

Forecasting PM_{2.5} Concentration in Medan City Using the ARIMAX Method with Meteorological Factors as Exogenous Variables

Fauzan Arrahman, Tessy Octavia Mukhti*, Dony Permana, Fenni Kurnia Mutiya

¹Departemen Statistika, Universitas Negeri Padang, Indonesia

*Corresponding author: tessyoctaviam@fmipa.unp.ac.id

Submitted : 17 Oktober 2025

Revised : 22 Oktober 2025

Accepted : 24 Oktober 2025

ABSTRACT

Particulate Matter 2.5 (PM_{2.5}) is a fine particle measuring less than 2.5 micrometers which is dangerous for human health because it can penetrate the respiratory system and cause cardiovascular disorders. High PM_{2.5} concentrations reflect a decline in air quality, so forecasting efforts are needed to support pollution control and environmental policies. This study aims to forecast daily PM_{2.5} concentrations in Medan City using the Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) method by considering meteorological factors as exogenous variables. The data used consist of PM_{2.5} concentrations and average temperature, humidity, rainfall, and wind speed data for the period from June 1, 2024 to June 10, 2025. The analysis results show that the best model is ARIMAX (4,1,0) with exogenous variables of average temperature and rainfall, where temperature has a positive effect and rainfall has a negative effect on PM_{2.5}. This model meets the assumptions of white noise and residual normality, with a MAPE value of 20.635%, indicating a fairly good level of forecasting accuracy. The forecasting results show PM_{2.5} concentrations in the range of 19–26 µg/m³ with a downward trend at the end of June 2025, indicating improved air quality in Medan City. Thus, the ARIMAX method with meteorological factors is considered effective in modeling and forecasting PM_{2.5} dynamics in urban areas.

Keywords: ARIMA, ARIMAX, Medan, Meteorologi, PM_{2.5}.



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Menurut World Health Organization (WHO), Sejak tahun 2006, polusi udara telah diakui sebagai ancaman lingkungan terbesar terhadap kesehatan manusia karena kontribusinya yang signifikan terhadap peningkatan beban penyakit secara global (WHO, 2021). Menurut Anandari dkk. (2024), polusi udara telah menjadi masalah lingkungan dengan jangkauan dan dampak yang sangat luas terhadap berbagai aspek kehidupan. Kondisi tersebut menunjukkan pentingnya pemantauan kualitas udara untuk menilai sejauh mana pencemaran udara berdampak terhadap kesehatan masyarakat.

Salah satu alat ukur yang digunakan secara global adalah Air Quality Index (AQI). AQI terdiri dari beberapa parameter polutan yang memengaruhinya, di mana *United States Environmental Protection Agency* (US EPA) menetapkan AQI, terdiri dari enam jenis polutan utama yaitu *Carbon Monoxide* (CO), *Ozone* (O₃), *Nitrogen Dioxide* (NO₂), *Sulfur Dioxide* (SO₂), dan *Particulate Matter* (PM₁₀ dan PM_{2.5}) (US EPA, 2024). Berdasarkan dari polutan-polutan yang memengaruhi AQI tersebut, secara global PM_{2.5} merupakan faktor risiko utama bagi kondisi pernapasan dan kardiovaskular, yang berkontribusi terhadap kematian dini dan morbiditas (Yu dkk, 2023).

Menurut IQAir (2024), PM_{2.5} merupakan partikel yang sangat halus dengan ukuran diameter ≤ 2,5 mikrometer, dilaporkan dalam satuan mikrogram per meter kubik (µg/m³). Di mana 1 mikrometer (1 µm) setara 0,001 milimeter atau sama dengan 10⁻⁶ meter. Ukuran yang sangat kecil ini menjadikan PM_{2.5} sebagai salah satu dari enam polutan utama yang diakui dan dipantau secara global karena dampaknya yang paling signifikan terhadap kesehatan manusia. Paparan PM_{2.5} baik jangka pendek maupun jangka panjang, dapat menimbulkan risiko berbagai penyakit pada manusia bahkan pada konsentrasi yang sangat rendah (Sangkham dkk, 2024). Oleh karena itu, PM_{2.5} secara luas diklasifikasikan sebagai "Pembunuh Tak Terlihat" di antara polutan udara lainnya (Basith dkk, 2022).

Kota Medan merupakan kota terbesar ketiga di Indonesia setelah Jakarta dan Surabaya (Yudha & Susilawati, 2025). Status ini ditunjukkan oleh jumlah penduduk yang melebihi 2,4 juta jiwa (BPS Kota Medan, 2024) dan posisinya sebagai pusat utama kawasan metropolitan Greater Medan (Mebidangro) (BPS Sumatera Utara, 2024). Selain itu, Medan juga memiliki peran strategis sebagai pusat ekonomi, perdagangan, dan pemerintahan di wilayah

Sumatera Utara. Sebagai kota besar dengan aktivitas penduduk yang tinggi, Medan kerap mengalami permasalahan polusi udara (Mustaqiman dkk, 2025).

Salah satu isu lingkungan penting adalah tingginya konsentrasi $PM_{2.5}$ di atmosfer. Konsentrasi ini tidak hanya dipengaruhi oleh sumber emisi, tetapi juga sangat ditentukan oleh kondisi meteorologi seperti suhu, kelembaban, curah hujan, dan angin. Faktor meteorologi dipilih dalam penelitian ini karena bersifat lebih dinamis secara harian, mudah diakses dalam bentuk data deret waktu, dan secara langsung memengaruhi pergerakan serta konsentrasi polutan di atmosfer. Berbeda dengan faktor emisi seperti industri atau transportasi yang relatif stabil per hari, sehingga variasinya sulit ditangkap dalam analisis jangka pendek (Liu dkk, 2020), variabel meteorologi justru memberikan gambaran yang lebih relevan terhadap fluktuasi konsentrasi $PM_{2.5}$ (Qonita dkk, 2025). Dengan demikian, analisis kondisi atmosfer menjadi penting untuk memahami dinamika kualitas udara di Medan.

Meningkatnya perhatian terhadap dampak kesehatan akibat polusi udara, khususnya paparan jangka panjang terhadap $PM_{2.5}$, mendorong perlunya sistem peringatan dini dan strategi pengendalian kualitas udara. Salah satu upaya yang dapat dilakukan adalah peramalan konsentrasi $PM_{2.5}$, guna mendukung langkah mitigasi seperti pembatasan aktivitas luar ruang, pengaturan lalu lintas, atau penyampaian imbauan kesehatan kepada masyarakat.

Melihat adanya pengaruh yang signifikan dari variabel meteorologi terhadap fluktuasi konsentrasi $PM_{2.5}$, maka diperlukan metode peramalan yang tidak hanya mengandalkan data historis $PM_{2.5}$, tetapi juga mempertimbangkan faktor-faktor tersebut. Peramalan yang komprehensif ini sangat penting untuk mendukung pengambilan keputusan yang cepat dan berbasis data dalam pengendalian polusi udara serta mitigasi dampak kesehatan.

Salah satu model yang umum digunakan dalam analisis deret waktu adalah *Autoregressive Integrated Moving Average* (ARIMA). Metode ARIMA pertama kali muncul pada abad ke-20 namun diformulasikan ulang oleh Box-Jenkins pada tahun 1970 sehingga metode ini juga dikenal dengan ARIMA Box-Jenkins (Khoiri, 2023). Metode ini bekerja berdasarkan pola historis dari variabel yang diamati. Namun, ARIMA memiliki keterbatasan karena tidak dapat mengakomodasi pengaruh variabel luar atau eksogen. Untuk mengatasi hal ini, dikembangkan model *Autoregressive Integrated Moving Average with Exogenous Variables* (ARIMAX) yang memungkinkan integrasi antara variabel utama dan variabel luar yang relevan, seperti faktor meteorologi.

Beberapa penelitian sebelumnya mendukung keunggulan ARIMAX dibandingkan dengan metode lainnya. Misalnya, Anggraeni dkk. (2017) menemukan ARIMAX lebih akurat dibanding VAR dalam prediksi harga komoditas yang dipengaruhi variabel eksogen. Hasil penelitian menunjukkan bahwa model ARIMAX dapat memprediksi harga konsumen padi dengan MAPE 0,15%. Hal ini lebih baik 15,27% dibandingkan model VAR. Muzakki dkk, (2024) menunjukkan bahwa peramalan dengan metode ARIMA menghasilkan MAPE 21,04 sedangkan peramalan dengan menggunakan metode ARIMAX menghasilkan MAPE sebesar 20,19. Penambahan variabel meteorologi ke dalam model ARIMAX meningkatkan akurasi prediksi konsentrasi $PM_{2.5}$ dibandingkan ARIMA biasa.

Dengan demikian, model ARIMAX menjadi model yang lebih unggul dalam memprediksi konsentrasi $PM_{2.5}$ yang dipengaruhi oleh kondisi eksternal yaitu faktor meteorologi.

II. METODE PENELITIAN

A. Jenis Penelitian dan Sumber Data

Jenis penelitian yang digunakan ialah penelitian terapan. Penelitian yang dilakukan adalah Peramalan Konsentrasi $PM_{2.5}$ di Kota Medan Menggunakan Metode ARIMAX dengan Faktor Meteorologi sebagai Variabel Eksogen. Data yang digunakan dalam penelitian ini diperoleh dari website <https://aqicn.org/here/id/> dan <https://dataonline.bmkg.go.id/>. Data tersebut terdiri dari data *Air Quality Indeks* (AQI $PM_{2.5}$) dan data faktor meteorologi yang diambil harian, mulai tanggal 1 Juni 2024 hingga 10 Juni 2025, dengan total sebanyak 375 hari pengamatan. Selanjutnya, nilai AQI $PM_{2.5}$ dikonversi menjadi konsentrasi $PM_{2.5}$ ($\mu g/m^3$) menggunakan rumus interpolasi linier berdasarkan standar dari US EPA.

B. Metode ARIMAX

Penelitian ini menggunakan metode *Autoregressive Integrated Moving Average with Exogenous Variables* (ARIMAX). ARIMAX yaitu model ARIMA yang diperbarui dengan menambah variabel eksogen (Muhaimin, 2024). Jika model ARIMA hanya memanfaatkan informasi masa lalu dari variabel endogen (Y_t) dan nilai residual masa lalu (ϵ_t), maka model ARIMAX juga memperhitungkan pengaruh dari satu atau lebih variabel eksogen (X_t) terhadap nilai Y_t .

1. Model *Autoregressive Integrated Moving Average* (ARIMA)

Model ARIMA adalah suatu model deret waktu yang mengombinasikan komponen *Autoregressive* (AR) dan *Moving Average* (MA) dengan parameter. Pada model ini, data deret waktu yang belum stasioner di *differencing* terlebih dahulu sebanyak d kali. Berikut bentuk umum dari ARIMA:

$$\phi(B)(1 - B)^d Z_t = \theta(B)\epsilon_t \quad (1)$$

dengan (Z_t) merupakan nilai pengamatan pada waktu ke- t , $\varphi(B)$ menggambarkan pengaruh nilai sebelumnya terhadap Z_t , sedangkan $(1 - B)^d$ menunjukkan bahwa data telah didiferensiasi supaya menjadi stasioner. Lalu, $\theta(B)$ dan ε_t menunjukkan bahwa error masa lalu juga ikut memengaruhi nilai Z_t sekarang. Jadi, Z_t terbentuk dari gabungan nilai masa lalu dan error masa lalu.

2. Model *Autoregressive Integrated Moving Average with Exogenous Variables* (ARIMAX)

Model ARIMAX yaitu model ARIMA yang diperbarui dengan menambah variabel eksogen (Muhaimin, 2024). Jika model ARIMA hanya memanfaatkan informasi masa lalu dari variabel endogen (Z_t) dan residual masa lalu (ε_t), maka model ARIMAX juga memperhitungkan pengaruh dari satu atau lebih variabel independen (X_t) pada waktu yang sama terhadap nilai Z_t . Secara umum, bentuk model ARIMAX sebagai berikut :

$$(1 - B)^d \phi(B) Z_t = \mu + \theta(B) \varepsilon_t + \alpha_1 X_{1t} + \alpha_2 X_{2t} + \dots + \alpha_k X_{kt} \quad (2)$$

dengan $(1 - B)^d \phi(B) Z_t$ menunjukkan bahwa nilai Z_t dipengaruhi oleh nilai historis atau masa lalunya setelah proses diferensiasi. Namun, selain komponen error $\theta(B) \varepsilon_t$ yang merepresentasikan pengaruh gangguan acak dari masa lalu, terdapat pula tambahan μ sebagai konstanta dan $\alpha_1 X_{1t} + \alpha_2 X_{2t} + \dots + \alpha_k X_{kt}$ yang merepresentasikan pengaruh variabel eksogen terhadap Z_t .

C. Langkah Analisis

Analisis data yang dilakukan menggunakan software R studio. Berikut adalah tahapan analisis (ARIMAX) :

1. Melakukan input data konsentrasi PM_{2.5} dan data faktor meteorologi di Kota Medan dari periode 1 Juni 2024 sampai 10 Juni 2025.
2. Melakukan eksplorasi data.
3. Melakukan imputasi missing value menggunakan Kalman Filter dengan bantuan *Software R Studio*.
4. Melakukan pembagian untuk data menjadi data latih (90%) dan data uji (10%)
5. Melakukan uji kestasioneran data :
 - a. Stasioner dalam ragam
Transformasi data menggunakan transformasi Box-Cox apabila varians tidak konstan menggunakan persamaan berikut :

$$T(Z_t) = Z_t^{(\lambda)} = \begin{cases} \frac{z_t^\lambda - 1}{\lambda}, \lambda \neq 0 \\ \ln Z_t, \lambda = 0 \end{cases} \quad (3)$$

- b. Stasioner dalam rata-rata
Uji stasioneritas dalam rata-rata pada data menggunakan uji *Augmented Dickey-Fuller* (ADF) dengan hipotesis awal yaitu terdapat unit root dalam data atau data tidak stasioner, menggunakan persamaan berikut:

$$T = \frac{\hat{\phi}}{S_{\hat{\phi}}} \quad (4)$$

dengan $\hat{\phi}$ yaitu nilai duga parameter *Autoregressive*, dan $S_{\hat{\phi}}$ standar errornya.

Kriteria Pengujian:

Jika $p\text{-value} > (\alpha = 0,05)$ maka gagal tolak H_0 artinya data tidak stasioner. Jika $p\text{-value} \leq (\alpha = 0,05)$ maka tolak H_0 artinya data stasioner

Jika data masih belum stasioner dalam rata-rata, maka dilakukan proses *differencing* pada data tersebut (Rahmayani, 2013). *Differencing* bertujuan untuk menghilangkan tren atau pola jangka panjang pada data. Proses ini dilakukan dengan menghitung selisih antara nilai saat ini dengan nilai sebelumnya.

6. Menentukan ordo model ARIMA:

Model ARIMA memiliki parameter (p, d, q) , dimana p merupakan nilai AR, q merupakan nilai MA, dan d yaitu jumlah *differencing* pada data.

a. *Autocorrelation Function* (ACF),

ACF adalah korelasi antara dua variabel pada periode waktu (t) dengan periode waktu sebelumnya ($t - 1$) yang disusun dalam plot setiap lag. ACF digunakan untuk menentukan ordo MA (*Moving Average*) (Muhaimin, 2024). Persamaan ACF sebagai berikut :

$$\rho_k = \frac{\sum_{t=1}^{n-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z})}{\sum_{t=1}^{n-k} (Z_t - \bar{Z})^2} \quad (5)$$

dengan ρ_k : koefisien *autocorrelation* pada lag ke- k , Z_t : nilai observasi pada periode ke- t , sedangkan Z_{t+k} : nilai observasi pada periode ke- $t + k$. Sementara itu, \bar{Z} : rata-rata dari seluruh nilai observasi.

- b. *Partial Autocorrelation Function* (PACF)
PACF merupakan korelasi antara deret pengamatan yang mengukur keeratan antara pengamatan suatu deret waktu (Muhaimin, 2024). PACF digunakan untuk menentukan ordo AR. Persamaan PACF sebagai berikut :

$$\hat{\phi}_{kk} = \frac{\hat{\rho}_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} \hat{\rho}_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} \hat{\rho}_j} \quad (6)$$

dengan $\hat{\phi}_{kk}$: koefisien korelasi antara z_t dengan z_{t+k} , sedangkan $\hat{\rho}_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} \hat{\rho}_{k-j}$ menunjukkan selisih antara koefisien *autocorrelation* pada lag ke- k ($\hat{\rho}_k$) dengan kontribusi *autocorrelation* dari semua lag sebelumnya (1 sampai $k - 1$) yang telah dikalikan dengan koefisien *partial autocorrelation* pada lag tersebut. $1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} \hat{\rho}_j$ sebagai faktor penyesuaian untuk memperhitungkan pengaruh lag sebelumnya secara keseluruhan

- c. Nilai d dari jumlah *differencing* yang dilakukan.
7. Setelah ordo untuk model ARIMA ditetapkan, maka berikutnya adalah melakukan proses estimasi pada parameter model. Teknik yang dipakai pada tahap ini adalah *Maximum Likelihood Estimation* (MLE), yaitu suatu metode untuk memaksimalkan fungsi likelihood. Pendekatan MLE menjadi salah satu metode penting dalam penaksiran parameter pada inferensi statistik. Metode ini memanfaatkan informasi dari distribusi probabilitas, baik diskrit maupun kontinu, untuk memperoleh estimasi parameter yang paling tepat (Rahmayani, 2013).
8. Diagnostik Model dan Pemilihan Model ARIMA terbaik
Setelah proses estimasi selesai, tahap berikutnya adalah evaluasi model melalui diagnostic checking untuk mengecek bahwa model yang diperoleh tersebut telah memenuhi uji asumsi dasar dan mampu merepresentasikan data dengan baik. Pengujian yang dilakukan yaitu:
- a. Uji *white noise*, bertujuan untuk memeriksa apakah nilai residual dari suatu model bersifat acak (tidak berkorelasi). Salah satu pendekatan yang digunakan dalam proses ini adalah uji Ljung-Box dengan hipotesis awal yaitu residual tidak ada yang berkorelasi menggunakan persamaan berikut:

$$Q = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k} \quad (7)$$

dengan Q : statistik uji Ljung-Box, n : jumlah observasi, h : banyak lag yang di uji, $\hat{\rho}_k^2$: autokorelasi residual pada lag ke- k

Kriteria pengujian :

Jika $p\text{-value} > (a = 0,05)$ maka gagal tolak H_0 artinya residual *white noise*. Jika $p\text{-value} \leq (a = 0,05)$ maka tolak H_0 artinya residual tidak *white noise*

- b. Uji normalitas residual, dilakukan untuk menilai apakah nilai residual dari suatu model mengikuti pola distribusi normal atau tidak. Salah satu pendekatan yang digunakan dalam proses ini adalah uji Kolmogorov-Smirnov dengan hipotesis awal yaitu residual mengikuti distribusi normal. Secara matematis, besaran statistik pada uji ini dapat diperoleh melalui perhitungan menggunakan rumus berikut:

$$D = \max |F_0(x) - F_t(x)| \quad (8)$$

dengan D merupakan deviasi maksimum antara kedua distribusi tersebut, $F_0(x)$ merupakan fungsi distribusi kumulatif empiris dari data residual, dan $F_t(x)$ merupakan fungsi distribusi kumulatif teoretis dari distribusi normal

Kriteria pengujian:

Jika hasil uji menunjukkan $p\text{-value}$ melebihi 0,05, maka H_0 dapat diterima, yang berarti data residual memiliki sebaran normal. Namun, jika $p\text{-value}$ kurang dari atau sama dengan 0,05, maka H_0 ditolak, sehingga dapat disimpulkan bahwa residual tidak menyebar secara normal.

Setelah memastikan karakteristik residual, pemilihan model terbaik selanjutnya dilakukan dengan mempertimbangkan nilai kriteria informasi sebagai dasar evaluasi.

- 1) *Akaike Information Criterion* (AIC) :

$$AIC = -2\ln(L) + 2k \quad (9)$$

- 2) *Bayesian Information Criterion* (BIC) :

$$BIC = -2\ln(L) + k\ln(n) \quad (10)$$

Model dengan AIC dan BIC terkecil serta residual yang tidak berautokorelasi dan berdistribusi normal dianggap sebagai model terbaik (Box dkk, 2016).

9. Membangun model ARIMAX berdasarkan hasil identifikasi ordo ARIMA. Ordo ARIMAX didapatkan dari hasil pemilihan model terbaik ARIMA. Setelah model ARIMAX dibentuk, langkah selanjutnya adalah melakukan estimasi parameter. Proses ini dilakukan dengan metode maksimum likelihood (MLE), serupa dengan proses estimasi pada model ARIMA. Perbedaanannya terletak pada keberadaan variabel eksogen dalam model.
10. Melakukan diagnostik model dengan memeriksa residual dengan menggunakan persamaan (7) dan (8).
11. Memilih model terbaik berdasarkan nilai AIC dan BIC terkecil menggunakan persamaan (9) dan (10).
12. Melakukan evaluasi model dengan menghitung nilai kesalahan peramalan menggunakan Mean Absolute Percentage Error (MAPE) dengan persamaan berikut :

$$MAPE = \left(\frac{1}{N} \sum_{t=1}^N \frac{|A_t - F_t|}{A_t} \right) \times 100 \quad (11)$$

dengan A_t : nilai aktual, F_t : nilai prediksi, N : jumlah pengamatan

13. Melakukan peramalan konsentrasi PM_{2.5} menggunakan model ARIMAX yang telah dibangun, dengan menggunakan data variabel eksogen aktual.

III. HASIL DAN PEMBAHASAN

A. Eksplorasi Data

Eksplorasi data dilakukan untuk melihat dan memahami karakteristik dari data PM_{2.5} dan meteorologi di Kota Medan dengan menggunakan analisis statistik deskriptif. Hasil analisis statistik deskriptif untuk data PM_{2.5} disajikan pada tabel 1 berikut.

Tabel 1. Statistika Deskriptif PM_{2.5} Kota Medan

Statistik	Nilai
Jumlah Data (n)	375
Mean	29,23
Median	28,00
Minimum	6,00
Maximum	112,00

Dari tabel 1 dilihat bahwa jumlah data PM_{2.5} di Kota Medan yang digunakan adalah sebanyak 375 observasi, yaitu dari 1 Juni 2024 - 10 Juni 2025. Selama rentang waktu tersebut tercatat bahwa konsentrasi tertinggi yaitu sebesar 112,00 µg/m³ dan konsentrasi terendah yaitu sebesar 6,00 µg/m³. Adapun dapat dilihat bahwa rata-rata PM_{2.5} di Kota Medan dari 1 Juni 2024 - 10 Juni 2025 adalah 29,23 µg/m³ dan dengan nilai median yaitu 28,00 µg/m³.

Selain data PM_{2.5}, dilakukan eksplorasi terhadap data variabel meteorologi yang digunakan sebagai variabel eksogen dalam model ARIMAX, yaitu suhu rata-rata, kelembapan rata-rata, curah hujan, dan kecepatan angin rata-rata. Hasil analisis statistik deskriptif untuk data meteorologi disajikan dalam tabel 2 berikut.

Tabel 2. Statistika Deskriptif Meteorologi di Kota Medan

Variabel	Mean	Median	Minimum	Maximum
Suhu rata-rata	28,03	27,90	24,30	32,70
Kelembapan rata-rata	82,47	83,00	60,00	97,00
Curah Hujan	11,02	1,50	0,00	136,00
Kecepatan Angin rata-rata	1,27	1,00	0,00	4,00

Berdasarkan tabel 2, dapat dilihat bahwa suhu rata-rata di Kota Medan selama periode pengamatan berada pada kisaran 24,30°C hingga 32,70°C dengan rata-rata sebesar 28,03°C dan median sebesar 27,90°C. Kelembapan udara berkisar antara 60% hingga 97%, dengan rata-rata sebesar 82,47% dan median sebesar 83%. Curah hujan menunjukkan nilai yang cukup bervariasi, dengan rata-rata sebesar 11,02 mm, median sebesar 1,50 mm, serta nilai maksimum mencapai 136,00 mm. Nilai median yang jauh lebih rendah dibandingkan rata-rata mengindikasikan bahwa sebagian besar hari memiliki curah hujan rendah atau bahkan tidak hujan, namun terdapat beberapa hari dengan curah hujan ekstrem tinggi. Sementara itu, kecepatan angin rata-rata berada pada kisaran 0,00 m/s hingga 4,00 m/s, dengan rata-rata sebesar 1,27 m/s.

B. Pembagian Data

Data $PM_{2.5}$ dan Meteorologi dikelompokkan menjadi dua set yaitu data latihan (training) yang mencakup 337 titik data (90%) dan data uji (testing) sebanyak 38 titik data (10%).

C. Identifikasi Model ARIMA

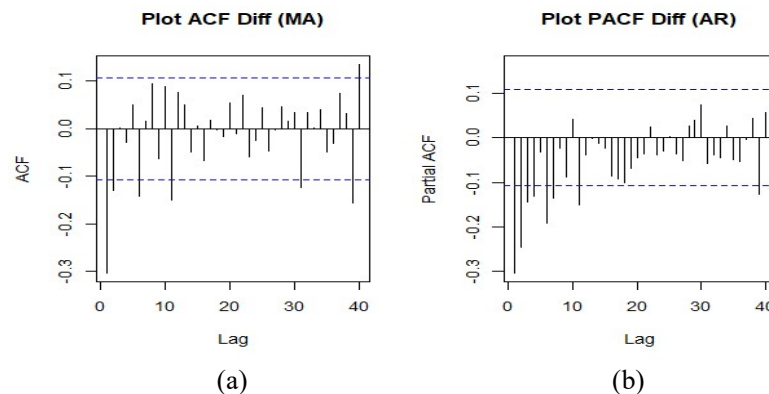
Pada pemodelan ARIMA tahap awal yang dilakukan adalah dengan mengecek apakah data yang digunakan sudah stasioner atau belum. Untuk menguji kestasioneran dalam ragam, digunakan uji Box-Cox sebagaimana ditunjukkan pada persamaan (3). Hasil pengujian tersebut disajikan pada tabel 3.

Tabel 3. Hasil Pengujian Stasioner Terhadap Ragam

$PM_{2.5}$	Lambda	Kesimpulan
Data Sebelum Transformasi	0.2073266	Belum Stasioner
Data Setelah Transformasi	1.218732	Stasioner

Dapat dilihat pada tabel 3 bahwa, data setelah dilakukan transformasi sudah stasioner dengan nilai lambda sudah mendekati 1. Selanjutnya adalah melakukan pengujian data untuk menentukan apakah stasioner dalam rata-rata dengan menggunakan persamaan (4) menggunakan tes ADF, menunjukkan nilai statistik uji *Dickey-Fuller* sebesar -4.8973 dengan lag optimal 6 serta *p-value* yaitu 0,01 setelah dilakukan *differencing* satu kali, yang berarti bahwa data $PM_{2.5}$ sudah stasioner dalam rata-rata. Setelah uji ADF, identifikasi orde p dan q dilakukan dengan mengamati pola pada plot ACF dan PACF

Data yang telah stasioner terhadap ragam dan rata-rata, selanjutnya akan dilakukan identifikasi ordo model ARIMA dengan menggunakan plot PACF dan ACF. Berikut plot ACF dan PACF data yang sudah di *differencing* sebanyak satu kali disajikan pada gambar 1 berikut.



Gambar 1. (a) Plot ACF dan (b) Plot PACF

Berdasarkan gambar 1, dapat dilihat bahwa pada plot ACF lag yang keluar melewati garis signifikan adalah lag 2, sehingga ordo q pada model ARIMA terdiri dari 0, 1 dan 2. Pada plot PACF lag yang melewati garis signifikan terdiri dari lag 1, lag 2, lag 3, dan lag 4. Sehingga, orde p pada model ARIMA terdiri dari 0, 1, 2, 3, dan 4. Pada proses pengujian stasioner terhadap rata-rata, data dilakukan *differencing* sebanyak 1 kali. Sehingga orde d pada model ARIMA adalah 1. Berdasarkan hasil identifikasi tersebut maka terbentuk 6 model ARIMA yang terdiri dari ARIMA (4,1,0), ARIMA (3,1,2), ARIMA (3,1,1), ARIMA (3,1,0), ARIMA (2,1,0), ARIMA (1,1,0).

D. Estimasi Parameter ARIMA

Tabel 4. Model ARIMA

Model ARIMA	Parameter	Koefisien	<i>p-value</i>	Keterangan
ARIMA (4,1,0)	AR-1	-0,443749	3,851e-16	Signifikan
	AR-2	-0,352159	1,896e-09	Signifikan
	AR-3	-0,208671	0,0003699	Signifikan
	AR-4	-0,140108	0,0100469	Signifikan
ARIMA (3,1,2)	AR-1	-0,179404	0,5745	Tidak Signifikan
	AR-2	0,265844	0,0723	Tidak Signifikan
	AR-3	0,075487	0,2006	Tidak Signifikan
	MA-1	-0,357661	0,2588	Tidak Signifikan
	MA-2	-0,574050	0,0548	Tidak Signifikan

ARIMA (3,1,1)	AR-1	0,421156	1,604e-12	Signifikan
	AR-2	0,015285	0,8017	Tidak Signifikan
	AR-3	0,061018	0,2963	Tidak Signifikan
	MA-1	-0,961058	< 2,2e-16	Signifikan
ARIMA (3,1,0)	AR-1	-0,421729	8,547e-15	Signifikan
	AR-2	-0,308269	5,183e-08	Signifikan
	AR-3	-0,148584	0,006208	Signifikan
ARIMA (2,1,0)	AR-1	-0,384487	4,960e-13	Signifikan
	AR-2	-0,250407	2,417e-06	Signifikan
ARIMA (1,1,0)	AR-1	-0,306490	4,299e-09	Signifikan

Berdasarkan hasil estimasi parameter pada tabel 4, model yang parameternya signifikan yaitu ARIMA (4,1,0), ARIMA (3,1,0), ARIMA (2,1,0) dan ARIMA (1,1,0) sehingga layak untuk digunakan. Oleh karena itu, model tersebut dipilih sebagai kandidat terbaik untuk tahap diagnostik selanjutnya.

E. Diagnostik Model ARIMA

Setelah didapatkan kandidat model yang layak digunakan, maka selanjutnya dilakukan uji *white noise* dengan menggunakan persamaan (7). Hasil pengujian *white noise* ditunjukkan pada tabel 5 berikut.

Tabel 5. Uji *White Noise* ARIMA

Model ARIMA	<i>p-value</i>	Keterangan
ARIMA (4,1,0)	0,9098	Terpenuhi
ARIMA (3,1,0)	0,6998	Terpenuhi
ARIMA (2,1,0)	0,4893	Terpenuhi
ARIMA (1,1,0)	0,1575	Terpenuhi

Berdasarkan tabel 5, dapat dilihat bahwa semua model yang di uji memiliki nilai *p-value* > 0,05, sehingga dapat disimpulkan bahwa residual dari masing-masing model bersifat *white noise*. Setelah uji *white noise* diperoleh, maka dilakukan uji normalitas residual dengan uji Kolmogorov-Smirnov menggunakan persamaan (8). Hasil pengujian normalitas residual ditunjukkan pada tabel 6 berikut.

Tabel 6. Uji Normalitas Residual ARIMA

Model ARIMA	<i>p-value</i>	Keterangan
ARIMA (4,1,0)	0,147	Terpenuhi
ARIMA (3,1,0)	0,1625	Terpenuhi
ARIMA (2,1,0)	0,04887	Tidak Terpenuhi
ARIMA (1,1,0)	0,2252	Terpenuhi

Berdasarkan tabel 6, sebagian besar model memiliki *p-value* lebih besar dari 0,05, sehingga memenuhi uji normalitas residual. Namun, model ARIMA (2,1,0) tidak memenuhi asumsi normalitas karena memiliki *p-value* < 0,05.

F. Pemilihan Model Terbaik ARIMA

Setelah diagnostik pada model maka didapatkan 3 model yang memenuhi asumsi *white noise* dan normalitas. Untuk memilih model ARIMA terbaik, selanjutnya dilakukan dengan mempertimbangkan nilai AIC dan BIC dengan menggunakan persamaan (9) dan (10). Berikut hasil nilai AIC dan BIC pada tabel 7 .

Tabel 7. AIC dan BIC

Model ARIMA	AIC	BIC
ARIMA (4,1,0)	618,648	638,270
ARIMA (3,1,0)	623,580	639,277
ARIMA (1,1,0)	651,017	658,865

Berdasarkan tabel 7, didapatkan model ARIMA terbaik yaitu ARIMA (4,1,0) dengan nilai AIC dan BIC yang terkecil.

G. Identifikasi dan Estimasi Parameter Model ARIMAX

Setelah didapatkan model ARIMA terbaik, maka dilakukan estimasi parameter untuk model ARIMAX dengan menambahkan variabel suhu rata-rata, kelembapan rata-rata, curah hujan, dan kecepatan angin rata-rata. Berikut hasil estimasi parameter ARIMAX pada tabel 8.

Tabel 8. Model ARIMAX

Model ARIMA	Parameter	Koefisien	<i>p-value</i>	Keterangan
ARIMAX (4,1,0)	AR-1	-0.4645372	< 2.2e-16	Signifikan
	AR-2	-0.4053816	2.573e-11	Signifikan
	AR-3	-0.2386304	5.860e-05	Signifikan
	AR-4	-0.1399890	0.0105545	Tidak Signifikan
	Suhu rata-rata	0.2355986	0.0001903	Signifikan
	Kelembapan rata-rata	0.0168269	0.2278976	Tidak Signifikan
	Curah hujan	-0.0073514	1.327e-07	Signifikan
	Kecepatan angin rata-rata	-0.0199919	0.6946478	Tidak Signifikan
ARIMAX (4,1,0)	AR-1	-0.4671736	< 2.2e-16	Signifikan
	AR-2	-0.4113049	1.094e-11	Signifikan
	AR-3	-0.2343376	7.539e-05	Signifikan
	AR-4	-0.1351070	0.01328	Signifikan
	Suhu rata-rata	0.1676328	9.808e-07	Signifikan
	Curah hujan	-0.0072932	1.633e-07	Signifikan

Berdasarkan hasil estimasi parameter pada tabel 8, model yang parameternya signifikan yaitu ARIMA(4,1,0) dengan variabel suhu rata-rata dan curah hujan sehingga layak untuk digunakan. Oleh karena itu, model tersebut dipilih terbaik untuk tahap diagnostik selanjutnya.

H. Diagnostik Model ARIMAX

Setelah didapatkan model yang layak digunakan, maka selanjutnya dilakukan uji *white noise* dengan menggunakan persamaan (7). Dari hasil pengujian yang dilakukan, diapatkan hasil nilai *p-value* yaitu sebesar 0,9123 yang berarti model ARIMA(4,1,0) dengan variabel suhu rata-rata dan curah hujan memenuhi uji *white noise*

Setelah pengujian *white noise*. Dilakukan uji normalitas residual pada model dengan uji Kolmogorov-Smirnov menggunakan persamaan (8) dan diapatkan nilai *p-value* yaitu 0,2588 yang berarti uji normalitas residual juga terpenuhi

model ARIMAX (4,1,0) dengan variabel eksogen suhu rata-rata dan curah hujan memenuhi asumsi normalitas, maka dapat disimpulkan bahwa model ARIMAX (4,1,0) dengan variabel eksogen suhu rata-rata dan curah hujan merupakan model terbaik.

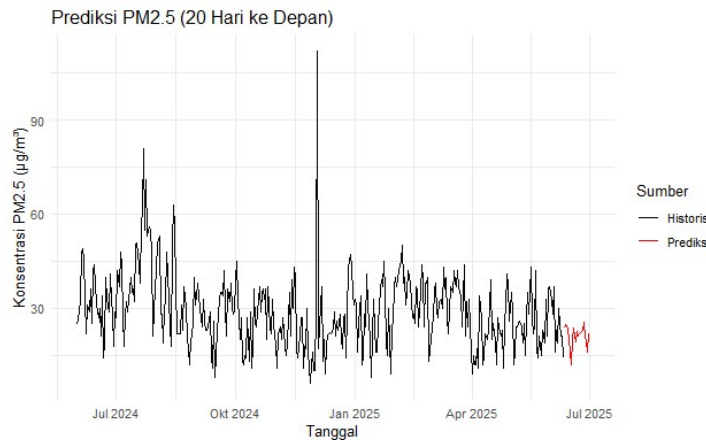
I. Evaluasi Model

Untuk melihat seberapa baik model ARIMAX yang telah didapatkan maka, dilakukan evaluasi model dengan menggunakan *Mean Absolute Percentage Error* (MAPE) dengan persamaan (11). Didapatkan nilai MAPE nya yaitu 20,635. Dimana dapat disimpulkan bahwa model ARIMAX (4,1,0) dengan variabel eksogen suhu rata-rata dan curah hujan cukup baik dalam meramalkan data PM_{2.5}

J. Peramalan**Tabel 9.** Hasil Peramalan PM_{2.5} Periode 11 Juni – 30 Juni 2025

Periode	Hasil Peramalan (µg/m ³)	Periode	Hasil Peramalan (µg/m ³)
11/06/2025	23,590	21/06/2025	22,631
12/06/2025	24,760	22/06/2025	21,228
13/06/2025	24,590	23/06/2025	21,860
14/06/2025	23,480	24/06/2025	22,610
15/06/2025	19,876	25/06/2025	22,897
16/06/2025	11,938	26/06/2025	25,549
17/06/2025	20,345	27/06/2025	20,674
18/06/2025	23,559	28/06/2025	19,827
19/06/2025	21,679	29/06/2025	15,941
20/06/2025	19,370	30/06/2025	21,916

Berdasarkan tabel 9, hasil peramalan menunjukkan bahwa konsentrasi $PM_{2.5}$ pada periode 11–30 Juni 2025 berfluktuasi di kisaran 19–26 $\mu g/m^3$. Nilai tertinggi diperkirakan terjadi pada 12 Juni 2025 (24,76 $\mu g/m^3$) dan 28 Juni 2025 (25,83 $\mu g/m^3$), sedangkan nilai terendah pada 29 Juni 2025 (15,94 $\mu g/m^3$). Secara grafis dapat dilihat pada gambar 2 berikut.



Gambar 2. Hasil Peramalan $PM_{2.5}$ Kota Medan Periode 11 Juni – 30 Juni 2025

Gambar 2, menunjukkan hasil peramalan konsentrasi $PM_{2.5}$ selama 20 hari ke depan. Garis hitam merepresentasikan data historis, sedangkan garis merah menunjukkan nilai prediksi dari model ARIMAX. Terlihat bahwa pola prediksi mengikuti tren data historis dengan fluktuasi yang relatif stabil tanpa lonjakan ekstrem. Hal ini menandakan bahwa model mampu menggambarkan pola perubahan $PM_{2.5}$ secara konsisten, dengan kecenderungan sedikit menurun di akhir periode, yang menunjukkan potensi perbaikan kualitas udara pada akhir Juni 2025

IV. KESIMPULAN

Berdasarkan hasil analisis data konsentrasi $PM_{2.5}$ di Kota Medan periode 1 Juni 2024 hingga 10 Juni 2025 menggunakan metode ARIMAX, diperoleh model terbaik yaitu ARIMAX (4,1,0) dengan variabel eksogen suhu rata-rata dan curah hujan. Kedua variabel tersebut terbukti signifikan memengaruhi konsentrasi $PM_{2.5}$, di mana suhu berpengaruh positif, sedangkan curah hujan berpengaruh negatif terhadap kadar $PM_{2.5}$. Model ARIMAX (4,1,0) yang dihasilkan telah memenuhi asumsi *white noise* dan normalitas residual, serta memiliki nilai MAPE sebesar 20,635%, yang menunjukkan bahwa model memiliki tingkat ketepatan peramalan yang cukup baik. Hasil peramalan menunjukkan bahwa konsentrasi $PM_{2.5}$ selama 11 Juni – 30 Juni 2025 berfluktuasi di kisaran 19–26 $\mu g/m^3$, dengan tren yang cenderung menurun di akhir periode, mengindikasikan perbaikan kualitas udara di Kota Medan menjelang akhir Juni 2025. Dengan demikian, metode ARIMAX dengan faktor meteorologi terbukti efektif untuk meramalkan dan memahami dinamika $PM_{2.5}$, serta dapat digunakan sebagai dasar dalam pengendalian kualitas udara di Kota Medan.

DAFTAR PUSTAKA

- Anandari, A. A., Wadjdi, A. F., & Harsono, G. (2024). *Dampak polusi udara terhadap kesehatan dan kesiapan pertahanan negara di Provinsi DKI Jakarta*. *Journal on Education*, 6(2), 10868–10884. <https://doi.org/10.31004/joe.v6i2.4880>
- Anggraeni, W., Andri, K. B., Sumaryanto, & Mahananto, F. (2017). *The Performance of ARIMAX Model and Vector Autoregressive (VAR) Model in Forecasting Strategic Commodity Price in Indonesia*. *Procedia Computer Science*, 124, 189–196. <https://doi.org/10.1016/j.procs.2017.12.146>
- Basith, S., Manavalan, B., Shin, T. H., Park, C. B., Lee, W. S., Kim, J., & Lee, G. (2022). *The impact of fine particulate matter 2.5 on the cardiovascular system: A review of the invisible killer*. *Nanomaterials*, 12 (15), 2656. <https://doi.org/10.3390/nano12152656>
- Badan Pusat Statistik Kota Medan. (2024). *Kota Medan dalam Angka 2025 (Medan Municipality in Figures 2024)*. Medan: Badan Pusat Statistik Kota Medan. <https://medankota.bps.go.id>
- IQAir. (2024). “2024 World Air Quality Report: Region & City PM_{2.5} Ranking”, Swiss Technology Company, AirVisual Platform, Switzerland. <https://www.iqair.com/id/world-air-quality-report>
- Khoiri, H. A. (2023, November 27). *Analisis Deret Waktu Univariat: Teori dan Pengolahan Data*.
- Liu, Z., Ciais, P., Deng, Z., Lei, R., Davis, S. J., Feng, S., et al. (2020). “Near-real-time monitoring of global CO₂ emissions reveals the effects of the COVID-19 pandemic”, *Nature Communications*, Vol. 11, No. 5172, hal. 1–12. <https://doi.org/10.1038/s41467-020-18922-7>
- Mustaqiman, A. N., Irawandani, T. D., Prayogo, W., & Suhardono, S. (2025). *Modeling air pollutant dispersion in urban areas: A HYSPLIT-based analysis of PM_{2.5} dynamics in Medan, Indonesia*. *Jurnal Lingkungan dan Rekayasa Kota*, X(1), 12004–12009. <https://doi.org/10.12345/hysplit.medan.2025>
- Muhaimin, Pandu Herdi. 2024. *Prediksi Hasil Produksi Tanaman Jagung Menggunakan Metode ARIMA Exogenous*. Skripsi. Malang: Universitas Islam Negeri Maulana Malik Ibrahim
- Muzakki, N. F., Putri, A. Z., Maruli, S., & Kartiasih, F. (2024). *Peramalan indeks kualitas udara dengan memanfaatkan beberapa faktor meteorologi menggunakan metode ARIMAX (Studi kasus: Kota Jakarta Pusat)*. *Jurnal JTik (Jurnal Teknologi Informasi dan Komunikasi)*, 8(3), 569–586. <https://doi.org/10.35870/jtik.v8i3.2012>
- Qonitaa, I. R., Santikayasa, I. P., & Turyanti, A. (2025). *Analisis pengaruh faktor meteorologi terhadap fluktuasi konsentrasi PM_{2.5} di Kota Pontianak*. *Positron*, 15 (1), 24–33. <https://doi.org/10.26418/positron.v15i1.91566>
- Rahmayani, Lina. (2013). *Model ARIMAX dan SARIMAX untuk Meramalkan Data Curah Hujan*. Skripsi. Jurusan Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Gunung Djati, Bandung
- Sangkham, S., Phairuang, W., Sherchan, S. P., Pansakun, N., Munkong, N., Sarndhong, K., Islam, M. A., & Sakunkoo, P. (2024). *An update on adverse health effects from exposure to PM_{2.5}*. *Environmental Advances*, 18, 100603. <https://doi.org/10.1016/j.envadv.2024.100603>
- The World Air Quality Index Project. (2024). “Air Quality Data Platform (AQICN): Real-time Air Quality Index (AQI) Data”, Available at: <https://aqicn.org/data-platform/register/>
- U.S. Environmental Protection Agency. (2018). “Technical Assistance Document for the Reporting of Daily Air Quality – The Air Quality Index (AQI)”, EPA-454/B-18-007, Office of Air Quality Planning and Standards, Air Quality Assessment Division, Research Triangle Park, NC.
- World Health Organization. (2021). *WHO global air quality guidelines: Particulate matter (PM_{2.5} and PM₁₀), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. World Health Organization. <https://apps.who.int/iris/handle/10665/345329>

- Yudha, M., & Susilawati. (2025). *Analisis kondisi polusi udara berdasarkan perubahan waktu menggunakan IoT dan logika fuzzy: Solusi mencegah dampak polusi terhadap kesehatan*. *INCODING: Journal of Informatics and Computer Science Engineering*, 3(2), 1–10.
- Yu, W., Ye, T., Zhang, Y., Xu, R., Lei, Y., Chen, Z., Yang, Z., Zhang, Y., Song, J., Yue, X., Li, S., & Guo, Y. (2023). *Global estimates of daily ambient fine particulate matter concentrations and unequal spatiotemporal distribution of population exposure: A machine learning modelling study*. *The Lancet Planetary Health*, 7(3), e209–e218. [https://doi.org/10.1016/S25425196\(23\)00008-6](https://doi.org/10.1016/S25425196(23)00008-6)