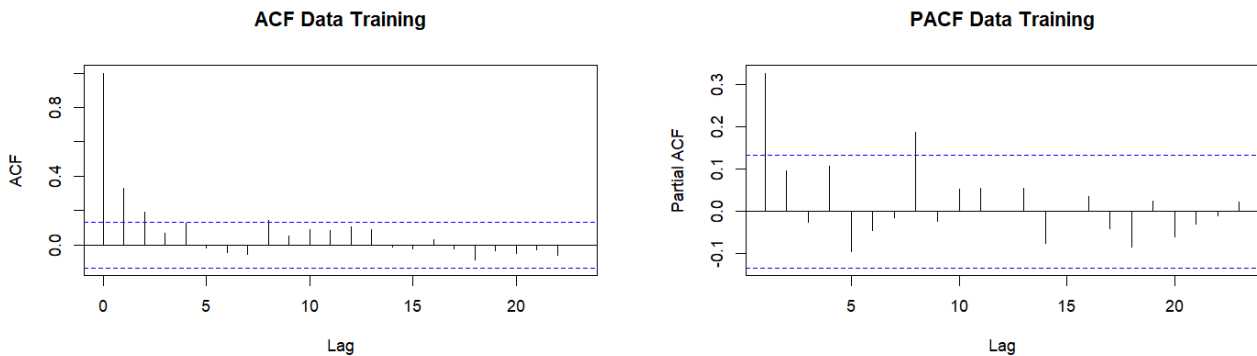


Gambar 5. Grafik Transformasi Boxcox

Nilai lambda Box-Cox sebesar 0,0489 yang mendekati nol menunjukkan bahwa data PM2.5 memiliki varians yang tidak konstan (heteroskedastis) sehingga memerlukan transformasi yang mendekati logaritmik. Perbandingan grafik sebelum dan sesudah transformasi memperlihatkan bahwa setelah Box-Cox, fluktuasi data menjadi lebih stabil dan lonjakan ekstrem berkurang, sehingga pola deret waktu terlihat lebih homogen. Dengan demikian, transformasi Box-Cox berhasil menstabilkan varians data PM2.5 dan membuat data lebih memenuhi asumsi pemodelan ARIMA untuk analisis dan peramalan.

Identifikasi Model

Berdasarkan (**Gambar 6.**) menampilkan plot ACF (Autocorrelation Function) dan PACF (Partial Autocorrelation Function), terlihat bahwa nilai autokorelasi menurun secara bertahap dan signifikan pada beberapa lag awal, sedangkan PACF menunjukkan pemotongan tajam setelah lag ke-1. Pola ini menunjukkan bahwa data PM2.5 memiliki keterkaitan jangka pendek antara periode waktu sebelumnya. Dengan demikian, indikasi awal menunjukkan bahwa model ARIMA yang sesuai kemungkinan memiliki komponen autoregressive (AR) orde 1 atau 2 dan/atau komponen moving average (MA) orde rendah. Hal ini menjadi dasar dalam pemilihan parameter (p, d, q) untuk model ARIMA terbaik yang akan dibangun pada tahap berikutnya.



Gambar 6. Plot ACF dan PACF

Fit Model ARIMA

Berdasarkan hasil pemodelan menggunakan Auto ARIMA (**Tabel 2.**), diperoleh model terbaik yaitu ARIMA(2,0,1) dengan mean tidak sama dengan nol, yang ditunjukkan dari nilai koefisien AR1 sebesar -0.4974, AR2 sebesar 0.3633, dan MA1 sebesar 0.8186 yang seluruhnya signifikan, menggambarkan bahwa nilai PM2.5 saat ini dipengaruhi secara linier oleh dua periode sebelumnya dan koreksi kesalahan satu periode sebelumnya. Nilai mean sebesar 34.3032 menunjukkan bahwa rata-rata level polusi PM2.5 berada pada kisaran tersebut setelah memperhitungkan komponen autoregresif dan moving average. Nilai $\sigma^2 = 101.1$ menunjukkan variabilitas residual model yang relatif moderat, sedangkan kriteria informasi AIC = 1616.30, AICc = 1616.58, dan BIC = 1633.17 menunjukkan bahwa model telah memberikan keseimbangan optimal antara tingkat kerumitan dan akurasi. Evaluasi akurasi berdasarkan training set memberikan hasil RMSE = 9.962, MAE = 7.964, dan MAPE = 26.72%, yang mengindikasikan bahwa model mampu melakukan peramalan dengan tingkat kesalahan sekitar 26%. Selain itu, nilai ACF1 sebesar 0.00245 mendekati nol menunjukkan bahwa residual tidak memiliki autokorelasi yang tersisa, sehingga model dapat dianggap sesuai (fit) dan residual bersifat white noise.

Model dengan Auto ARIMA.

Tabel 2. Hasil Fit Model ARIMA (2,0,1)

Komponen	Parameter	Estimasi	Std.Error	Statistik Model	Nilai	Training Set Error	Nilai
AR(1)	ar1	-0.4974	0.1152	Sigma ²	101.1	ME	0.0112
AR(2)	ar2	0.3633	0.0655	Log-Likelihood	-803.15	RMSE	9.9623
MA(1)	ma1	0.8186	0.1059	AIC	1616.3	MAE	7.9644
Mean	mean	34.3032	1.0838	AICc	1616.58	MPE	-9.7482

BIC	1633.17	MAPE	26.7190
		MASE	0.8324
		ACF1	0.00245

Uji Diagnostik

Uji diagnostik dilakukan untuk memastikan bahwa residual dari model ARIMA bersifat acak (*white noise*) dan memenuhi asumsi kestasioneran. Berdasarkan hasil plot residual dan ACF residual, tidak ditemukan pola yang sistematis atau autokorelasi signifikan pada lag mana pun. Hal ini menunjukkan bahwa model telah mampu menangkap pola utama dalam data dengan baik. Selain itu, hasil *Ljung-Box Test* menghasilkan nilai *p-value* > 0,05, yang berarti tidak ada autokorelasi yang tersisa pada residual model. Dengan demikian, model ARIMA yang digunakan dapat dinyatakan valid dan layak digunakan untuk melakukan peramalan nilai PM2.5 di periode berikutnya.

1. Heteroskedastisitas / Uji ARCH

Uji heteroskedastisitas menggunakan ARCH LM-Test pada residual model ARIMA (**Tabel 3.**) menghasilkan nilai Chi-Squared = 10.857, df = 12, dan *p-value* = 0.5412. Karena *p-value* > 0.05, maka gagal menolak H_0 , yang berarti tidak terdapat efek ARCH atau tanda-tanda heteroskedastisitas pada residual. Dengan demikian, varians residual dapat dianggap konstan (homoskedastis), sehingga model ARIMA yang dibangun memenuhi asumsi kestabilan varians dan layak digunakan untuk peramalan.

Tabel 3. Hasil Uji ARCH-LM pada Residual Model ARIMA

Statistik Uji	Nilai
Chi-squared	10.857
df	12
p-value	0.5412

2. Uji residual menggunakan Ljung-Box dan ACF residual

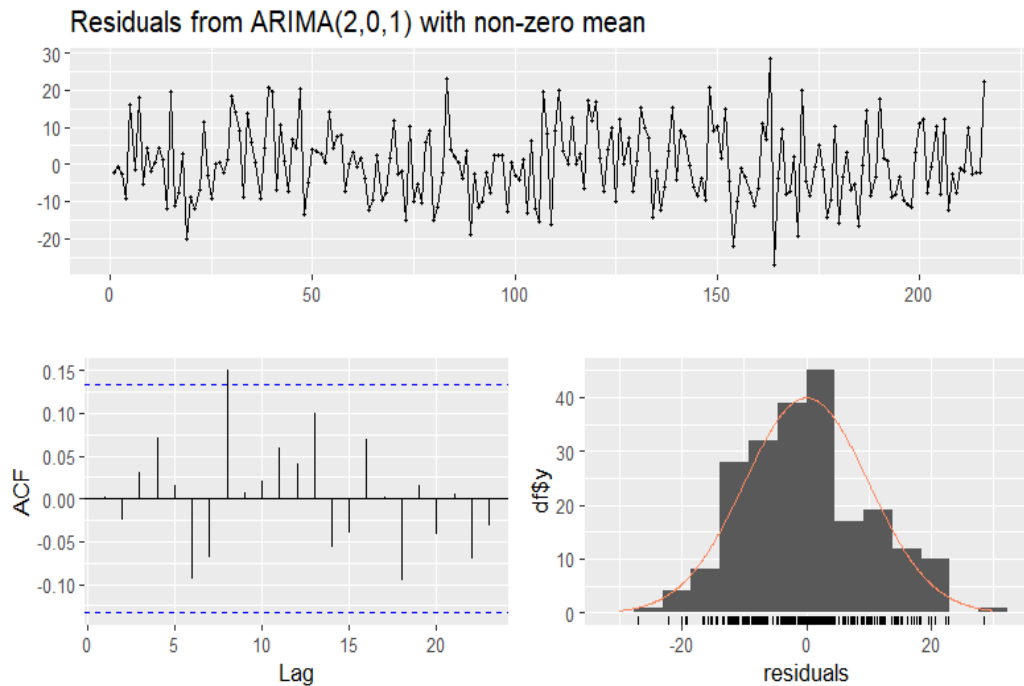
Hasil diagnostik residual menunjukkan nilai Ljung-Box = 9.771, df = 7, dan *p-value* = 0.2019 pada uji Ljung-Box. Karena *p-value* > 0.05, maka gagal menolak H_0 , yang berarti tidak terdapat autokorelasi signifikan pada residual model ARIMA. Dengan demikian, residual dapat dianggap sebagai *white noise*, sehingga model ARIMA sudah cukup baik dalam menangkap pola data dan layak digunakan untuk proses peramalan.

Tabel 4. Hasil Uji Ljung-Box pada Residual Model ARIMA

Statistik Uji	Nilai
Ljung-Box	9.771
df	7
p-value	0.2019
df Model	3
Total lag	10

Pada (**Gambar 7.**) Plot diagnostik residual dari model ARIMA(2,0,1) menunjukkan bahwa residual menyebar secara acak di sekitar nilai nol dan tidak membentuk pola tertentu, sehingga mengindikasikan bahwa model sudah mampu menangkap struktur data dengan baik. Plot ACF

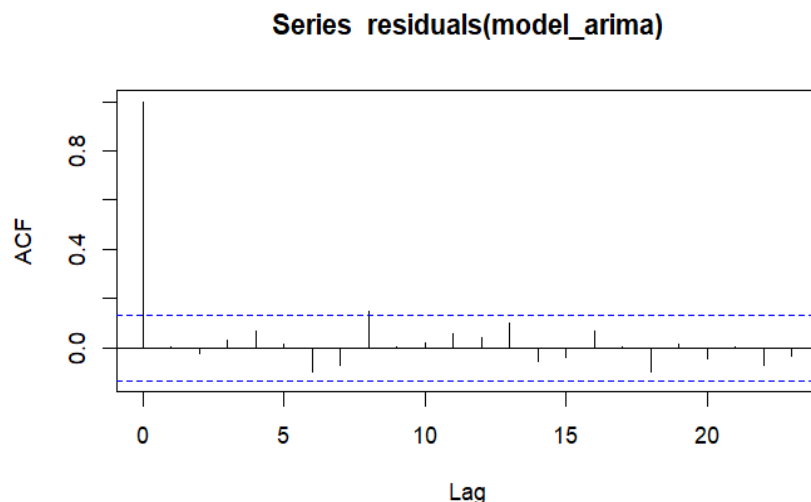
residual memperlihatkan bahwa seluruh lag berada dalam batas signifikansi (garis biru putus-putus), menegaskan bahwa tidak terdapat autokorelasi yang berarti. Selain itu, histogram residual yang mengikuti kurva normal menunjukkan bahwa distribusi residual mendekati normalitas. Secara keseluruhan, hasil ini mengonfirmasi bahwa model ARIMA yang dipilih telah memenuhi asumsi diagnostik model time series dan layak digunakan untuk proses peramalan.



Gambar 7. Plot diagnostik residual ARIMA

3. Autokorelasi residual

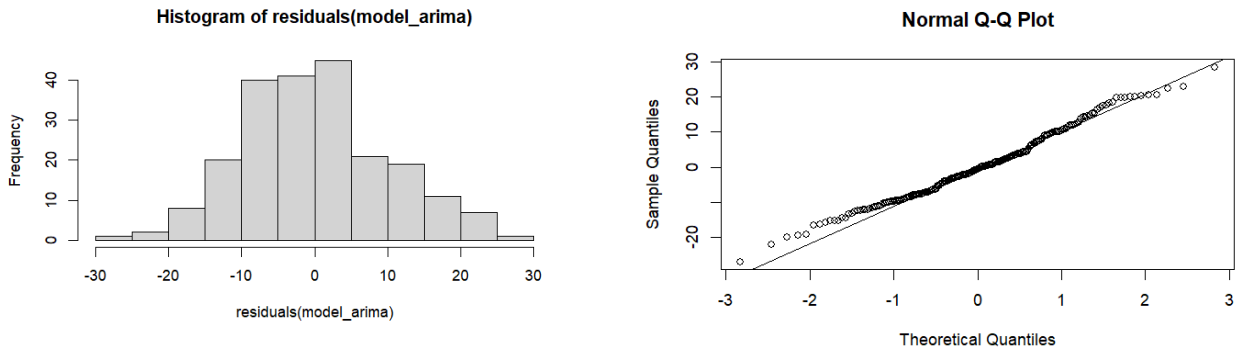
Berdasarkan (**Gambar 8.**) Plot ACF residual menunjukkan bahwa seluruh batang (lag residual) berada di dalam batas signifikansi (garis biru putus-putus), kecuali lag 0 yang memang selalu bernilai 1. Hal ini menunjukkan bahwa tidak terdapat autokorelasi yang signifikan pada residual model ARIMA, sehingga residual dapat dianggap sebagai deret acak (white noise). Dengan tidak adanya autokorelasi yang tersisa, berarti model ARIMA telah berhasil menangkap pola dalam data secara memadai dan tidak memerlukan penambahan komponen AR, MA, atau differencing lebih lanjut. Kesimpulannya, model yang digunakan sudah memenuhi salah satu asumsi penting pada analisis time series, yaitu tidak adanya autokorelasi pada residual.



Gambar 8. Plot ACF residual

4. Normalitas

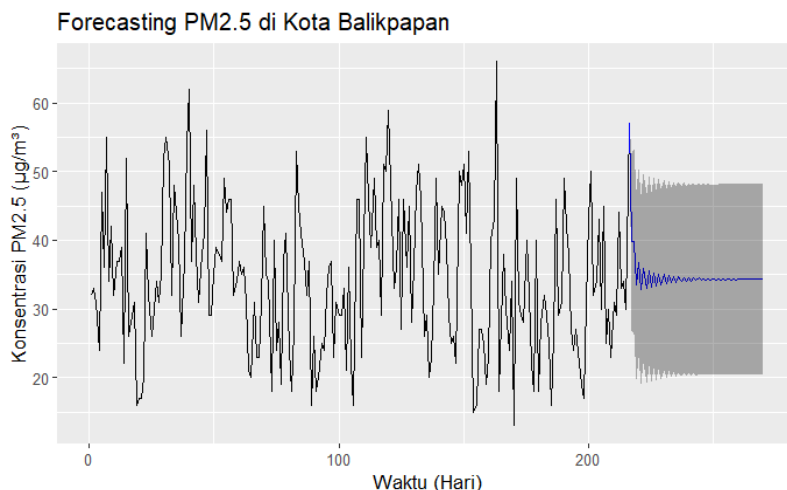
Histogram residual menunjukkan pola distribusi yang menyerupai bentuk lonceng (bell-shaped), di mana sebagian besar nilai residual berkumpul di sekitar nol dan menyebar simetris ke kanan dan kiri. Hal ini mengindikasikan bahwa residual cenderung mengikuti distribusi normal. Selain itu, Q-Q plot memperlihatkan bahwa titik-titik residual sebagian besar berada di sepanjang garis diagonal, yang menunjukkan kesesuaian dengan distribusi normal teoritis. Meskipun terdapat sedikit deviasi pada bagian ekor kiri dan kanan, penyimpangan tersebut masih dalam batas yang dapat diterima untuk model deret waktu. Secara keseluruhan, hasil ini menunjukkan bahwa asumsi normalitas residual telah terpenuhi, sehingga model ARIMA dapat dikatakan valid dan layak digunakan untuk proses peramalan.



Gambar 9. Histogram residual dan Q-Q plot

Forecasting

Grafik forecasting pada (**Gambar 10.**) menunjukkan hasil forecasting konsentrasi PM2.5 di Kota Balikpapan menggunakan model ARIMA, yang menampilkan dua bagian utama, yaitu data aktual pada periode training (garis hitam) dan hasil peramalan pada periode testing (garis biru). Hasil peramalan memperlihatkan bahwa nilai PM2.5 bergerak menuju rata-rata jangka panjangnya, yang merupakan karakteristik umum model ARIMA ketika pola musiman tidak dominan. Area bayangan berwarna abu-abu menggambarkan interval kepercayaan (confidence interval) yang menunjukkan tingkat ketidakpastian prediksi, di mana rentang tersebut semakin melebar seiring bertambahnya horizon waktu. Hal ini mengindikasikan bahwa prediksi jangka panjang memiliki tingkat ketidakpastian yang lebih tinggi dibanding jangka pendek. Secara keseluruhan, model ARIMA mampu mengikuti pola historis data PM2.5 dan menghasilkan prediksi yang stabil, sehingga dapat dijadikan dasar dalam pemantauan dan pengambilan keputusan terkait kualitas udara di masa mendatang.



Gambar 10. Grafik forecasting

Evaluasi Model

Hasil evaluasi menunjukkan bahwa model ARIMA (**Tabel 5.**) memiliki kinerja prediksi yang cukup baik. Nilai RMSE sebesar 11.52 dan MAE sebesar 9.30 mengindikasikan bahwa rata-rata kesalahan prediksi model berada di kisaran 9–12 $\mu\text{g}/\text{m}^3$. Nilai MAPE sebesar 26.81% menunjukkan tingkat kesalahan persentase yang termasuk kategori moderat, sehingga model masih cukup mampu menggambarkan pola PM2.5 meskipun terdapat variabilitas data yang tinggi. Yang paling penting, nilai MASE sebesar 0.43, yang berada di bawah angka 1, menandakan bahwa model ARIMA memiliki performa lebih baik dibandingkan metode naive forecasting. Dengan demikian, model ARIMA dapat dianggap efektif dan reliabel dalam memprediksi konsentrasi PM2.5 pada data uji.

Tabel 5. Hasil Evaluasi Model ARIMA

Ukuran Error	Nilai
RMSE	11.517
MAE	9.3
MAPE	26.8%
MASE	0.4305

IMPLIKASI KEBIJAKAN

Temuan penelitian mengenai pola fluktuasi dan hasil peramalan konsentrasi PM2.5 di Kota Balikpapan memberikan dasar penting bagi pemerintah daerah dalam merumuskan kebijakan pengendalian kualitas udara. Karena hasil analisis menunjukkan adanya peningkatan konsentrasi PM2.5 pada periode tertentu, maka diperlukan strategi mitigasi yang bersifat adaptif terhadap dinamika musiman. Pemerintah dapat memanfaatkan hasil prediksi sebagai dasar penyusunan sistem peringatan dini kualitas udara dan pengaturan aktivitas industri maupun transportasi yang berpotensi meningkatkan beban polusi pada periode risiko tinggi.

Kebijakan pengendalian kualitas udara harus selaras dengan regulasi nasional, seperti Peraturan Pemerintah Nomor 22 Tahun 2021 tentang Penyelenggaraan Perlindungan dan Pengelolaan Lingkungan Hidup, yang menetapkan baku mutu udara ambien, serta Permen LHK Nomor 14 Tahun 2020 tentang Indeks Standar Pencemar Udara (ISPU) yang menjadi acuan penilaian mutu udara untuk perlindungan kesehatan masyarakat. Selain itu, PP Nomor 41 Tahun 1999 tentang Pengendalian Pencemaran Udara menegaskan kewajiban pemerintah daerah untuk melakukan pengendalian sumber pencemar melalui penegakan regulasi dan implementasi teknologi pengendalian emisi.

Untuk meningkatkan ketepatan kebijakan, diperlukan perluasan jaringan pemantauan kualitas udara secara real-time terutama pada kawasan industri, pelabuhan, dan pusat kegiatan ekonomi. Pemerintah daerah perlu memperkuat kolaborasi dengan Dinas Lingkungan Hidup, Dinas Kesehatan, institusi pemantau cuaca, dan akademisi untuk mendukung perencanaan berbasis data. Literasi publik terkait risiko kesehatan akibat paparan PM2.5 juga perlu ditingkatkan melalui kampanye kesehatan dan penyampaian informasi ISPU yang mudah diakses oleh masyarakat, mengingat penelitian menunjukkan bahwa paparan PM2.5 dalam jangka panjang meningkatkan risiko penyakit pernapasan dan kardiovaskular. Dengan demikian, hasil penelitian ini dapat menjadi rujukan bagi pemerintah dalam mengembangkan kebijakan pengendalian pencemaran udara yang berbasis bukti ilmiah (evidence-based policy), responsif terhadap kondisi aktual, serta berfokus pada perlindungan kesehatan masyarakat dan keberlanjutan lingkungan Kota Balikpapan. Integrasi data peramalan ke dalam perencanaan pembangunan dan tata ruang menjadi langkah strategis untuk mewujudkan kualitas udara yang lebih baik dan aman bagi generasi sekarang maupun mendatang.

KESIMPULAN

Penelitian ini bertujuan untuk menganalisis pola historis dan melakukan peramalan konsentrasi polutan udara PM_{2.5} di Kota Balikpapan menggunakan metode deret waktu ARIMA. Berdasarkan hasil eksplorasi data, diketahui bahwa konsentrasi PM_{2.5} selama periode 1 Januari hingga 30 September 2025 memiliki fluktuasi yang cukup tinggi dengan rata-rata sebesar 34,84 $\mu\text{g}/\text{m}^3$, nilai minimum 13 $\mu\text{g}/\text{m}^3$, dan maksimum 70 $\mu\text{g}/\text{m}^3$. Pola data menunjukkan adanya variasi harian tanpa komponen musiman yang dominan, sehingga sesuai dianalisis menggunakan model ARIMA.

Hasil pengujian stasioneritas menggunakan Augmented Dickey-Fuller (ADF) menunjukkan bahwa data telah bersifat stasioner, sehingga proses differencing tidak diperlukan. Melalui identifikasi model menggunakan analisis ACF dan PACF, serta pemilihan model berdasarkan nilai AIC dan BIC, diperoleh model terbaik yaitu ARIMA(2,0,1). Model ini memberikan performa yang baik dengan nilai RMSE sebesar 9.96, MAE 7.96, dan MAPE 26.72% pada data training, serta RMSE 11.52 dan MAE 9.30 pada data testing. Uji diagnostik residual melalui Ljung-Box Test menunjukkan bahwa residual telah mengikuti pola white noise dan tidak terdapat autokorelasi yang signifikan sehingga model dinyatakan valid untuk peramalan.

Hasil forecasting memperlihatkan bahwa nilai prediksi PM_{2.5} mengikuti pola pergerakan historis serta cenderung mendekati nilai rata-rata jangka panjangnya, dengan tingkat ketidakpastian yang meningkat seiring bertambahnya horizon waktu. Grafik prediksi menunjukkan bahwa model ARIMA mampu merepresentasikan pola data dan memberikan hasil estimasi yang stabil untuk periode mendatang. Secara keseluruhan, penelitian ini berkontribusi dalam menyediakan informasi ilmiah mengenai kondisi kualitas udara di Kota Balikpapan dan membuktikan bahwa model ARIMA dapat digunakan sebagai alat prediksi jangka pendek untuk pemantauan polusi udara. Temuan ini dapat menjadi dasar bagi pemerintah daerah dalam merencanakan strategi mitigasi pencemaran udara dan mendukung kebijakan lingkungan yang berkelanjutan.

DAFTAR PUSTAKA

- Aladađ, C.H. et al. (2021). Forecasting of Particulate Matter with a Hybrid ARIMA Model. <https://www.sciencedirect.com/science/article/abs/pii/S2212095521001607>
- Brockwell, P. J., & Davis, R. A. (2016). Introduction to time series and forecasting (3rd ed.). Springer.
- Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and practice (3rd ed.). OTexts.
- Kumar, R., & Goyal, P. (2020). Forecasting of air quality in Delhi using ARIMA model. Environmental Monitoring and Assessment, 192(2), 1–14.
- Kurniawan, A. (2024). Predictive Performance Evaluation of ARIMA and Hybrid Models on PM_{2.5} Data in Urban Areas. <https://join.if.uinsgd.ac.id/index.php/join/article/view/1318>
- Putra, R.A. (2024). Prediksi Konsentrasi PM_{2.5} Menggunakan Model ARIMA-ANN di DKI Jakarta.
- Rita, R., Lestiani, D. D., Panjaitan, E. H., Santoso, M., & Yulinawati, H. (2016). Kualitas Udara (Pm10 Dan Pm2. 5) Untuk Melengkapi Kajian Indeks Kualitas Lingkungan Hidup. *Ecolab*, 1(1), 1-7.
- Shumway, R. H., & Stoffer, D. S. (2017). Time series analysis and its applications: With R examples (4th ed.). Springer.
- World Air Quality Index Project. (2025). Kualitas udara di Balikpapan (Stasiun Sepinggan Baru). <https://aqicn.org/station/indonesia-sepinggan-baru-balikpapan-sepinggan/id/>
- World Health Organization (2021). *WHO Global Air Quality Guidelines: Particulate Matter (PM_{2.5} and PM₁₀)*. [WHO global air quality guidelines: particulate matter \(PM_{2.5} and PM₁₀\), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide](https://www.who.int/publications/m/item/who-global-air-quality-guidelines-particulate-matter-pm2.5-and-pm10)
- Zhang, Y., Li, C., Wang, X., & Zhang, J. (2018). Time series forecasting of air quality using ARIMA models. Atmospheric Pollution Research, 9(4), 708–716.

Peraturan/Undang- Undang

Kementerian Lingkungan Hidup dan Kehutanan. (2020). Peraturan Menteri Lingkungan Hidup dan Kehutanan Nomor 14 Tahun 2020 tentang Indeks Standar Pencemar Udara (ISPU).

Pemerintah Republik Indonesia. (1999). *Peraturan Pemerintah Republik Indonesia Nomor 41 Tahun 1999 tentang Pengendalian Pencemaran Udara.*

Pemerintah Republik Indonesia. (2021). Peraturan Pemerintah Republik Indonesia Nomor 22 Tahun 2021 tentang Penyelenggaraan Perlindungan dan Pengelolaan Lingkungan Hidup.

Siregar, S., et al. (2024). Association between long-term PM_{2.5} exposure and mortality: Evidence from Sumatra Island, Indonesia. *Environmental Monitoring and Assessment.*

