
Pemodelan Indeks Kualitas Udara PM_{2.5} di Kemayoran, Jakarta, dengan Faktor Meteorologi Menggunakan ARFIMAX

Oktavia Laras Dianingati¹, Mahmudi², Dhea Urfina Zulkifli³

oktavialaras55@gmail.com¹, mahmudi@uinjkt.ac.id², dhea.urfina@uinjkt.ac.id³

^{1,2,3} UIN Syarif Hidayatullah Jakarta

Informasi Artikel	Abstrak
Diterima : 23 Jan 2025 Direvisi : 21 Feb 2025 Disetujui : 28 Feb 2025	Kualitas udara yang buruk yang berdampak signifikan terhadap kesehatan akibat partikel halus PM _{2.5} menjadi salah satu masalah serius di kawasan perkotaan seperti Kemayoran, Jakarta. Penelitian ini bertujuan memprediksi Indeks Kualitas Udara polutan PM _{2.5} di Kemayoran, Jakarta, dengan model ARFIMAX yang akan dibandingkan dengan model ARFIMA. Pemodelan ARFIMAX melibatkan faktor meteorologi sebagai variabel eksogen. Hasil penelitian menunjukkan bahwa model ARFIMAX(1,0,33,1) dengan variabel eksogen yang signifikan, yaitu suhu minimum, suhu rata-rata, dan arah angin pada kecepatan maksimum memberikan akurasi prediksi lebih baik dengan nilai Mean Absolute Percentage Error (MAPE) sebesar 23,69%, dibandingkan ARFIMA dengan nilai MAPE 25,76%. Penurunan nilai MAPE ini mengindikasikan bahwa penambahan variabel eksogen dalam model dapat meningkatkan akurasi prediksi kualitas udara.
Kata Kunci Indeks Kualitas Udara (IKU) PM _{2.5} , ARFIMA, ARFIMAX, variabel eksogen, prediksi	

Keywords	Abstract
<i>Air Quality Index (AQI) PM_{2.5}, ARFIMA, ARFIMAX, exogenous variables, forecast</i>	<i>Poor air quality that significantly impacts on health due to fine particles PM_{2.5} has become one of the serious problems in urban areas such as Kemayoran, Jakarta. This study aims to predict the Air Quality Index of PM_{2.5} pollutant in Kemayoran, Jakarta, with the ARFIMAX model which will be compared with the ARFIMA model. The ARFIMAX modeling involves meteorological factors as exogenous variables. The research results showed that the ARFIMAX(1,0,33,1) model with significant exogenous variables, namely minimum temperature, average temperature, and wind direction at maximum speed provides better prediction accuracy with a Mean Absolute Percentage Error (MAPE) value of 23.69%, compared to ARFIMA with a MAPE value of 25.76%. This decrease in MAPE value indicates that the addition of exogenous variables in the model can improve the accuracy of air quality forecast.</i>

A. Pendahuluan

Salah satu polutan udara yang berbahaya, *particulate matter* (PM), terutama PM_{2.5}, memiliki ukuran yang sangat kecil dengan diameter aerodinamis yang sama dengan atau kurang dari 2,5 μm . PM_{2.5} dapat dengan mudah masuk ke dalam saluran pernapasan dan berdampak pada kondisi kesehatan manusia seperti penyakit paru-paru dan kardiovaskular. Organisasi Kesehatan Dunia (WHO) menyatakan bahwa konsentrasi rata-rata tahunan polutan PM_{2.5} di udara tidak boleh melebihi 5 $\mu\text{g}/\text{m}^3$, dan rata-rata harian tidak boleh melebihi 15 $\mu\text{g}/\text{m}^3$ [1].

Konsentrasi PM_{2.5}, yang berdampak signifikan pada kesehatan masyarakat, masih menjadi perhatian utama di Jakarta terkait kualitas udara. Berdasarkan data pengamatan dari Stasiun Pemantauan Kualitas Udara BMKG Kemayoran dari 1 hingga 27 Agustus 2023, konsentrasi harian PM_{2.5} berada pada kategori sedang hingga tidak sehat. Menurut Fachri Radjab, Kepala Pusat Informasi Perubahan Iklim BMKG, konsentrasi maksimum PM_{2.5} tercatat sebesar 164,6 $\mu\text{g}/\text{m}^3$ pada 8 Agustus 2023 pukul 05.00 WIB, dan konsentrasi rata-rata PM_{2.5} mencapai 60,4 $\mu\text{g}/\text{m}^3$ dari bulan Agustus hingga 27 Agustus 2023. Kondisi ini merupakan peningkatan yang signifikan dibandingkan periode yang sama tahun 2022, ketika nilai konsentrasi rata-rata adalah 44,3 $\mu\text{g}/\text{m}^3$. Siklus harian menunjukkan bahwa konsentrasi PM_{2.5} cenderung lebih tinggi pada malam hingga pagi hari[2]. Kondisi ini menegaskan pentingnya pemantauan dan analisis lebih lanjut berkaitan dengan Indeks Kualitas Udara (IKU), khususnya PM_{2.5}.

Indeks Kualitas Udara (IKU), khususnya yang berfokus pada PM_{2.5}, telah menjadi perhatian dalam berbagai penelitian karena dampaknya terhadap kesehatan manusia. Studi terdahulu menunjukkan bahwa faktor-faktor meteorologi memainkan peranan penting dalam memengaruhi konsentrasi PM_{2.5}[3]. Berbagai metode statistik dan model prediktif telah digunakan untuk menganalisis dan memprediksi IKU, seperti penelitian menggunakan ARIMAX[3], Support Vector Regression (SVR)[4], SARIMA[5], dan *hybrid* model[6].

Salah satu penelitian terdahulu yang relevan dilakukan oleh Muzakki, dkk[3] yang menggunakan model ARIMAX untuk memprediksi PM_{2.5} di Jakarta Pusat dengan variabel eksogen yang signifikan, yaitu suhu rata-rata, kelembapan rata-rata, dan kecepatan angin rata-rata yang dibandingkan dengan model ARIMA. Penelitian ini menghasilkan akurasi dengan MAPE sebesar 21.046% untuk model ARIMA dan 20.190% untuk ARIMAX. Hal ini membuktikan bahwa penggunaan variabel eksogen dapat meningkatkan akurasi model. Namun, penelitian ini tidak mempertimbangkan arah angin pada kecepatan maksimum sebagai variabel eksogen, yang berpotensi memberikan pengaruh signifikan terhadap kualitas udara. Selain itu, penelitian ini belum memperhatikan adanya sifat ketergantungan jangka panjang pada data PM_{2.5}. Oleh karena itu, diperlukan penelitian lanjutan yang mengakomodasi sifat ketergantungan jangka panjang dan pelibatan variabel eksogen dengan penggunaan model ARFIMAX.

Pendekatan yang dapat menangani ketergantungan jangka panjang pada data runtun waktu adalah model ARFIMA (AutoRegressive Fractional Integrated Moving Average), yang telah digunakan pada penelitian sebelumnya terhadap suku bunga PUAB[7], kecepatan angin[8], dan *temperature*[9]. Selain ARFIMA, model

ARFIMAX, yang mengintegrasikan variabel eksogen, telah digunakan untuk meningkatkan akurasi analisis. Penelitian Degiannakis[10] menerapkan ARFIMAX dan ARFIMAX-TARCH dalam memprediksi volatilitas indeks saham CAC40 dan DAX30. Sementara Sarkar, dkk[11] menunjukkan keakuratan ARFIMAX dalam memodelkan harga kentang maksimum dan harga modal harian di pasar Arga, India, dengan variabel eksogen seperti kedatangan pasar. Selain itu, penelitian lain menemukan bahwa variabel eksogen meningkatkan akurasi model, terutama dalam peramalan curah hujan yang menggunakan SARIMAX dan hybrid SARIMAX-ANN[12]. Hal ini melatarbelakangi pelibatan variabel eksogen dalam pemodelan pada penelitian ini. Dengan demikian, pendekatan ARFIMAX yang melibatkan variabel eksogen diharapkan dapat meningkatkan akurasi dalam memprediksi $PM_{2.5}$ dibandingkan dengan ARFIMA.

Penelitian ini berfokus pada wilayah Kemayoran, Jakarta, yang merupakan lokasi Stasiun Pemantauan BMKG. Model ARFIMAX digunakan untuk memprediksi konsentrasi $PM_{2.5}$ dengan melibatkan faktor meteorologi sebagai variabel eksogen. Untuk mengetahui pengaruh variabel eksogen dalam model, hasil akurasi ARFIMAX akan dibandingkan dengan model ARFIMA. Penelitian ini bertujuan untuk menganalisis apakah pelibatan variabel eksogen meteorologi dapat meningkatkan akurasi prediksi $PM_{2.5}$ di Kemayoran, Jakarta. Penelitian ini juga diharapkan memberikan kontribusi bagi pengembangan kebijakan lingkungan dalam pengelolaan kualitas udara serta mitigasi dampak kesehatan akibat polusi udara.

B. Metode Penelitian

1. Data

Data yang digunakan pada penelitian ini adalah data harian indeks kualitas udara (IKU) $PM_{2.5}$ ($\mu\text{g}/\text{m}^3$) yang diperoleh dari World Air Quality Project Historical Air Quality Data Platform (aqicn.org), data harian suhu udara minimum ($^{\circ}\text{C}$), suhu udara maksimum ($^{\circ}\text{C}$), suhu udara rata-rata ($^{\circ}\text{C}$), kelembapan udara rata-rata (persen), lamanya penyinaran sinar matahari (jam), kecepatan angin maksimum (m/s), kecepatan angin rata-rata (m/s), dan arah angin saat kecepatan maksimum ($^{\circ}\text{C}$) yang diperoleh dari Stasiun Meteorologi Kemayoran dalam Data Online Pusat Database BMKG (dataonline.bmkg.go.id). Data yang digunakan pada penelitian merupakan data harian dari 1 Mei 2023 hingga 8 Juli 2024 dengan total 435 data. Data yang digunakan dibagi menjadi data *training* (80%) dari 1 Mei 2023 - 12 April 2024 untuk pembentukan model dan data *testing* (20%) dari 13 April 2024 - 8 Juli 2024 untuk memvalidasi model terpilih.

2. Model ARFIMA

Model ARFIMA (Autoregressive Fractionally Integrated Moving Average) dirancang untuk menganalisis data deret waktu yang memiliki sifat *long memory*. Fenomena *long memory* terjadi ketika autokorelasi data menurun secara perlahan seiring dengan peningkatan lag atau meluruh secara hiperbolik[13]. Model ARFIMA merupakan pengembangan dari ARIMA dengan menambahkan komponen differencing fraksional d . Perbedaan mendasar dari ARIMA adalah bahwa parameter d dalam ARFIMA tidak terbatas pada bilangan bulat, melainkan dapat berupa nilai riil[14]. Secara matematis, model ARFIMA dengan orde (p,d,q) dapat dinyatakan sebagai:

$$\Phi(L)(1-L)^d(y_t) = \Theta(L)e_t \quad (1)$$

dengan y_t adalah data deret waktu, $e_t \sim i.i.d(0, \sigma_\varepsilon^2)$ adalah proses *white noise*, $\Phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$, ϕ_1, \dots, ϕ_p koefisien autoregresif pada lag 1 hingga orde p , $\Theta(L) = 1 + \theta_1 L + \dots + \theta_q L^q$, $\theta_1, \dots, \theta_q$ koefisien moving average pada lag 1 hingga orde q , L adalah lag operator, dan d adalah operator *fractional difference* yang merepresentasikan memori jangka panjang[13],[11].

Operator *fractional difference* didefinisikan[15]:

$$\begin{aligned} \nabla^d &= (1 - L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-L)^k \\ &= 1 - dL - \frac{1}{2}d(1-d)L^2 - \frac{1}{6}d(1-d)(2-d)L^3 - \dots \end{aligned} \quad (2)$$

3. Model ARFIMAX

Model ARFIMAX (Autoregressive Fractionally Integrated Moving Average Model with Exogenous variable) merupakan pengembangan lebih lanjut dari model ARFIMA. Variabel eksogen ini ditambahkan ke dalam model ARFIMA untuk kinerja model yang lebih baik[11]. Degiannakis menerapkan model ARFIMAX[10], model ARFIMAX dengan k variabel eksogen dapat dinyatakan sebagai[10]:

$$\Phi(L)(1 - L)^d(y_t - x'_t \beta) = \Theta(L)e_t \quad (3)$$

dengan, y_t adalah data deret waktu variabel endogen/dependen, x_t vektor dari variabel eksogen pada waktu t , dan β vektor koefisien yang berhubungan dengan variabel eksogen, dan keterangan notasi lainnya tersebut sama seperti pada model ARFIMA[11].

4. Pengujian Diagnostik Model

Setelah nilai parameter model diestimasi, tahap selanjutnya pemeriksaan diagnostik model yaitu pengujian kenormalan pada residual, pengujian autokorelasi pada residual, dan pengujian heteroskedastisitas pada residual. Pengujian kenormalan pada residual menggunakan uji Kolmogorov-Smirnov. Jika $p\text{-value} >$ tingkat signifikan(α) maka residual berdistribusi normal. Pengujian autokorelasi residual menggunakan uji Ljung-Box. Jika $p\text{-value} >$ tingkat signifikan(α) maka tidak terdapat autokorelasi pada residual atau memenuhi asumsi *white noise*. Sedangkan, pengujian heteroskedastisitas menggunakan uji ARCH-LM. Jika $p\text{-value} >$ tingkat signifikan(α) maka tidak ada heteroskedastisitas dalam residual atau residual konstan.

5. Pemilihan Model Terbaik

Pemilihan model dilihat dari nilai *Akaike Info Criterion*(AIC) terkecil. AIC dinyatakan sebagai[16]:

$$AIC = -2\log\mathcal{L}(\hat{\theta}) + 2r \quad (4)$$

dengan $\mathcal{L}(\theta)$ adalah likelihood dari data, $\hat{\theta}$ adalah estimasi maksimum likelihood, dan r adalah jumlah parameter model yang diestimasi.

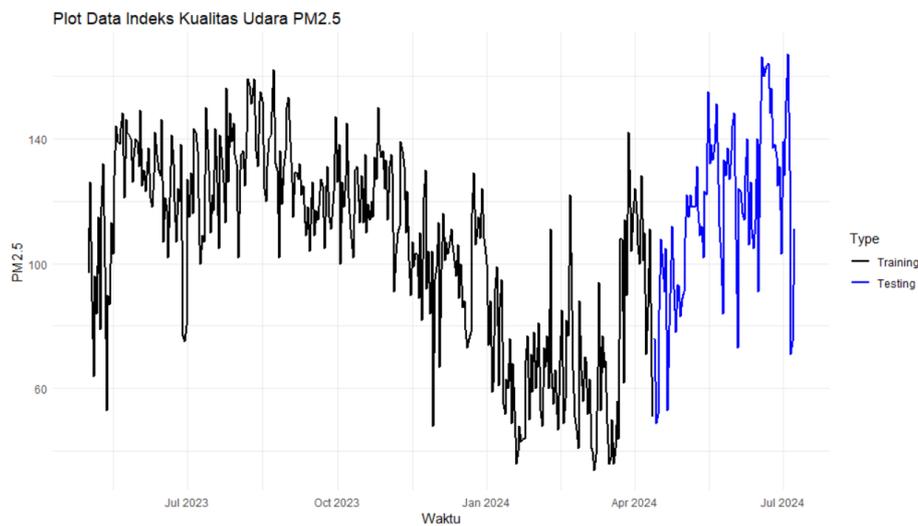
Nilai *Mean Absolute Percentage Error*(MAPE) merupakan salah satu cara untuk mengukur keakuratan peramalan atau akurasi model. Rumus MAPE dinyatakan sebagai[17]:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100\% \quad (5)$$

dengan y_t adalah nilai pengamatan pada waktu ke- t , \hat{y}_t adalah nilai hasil peramalan pada waktu ke- t , dan n adalah banyaknya data pengamatan.

C. Hasil dan Pembahasan
1. Analisis Deskripsi Data

Data harian indeks kualitas udara $PM_{2.5}(\mu g/m^3)$ sebagai variabel endogen dan data harian suhu udara minimum ($^{\circ}C$), suhu udara maksimum ($^{\circ}C$), suhu udara rata-rata ($^{\circ}C$), kelembapan udara rata-rata (persen), lamanya penyinaran sinar matahari (jam), kecepatan angin maksimum (m/s), kecepatan angin rata-rata (m/s), dan arah angin saat kecepatan maksimum ($^{\circ}C$) sebagai variabel eksogen. Sebelum pemodelan dilakukan, jika terdapat missing value permasalahan tersebut ditangani terlebih dahulu dengan menggunakan nilai rata-rata dari data yang ada. Berikut plot data indeks kualitas udara $PM_{2.5}$ yang digunakan dalam penelitian ini.



Gambar 1. Plot Data Indeks Kualitas Udara $PM_{2.5}$ di Kemayoran, Jakarta

Berdasarkan Gambar 1, grafik berwarna hitam merupakan data *training* yang berjumlah 348 data yang digunakan untuk pembentukan model, sedangkan grafik berwarna biru merupakan data *testing* yang berjumlah 87 data yang digunakan untuk memvalidasi model terpilih. Berikut statistik deskriptif data *training* yang disajikan pada Tabel 1.

Tabel 1. Statistik Deskriptif Data

Variabel	Nilai Minimum	Nilai Rata-rata	Nilai Maksimum	Standar Deviasi
Indeks kualitas udara $PM_{2.5}$	34	105.43	162	31.304
Suhu udara minimum	24	26.26	28.10	0.859
Suhu udara maksimum	27	32.88	36.80	1.285
Suhu udara rata-rata	26	29.09	31.8	0.931
Kelembapan udara rata-rata	56	74.72	94	6.50
Lamanya penyinaran sinar matahari	0	4.481	9.9	2.633
Kecepatan angin maksimum	0	4.853	11	1.849
Kecepatan angin rata-rata	0	1.118	3	0.598
Arah angin saat kecepatan maksimum	0	274.37	360	88.414

Data terkait indeks kualitas udara $PM_{2.5}$ menunjukkan bahwa nilai rata-rata mencapai 105,43, yang mencerminkan kualitas udara cenderung berada dalam kategori tidak sehat. Namun, kualitas udara tidak selalu berada dalam kategori tersebut, karena nilai minimum indeks kualitas udara $PM_{2.5}$ menunjukkan nilai sebesar 34, yang dikategorikan sebagai baik.

2. Pengujian Kestasioneran Data dan Identifikasi Long Memory

Sebelum melakukan pemodelan, data diperiksa apakah data sudah stasioner. Pengujian terhadap rata-rata menggunakan uji ADF, dengan $\alpha = 0.05$. Kriteria pengambilan keputusan apabila $p\text{-value} < \alpha$ maka tolak H_0 , sehingga dapat disimpulkan bahwa data stasioner. Berikut hasil pengujian ADF pada data variabel endogen dan variabel eksogen yang disajikan pada Tabel 2.

Tabel 2. Hasil Uji ADF

Variabel	Dickey-Fuller	$p\text{-value}$
Indeks kualitas udara $PM_{2.5}$	-3.6663	0.02691
Suhu udara minimum	-4.3207	0.01
Suhu udara maksimum	-4.9415	0.01
Suhu udara rata-rata	-4.3221	0.01
Kelembapan udara rata-rata	-4.2592	0.01
Lamanya penyinaran sinar matahari	-4.1805	0.01
Kecepatan angin maksimum	-4.7101	0.01
Kecepatan angin rata-rata	-5.2864	0.01
Arah angin saat kecepatan maksimum	-5.2740	0.01

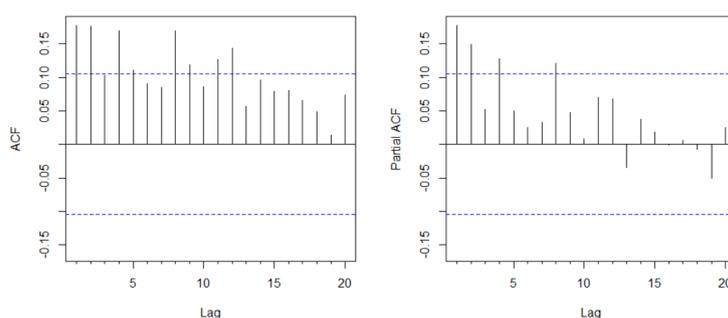
Berdasarkan Tabel 2 diperoleh nilai $p\text{-value}$ pada semua variabel < 0.05 , yang berarti bahwa data pada semua variabel sudah stasioner terhadap rata-rata.

Tahap selanjutnya yaitu melakukan identifikasi long memory pada data indeks kualitas udara $PM_{2.5}$ menggunakan uji statistik Hurst(H). Dari hasil perhitungan, diperoleh nilai $H = 0.83$ yang berada pada rentang $0.5 < H < 1$. Berdasarkan kriteria uji statistik Hurst[18] dapat disimpulkan bahwa terdapat ketergantungan jangka panjang pada data $PM_{2.5}$.

3. Identifikasi Model ARFIMA

Pada model ARFIMA penentuan parameter d dilakukan dengan menggunakan metode *Rescaled Range Statistic*(R/S). Nilai d dihitung dengan formula $d = H - 0.5$. Berdasarkan perhitungan sebelumnya diperoleh nilai $d = 0.33$. Hasil estimasi nilai d terletak antara $0 < d < 0.5$ yang menunjukkan data merupakan *stationary long memory model*[13].

Identifikasi model dilakukan dengan melihat plot ACF dan PACF dari data yang telah *didifferencing* dengan $d = 0.33$. Berikut disajikan plot ACF dan PACF dari data indeks kualitas udara $PM_{2.5}$ yang telah *didifferencing* pada Gambar 2.



Gambar 2. Plot ACF dan PACF Data didifferencing ($d = 0.33$)

Berdasarkan Gambar 2, plot ACF *cut-off* pada lag 1,2,4,5,8,9,11, dan 12, dan plot PACF *cut-off* pada lag 1,2,4, dan 8. Menurut prinsip Parsimoni, model yang lebih sederhana lebih disukai daripada model dengan banyak parameter [19]. Oleh karena itu, pada penelitian ini kandidat model dibatasi hingga lag 2. Kandidat model ARFIMA yang dianalisis lebih lanjut adalah ARFIMA(0,d,1), ARFIMA(0,d,2), ARFIMA(1,d,0), ARFIMA(1,d,1), ARFIMA(1,d,2), ARFIMA(2,d,0), ARFIMA(2,d,1), dan ARFIMA(2,d,2), dengan $d = 0.33$.

4. Estimasi parameter model ARFIMA

Tahapan selanjutnya adalah estimasi parameter model dan uji signifikansi parameter model. Uji signifikansi parameter dilakukan dengan taraf signifikansi ($\alpha = 0.05$). Parameter dikatakan signifikan apabila nilai $p\text{-value} < \alpha$. Berdasarkan kandidat model yang diperoleh pada tahap sebelumnya, berikut disajikan hasil estimasi dan uji signifikansi parameter dan modelnya.

Tabel 3. Estimasi Parameter ARFIMA $d = 0.33$

No	Model ARFIMA	Parameter	Estimasi Parameter	$p\text{-value}$	Signifikansi
1	(0,d,1)	θ_1	-0.1379	0.0030	Ya
2	(0,d,2)	θ_1	-0.1501	0.0056	Ya
		θ_2	-0.1352	0.0054	Ya
3	(1,d,0)	ϕ_1	0.1796	0.0007	Ya
4	(1,d,1)	ϕ_1	0.9535	<2.22e-16	Ya
		θ_1	0.8660	<2.22e-16	Ya
		ϕ_1	0.9611	<2.22e-16	Ya
5	(1,d,2)	θ_1	0.8550	<2.22e-16	Ya
		θ_2	0.0276	0.62894	Tidak Signifikan
6	(2,d,0)	ϕ_1	0.1526	0.0041	Ya
		ϕ_2	0.1541	0.0040	Ya
		ϕ_1	0.9979	<2.22e-16	Ya
7	(2,d,1)	ϕ_2	-0.0343	0.60763	Tidak Signifikan
		θ_1	0.8893	<2.22e-16	Ya
		ϕ_1	-0.2254	2.9003e-06	Ya
8	(2,d,2)	ϕ_2	0.7737	<2.22e-16	Ya
		θ_1	-0.3954	1.2081e-14	Ya
		θ_2	0.5979	<2.22e-16	Ya

Berdasarkan Tabel 3, terdapat enam model dengan seluruhnya parameter yang signifikan yaitu ARFIMA(0,0.33,1), ARFIMA(0,0.33,2), ARFIMA(1,0.33,0), ARFIMA(1,0.33,1), ARFIMA(2,0.33,0), dan ARFIMA(2,0.33,2). Keenam model tersebut akan dilakukan pengujian diagnostik model pada tahap selanjutnya.

5. Diagnostik Model ARFIMA

Model ARFIMA yang memenuhi kriteria parameter yang signifikan selanjutnya dilakukan pengujian diagnostik model. Pengujian diagnostik meliputi uji asumsi nilai residual berdistribusi normal dengan menggunakan uji Kolmogorov-Smirnov, uji autokorelasi terhadap residual model dengan uji L-jung Box, dan uji heteroskedastisitas dalam residual dengan uji ARCH-LM.

Tabel 4. Hasil Diagnostik Model ARFIMA dan Nilai AIC Model

Model ARFIMA	p-value			AIC
	Uji Kolmogorov-Smirnov	Uji L-jung Box	Uji ARCH-LM	
(0,0.33,1)	0.9128	0.699	0.2105	2012.964
(0,0.33,2)	0.7904	0.9296	0.2093	2007.347
(1,0.33,0)	0.8999	0.5959	0.2134	2010.246
(1,0.33,1)	0.8735	0.6712	0.3865	1993.458
(2,0.33,0)	0.9545	0.8857	0.3019	2004.071
(2,0.33,2)	0.8746	0.4041	0.3644	2004.003

Berdasarkan hasil diagnostik model yang disajikan dalam Tabel 4, dapat dilihat bahwa seluruh model ARFIMA yang diuji memiliki nilai *p-value* lebih besar dari tingkat signifikansi ($\alpha = 0.05$) pada masing-masing uji diagnostik. Hal ini menunjukkan bahwa semua model yang diuji memenuhi semua asumsi diagnostik model, yaitu residual berdistribusi normal, tidak terdapat autokorelasi pada residual, serta tidak ada heteroskedastisitas dalam residual. Dengan demikian, kandidat model pada Tabel 4 yang diuji telah memenuhi uji diagnostik model.

Selanjutnya, pemilihan model terbaik dilakukan dengan membandingkan nilai *Akaike Information Criterion* (AIC) dari setiap model yang memenuhi uji diagnostik model. Nilai AIC digunakan untuk menentukan model yang paling optimal. Berdasarkan nilai AIC, model ARFIMA(1,0.33,1) memiliki nilai AIC terkecil yaitu 1993.456 dibandingkan model ARFIMA lainnya. Dengan demikian, model ARFIMA(1,0.33,1) dipilih untuk digunakan dalam pengembangan model lanjutan, yaitu ARFIMAX.

6. Model ARFIMAX

Pada pemodelan dengan ARFIMAX, model ARFIMA terpilih akan ditambahkan variabel eksogen yaitu faktor-faktor meteorologi berupa suhu udara minimum (Tn), suhu udara maksimum (Tx), suhu udara rata-rata (Tavg), kelembaban udara rata-rata (RH_avg), lamanya penyinaran sinar matahari (ss), kecepatan angin maksimum (ff_x), kecepatan angin rata-rata (ff_avg), dan arah angin saat kecepatan maksimum (ddd_x).

Pemodelan dilakukan dengan pendekatan *trial and error* untuk mengevaluasi setiap kombinasi model yang memungkinkan. Proses ini bertujuan untuk menentukan kombinasi variabel eksogen terbaik yang dapat memberikan hasil prediksi yang paling akurat. Hasil estimasi parameter model ARFIMA(1,0.33,1) pada berbagai kombinasi variabel eksogen diuji apakah parameter berpengaruh signifikan pada model. Taraf signifikansi pada pengujian signifikansi parameter model ARFIMAX yang digunakan yaitu $\alpha = 0.1$. Hasil pengujian signifikansi parameter model ARFIMAX disajikan pada Tabel berikut.

Tabel 5. Hasil Pengujian Signifikansi Parameter ARFIMAX(1, 0.33,1)

Model	Variabel Eksogen	Signifikansi Parameter ($\alpha = 0.1$)
1	Tn, RH_avg	RH_avg tidak signifikan
2	Tx, ddd_x	Tx tidak signifikan
3	ff_x, ddd_x	ff_x tidak signifikan
4	Tn, Tavg, ff_x	ff_x tidak signifikan
5	Tn, Tavg, ff_avg	ff_avg tidak signifikan
6	Tn, Tavg, ss	Signifikan
7	Tn, Tavg, ddd_x	Signifikan
8	Tn, RH_avg, ddd_x	RH_avg tidak signifikan
9	Tx, Tavg, ss	Tx, ss tidak signifikan
10	Tavg, ff_x, ddd_x	ff_x tidak signifikan
11	Tn, Tavg, ff_x, ff_avg	ff_x, ff_avg tidak signifikan
12	Tn, Tavg, ss, ddd_x	ss, ddd_x tidak signifikan
13	Tn, RH_avg, ss, ddd_x	RH_avg, ss, ddd_x tidak signifikan
14	Tx, Tavg, ff_x, ddd_x	Tx, ff_x tidak signifikan
15	Tx, Tavg, ff_avg, ss	Tx, ff_avg, ss tidak signifikan
16	Tavg, ff_x, ss, ddd_x	ff_x, ss tidak signifikan
17	Tavg, ff_avg, ss, ddd_x	ff_avg, ss tidak signifikan
18	Tn, Tx, Tavg, ff_x, ss	Tx, ff_x tidak signifikan
19	Tn, Tx, Tavg, ff_x, ddd_x	Tx, ff_x tidak signifikan
20	Tn, Tx, Tavg, ff_avg, ss	Tx, ff_avg tidak signifikan

Berdasarkan Tabel 5, ditemukan bahwa model yang seluruh parameternya signifikan adalah model 6 yaitu ARFIMAX(1,0.33,1) dengan variabel eksogen suhu udara minimum (Tn), suhu udara rata-rata (Tavg), dan lamanya penyinaran sinar matahari (ss), dan model 7 yaitu ARFIMAX(1,0.33,1) dengan variabel eksogen suhu udara minimum (Tn), suhu udara rata-rata (Tavg), dan arah angin saat kecepatan maksimum (ddd_x). Kedua model ARFIMAX tersebut memenuhi kriteria signifikansi parameter. Kedua model tersebut akan dilakukan pengujian diagnostik model pada tahap selanjutnya. Hasil pengujian diagnostik model disajikan pada tabel berikut.

Tabel 6. Hasil Diagnostik Model ARFIMAX(1,0.33,1)

Model	p-value			AIC
	Uji Kolmogorov-Smirnov	Uji L-jung Box	Uji ARCH-LM	
ARFIMAX(1,0.33,1) dengan variabel eksogen suhu udara minimum, suhu udara rata-rata, dan lamanya penyinaran sinar matahari	0.8225	0.66	0.4533	1988.88
ARFIMAX(1,0.33,1) dengan variabel eksogen suhu udara minimum, suhu udara rata-rata, dan arah angin saat kecepatan maksimum	0.9641	0.5929	0.2571	1988.74

Berdasarkan Tabel 6, hasil pengujian diagnostik kedua model ARFIMAX(1,0.33,1) telah memenuhi asumsi residual model. Hal ini ditunjukkan dari nilai *p-value* pada semua uji diagnostik model diatas 0.05. Dengan demikian, kedua residual model ARFIMAX(1,0.33,1) berdistribusi normal, tidak ada autokorelasi, dan variansinya homogen.

Selanjutnya model ARFIMAX terbaik dipilih berdasarkan nilai AIC terkecil. Berdasarkan Tabel 6, model ARFIMAX(1,0.33,1) dengan variabel eksogen suhu udara minimum, suhu udara rata-rata, dan arah angin saat kecepatan maksimum memiliki nilai AIC sebesar 1988.74. Nilai ini lebih kecil dibandingkan model

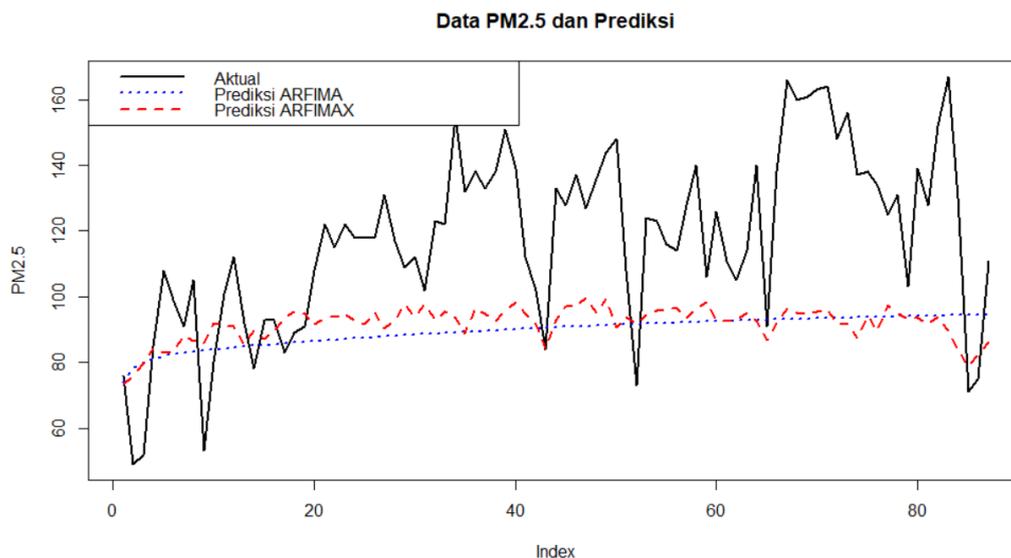
ARFIMAX(1,0.33,1) yang menggunakan variabel eksogen suhu udara minimum, suhu udara rata-rata, dan lamanya penyinaran sinar matahari, yang memiliki nilai AIC sebesar 1988.88. Oleh karena itu, model ARFIMAX(1,0.33,1) dengan variabel eksogen berupa suhu udara minimum, suhu udara rata-rata, dan arah angin saat kecepatan maksimum dipilih sebagai model terbaik untuk digunakan pada tahap analisis selanjutnya.

Model ARFIMAX(1,0.33,1) dengan variabel eksogen suhu udara minimum (T_n), suhu udara rata-rata (T_{avg}), dan arah angin saat kecepatan maksimum (ddd_x) yang diperoleh dapat dituliskan sebagai berikut.

$$y_t = -20.8323 + 1.2913y_{t-1} - 0.4272y_{t-2} + 0.1657y_{t-3} + \dots + e_t + 0.8827e_{t-1} + 2.1520T_n + 2.0638T_{avg} + 0.0171ddd_x \quad (7)$$

7. Validasi Model

Pada tahapan ini akan dilakukan prediksi nilai indeks kualitas udara $PM_{2.5}$ menggunakan model ARFIMA(1,0.33,1) dan ARFIMAX(1,0.33,1) dengan variabel eksogen suhu udara minimum, suhu udara rata-rata, dan arah angin saat kecepatan maksimum pada data *testing*. Hasil prediksi dari masing-masing model digunakan untuk menghitung akurasi model menggunakan nilai MAPE. Berikut disajikan grafik hasil prediksi pada masing-masing model.



Gambar 4. Plot Hasil Prediksi Data Indeks Kualitas Udara $PM_{2.5}$ pada Data *Testing*

Berdasarkan hasil analisis, nilai MAPE untuk model ARFIMA(1,0.33,1) adalah sebesar 25,76%, sedangkan untuk model ARFIMAX(1,0.33,1) dengan variabel eksogen berupa suhu udara minimum, suhu udara rata-rata, dan arah angin saat kecepatan maksimum adalah sebesar 23,69%. Dari hasil ini, terlihat bahwa MAPE pada model ARFIMAX(1,0.33,1) tersebut lebih kecil dibandingkan dengan model ARFIMA(1,0.33,1), yang menunjukkan bahwa pelibatan variabel eksogen berkontribusi dalam meningkatkan akurasi prediksi. Hal ini sejalan dengan penelitian Indeks Kualitas Udara $PM_{2.5}$ di Kota Jakarta Pusat, yang juga menunjukkan bahwa melibatkan variabel eksogen dalam model secara signifikan dapat meningkatkan akurasi prediksi[3].

D. Simpulan

Berdasarkan hasil penelitian, data Indeks Kualitas Udara $PM_{2.5}$ di kawasan Kemayoran, Jakarta, menunjukkan adanya ketergantungan jangka panjang sehingga dapat dimodelkan menggunakan ARFIMA dan ARFIMAX. Model ARFIMAX(1,0.33,1) dengan variabel eksogen berupa suhu udara minimum, suhu udara rata-rata, dan arah angin pada kecepatan maksimum terbukti memiliki akurasi yang lebih baik dibandingkan model ARFIMA(1,0.33,1), dengan nilai MAPE masing-masing sebesar 23,69% dan 25,76%. Pada model ARFIMAX variabel suhu udara minimum, suhu udara rata-rata, dan arah angin pada kecepatan maksimum berpengaruh signifikan terhadap akurasi prediksi indeks kualitas udara $PM_{2.5}$.

Penelitian selanjutnya untuk meningkatkan akurasi model dapat menambahkan variabel eksogen lain misalnya curah hujan dan volume lalu lintas kendaraan yang beroperasi. Selain itu, mengombinasikan ARFIMAX dengan metode *machine learning* atau pendekatan *hybrid* lainnya yang dapat meningkatkan akurasi prediksi juga dapat dieksplorasi.

E. Referensi

- [1] World Health Organization, *Air Quality Guidelines - Update 2021*. 2021.
- [2] Khadijah Nur Azizah, "BMKG Catat Konsentrasi $PM_{2.5}$ Ibukota Paling Tidak Sehat di Periode Ini." Accessed: Jan. 15, 2025. [Online]. Available: <https://health.detik.com/berita-detikhealth/d-6899650/bmkg-catat-konsentrasi-pm2-5-ibukota-paling-tidak-sehat-di-periode-ini>
- [3] N. F. Muzakki, A. Z. Putri, S. Maruli, and F. Kartiasih, "Forecasting the Air Quality Index by Utilizing Several Meteorological Factors Using the ARIMAX Method (Case Study: Central Jakarta City)," *J. JTJK. Teknol. Inf. Dan Komun.*, vol. 8, no. 3, pp. 569–586, Apr. 2024, doi: 10.35870/jtik.v8i3.2012.
- [4] R. Wahyudi, S. Annas, and Z. Rais, "Analisis Support Vector Regression (SVR) untuk Meramalkan Indeks Kualitas Udara di Kota Makassar," *J. VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 5, no. 3, pp. 104–117, Dec. 2023, doi: <https://doi.org/10.35580/variasiunm107>.
- [5] A. Pant, R. C. Joshi, S. Sharma, and K. Pant, "Predictive Modeling for Forecasting Air Quality Index (AQI) Using Time Series Analysis," *Avicenna J. Environ. Health Eng.*, vol. 10, no. 1, pp. 38–43, Sep. 2023, doi: 10.34172/ajehe.2023.5376.
- [6] R. Yunis, A. Andri, and D. Djoni, "Hybridization Model for Air Pollution Prediction Using Time Series Data," *CogITO Smart J.*, vol. 10, no. 1, pp. 422–435, Jun. 2024, doi: 10.31154/cogito.v10i1.619.422-435.
- [7] D. Hartini and N. N. Nurmaleni, "Penerapan Model Autoregressive Fractionally Integrated Moving Average (ARFIMA) dalam Prakiraan Data Suku Bunga PUAB (Pasar Uang Antar Bank)," *LOGIK@*, vol. 8, no. 1, pp. 24–35, 2018.
- [8] D. I. Octaviyani, M. Y. Wijaya, and N. Fitriyati, "Estimation Parameter d in Autoregressive Fractionally Integrated Moving Average Model in Predicting Wind Speed," *Inpr. Indones. J. Pure Appl. Math.*, vol. 1, no. 2, Dec. 2019, doi: 10.15408/inprime.v1i2.13676.
- [9] R. E. Caraka, W. Sugiyarto, H. Yasin, and Z. Mahmud, "Long Memory Models to Forecasting Temperature," *Pros. Semin. Nas. Meteorol. Dan Klimatol.*, 2016.

- [10] S. Degiannakis, "ARFIMAX and ARFIMAX-TARCH realized volatility modeling," *J. Appl. Stat.*, vol. 35, no. 10, pp. 1169–1180, Oct. 2008, doi: 10.1080/02664760802271017.
- [11] K. P. Sarkar *et al.*, "Forecasting long range dependent time series with exogenous variable using ARFIMAX model," *Indian J. Agric. Sci.*, vol. 90, no. 7, pp. 1302–1305, Oct. 2020, doi: 10.56093/ijas.v90i7.105599.
- [12] Mahmudi, H. Mehri, and S. Inna, "Pengaruh Variabel Eksogen terhadap Pemodelan Curah Hujan di Kota Bandung," *J. Absis J. Pendidik. Mat. Dan Mat.*, vol. 6, no. 2, pp. 988–1005, 2024.
- [13] M. Boutahar and R. Khalfaoui, "Estimation of the long memory parameter in non stationary models: A Simulation Study," *GREQAM - Group. Rech. En Econ. Quant. Aix-Marseille*, May 2011.
- [14] C. W. J. Granger and R. Joyeux, "An Introduction To Long-Memory Time Series Models and Fractional Differencing," *J. Time Ser. Anal.*, vol. 1, no. 1, pp. 15–29, Jan. 1980, doi: 10.1111/j.1467-9892.1980.tb00297.x.
- [15] J. R. M. Hosking, "Fractional differencing," *Biometrika*, vol. 68, no. 1, pp. 165–176, 1981, doi: 10.1093/biomet/68.1.165.
- [16] W. Palma, *Time Series Analysis*. New Jersey: John Wiley & Sons, 2016.
- [17] D. C. Montgomery, C. L. Jennings, and M. Kulahci, *Introduction to Time Series Analysis and Forecasting*. New Jersey: John Wiley & Sons, 2015.
- [18] B. Qian and K. Rasheed, "Hurst exponent and financial market predictability," *Proc. IASTED Int. Conf. Financ. Eng. Appl.*, pp. 203–209, 2004.
- [19] V. B. Sitorus, S. Wahyuningsih, and M. N. Hayati, "Peramalan dengan Metode Seasonal Autoregressive Integrated Moving Average (SARIMA) di Bidang Ekonomi," *J. EKSPONENSIAL*, vol. 8, no. 1, pp. 17–26, 2017.