

Analisis Prediktif Menggunakan Metode *Hybrid Seasonal Autoregressive Integrated Moving Average – Artificial Neural Network* pada Data Konsentrasi PM_{2.5} Harian di DKI Jakarta

I Kadek Pasek Kusuma Adi Putra¹, Sediono^{2✉}, M. Fariz Fadillah Mardianto³, Elly Pusporani⁴

^{1, 2, 3, 4} Statistika, Fakultas Sains dan Teknologi, Universitas Airlangga, Indonesia

Informasi Artikel

Riwayat Artikel

Diserahkan : 06-01-2024

Direvisi : 13-01-2024

Diterima : 15-01-2024

Kata Kunci:

DKI Jakarta, *Hybrid*
ARIMA-ANN, PM_{2.5},
Polusi Udara, SDGs

Keywords :

DKI Jakarta, *Hybrid*
ARIMA-ANN, PM_{2.5}, Air
Pollution, SDGs

Corresponding Author :

Sediono

Statistika, Fakultas Sains dan Teknologi, Universitas Airlangga, Indonesia

Jl. Dr. Ir. H. Soekarno, Mulyorejo, Kec. Mulyorejo, Surabaya, Jawa Timur 60115

Email: sediono101@gmail.com

ABSTRAK

Polusi udara, khususnya PM_{2.5} merupakan ancaman tingkat global yang sulit untuk ditangani. Partikulat ini memberikan dampak yang serius secara lingkungan, kesehatan, dan juga ekonomi di DKI Jakarta. Tujuan dari penelitian ini adalah untuk memprediksi konsentrasi PM_{2.5} dengan sangat akurat. Penelitian ini menghadirkan model prediksi inovatif, yaitu *Hybrid* ARIMA-ANN. Dibandingkan dengan ARIMA dan ANN secara terpisah, model ini menunjukkan kinerja luar biasa dengan R² mencapai 0,9012, MAPE 13,603%, dan RMSE 4,061 pada data *train*. Evaluasi pada data uji menghasilkan RMSE 5,961 dan MAPE 7,622%, menandakan kemampuan prediksi yang sangat baik selama 14 periode ke depan. Temuan ini memberikan panduan penting bagi pemerintah untuk memprediksi kualitas udara di masa depan dan meningkatkan kebijakan penanganan polusi udara di DKI Jakarta. Implikasi kesimpulan ini juga mendukung pencapaian SDGs, terutama poin ke-3, ke-13, dan ke-15.

ABSTRACT

Air pollution, particularly PM_{2.5}, poses a global threat that proves challenging to address. This particulate matter has significant environmental, health, and economic repercussions in DKI Jakarta. The aim of this research is to predict PM_{2.5} concentrations with exceptional accuracy. The study introduces an innovative prediction model, namely *Hybrid* ARIMA-ANN. In comparison to ARIMA and ANN individually, this model demonstrates outstanding performance, achieving an R² of 0.9012, MAPE of 13.603%, and RMSE of 4.061 on the training data. Evaluation on test data yields an RMSE of 5.961 and MAPE of 7.622%, indicating excellent predictive capabilities over the next 14 periods. These findings offer crucial insights for the government to forecast air quality in the future and enhance policies addressing air pollution in DKI Jakarta. The implications of these conclusions also align with the achievement of SDGs, particularly goals 3, 13, and 15.

PENDAHULUAN

Polusi udara merupakan ancaman tingkat global yang berdampak pada kesehatan manusia, ekosistem daratan, iklim, dan lingkungan (Manisolidis, 2020). Menurut World Health Organization (WHO), hampir setiap tahun ada kematian sejumlah 4.200 jiwa dan 91% populasi dunia hidup di kondisi udara yang di atas standar WHO (Jia, 2019). Sehingga, keberadaan polusi udara berlebih akan menyebabkan kerusakan yang bahkan dapat menghambat tercapainya beberapa poin *Sustainable Development Goals* (SDGs), seperti poin ke-3 mengenai kehidupan sehat dan sejahtera, poin ke-13 penanganan perubahan iklim, dan poin ke-15 mengenai ekosistem daratan.

Berdasarkan situs Indeks Kualitas Udara global (iqair.com) menyatakan bahwa kualitas udara di Indonesia per tahun 2022 memiliki rata – rata indeks kualitas udara sebesar 89 dimana indeks ini termasuk kategori “Sedang”. Selain itu, Indonesia pernah menduduki peringkat 1 dengan tingkat konsentrasi partikulat PM_{2.5} tertinggi se-Asia Tenggara (Safira, 2020). PM_{2.5} merupakan partikel udara atau debu yang memiliki diameter berukuran kurang lebih 2,5 μm (mikrometer). Menurut WHO, PM_{2.5} merupakan partikulat debu yang diklasifikasikan sebagai zat pemicu kanker dan polutan yang paling berbahaya (Rodriguez-Urrego, 2020).

Ambang batas normal konsentrasi PM_{2.5} harian adalah sebesar 15 $\mu\text{g}/\text{m}^3$ dan untuk tahunan sebesar 5 $\mu\text{g}/\text{m}^3$ (Burki, 2021). Akan tetapi, di Indonesia pemantauan yang berfokus pada partikulat PM_{2.5} baru dimulai semenjak tahun 2020 dan batas ambang normalnya ditetapkan sebesar 65 $\mu\text{g}/\text{m}^3$ (BMKG, 2020). Hal ini menunjukkan bahwa batas normal PM_{2.5} Indonesia jauh di atas dari ketentuan internasional.

Pada tahun 2022 Daerah Khusus Ibukota (DKI) Jakarta memiliki konsentrasi PM_{2.5} tahunan tertinggi di Indonesia mencapai 36,2 $\mu\text{g}/\text{m}^3$ (Ahdiat, 2023). Hal ini tentu sangat berbahaya karena DKI Jakarta merupakan pusat aktivitas dan menjadi kota terpadat di Indonesia. Selain itu, tercatat bahwa konsentrasi PM_{2.5} harian di Jakarta pada akhir September 2023 mencapai 72,7 $\mu\text{g}/\text{m}^3$ dan angka tersebut masuk ke dalam kategori “Tidak Sehat” (dilansir dari situs iqair.com).

Faktanya, pada tahun 2023 terdapat ± 11.000 kematian di Kota Jakarta akibat polusi udara (IQAir, 2023). Hal ini juga berdampak pada ekonomi di Kota Jakarta dimana terdapat kerugian sekitar 2,1 miliar US Dolar dari kerusakan akibat polusi udara yang dialami (Putra, 2023). Pemerintah DKI Jakarta sudah pernah memberlakukan berbagai kebijakan, tetapi belum ada kebijakan yang berfokus untuk memprediksi kualitas udara di DKI Jakarta secara periodik.

Sebagai referensi, terdapat beberapa penelitian terdahulu yang membahas kajian prediksi kualitas udara di DKI Jakarta. Terdapat penelitian yang memodelkan konsentrasi PM_{2.5} bulanan di DKI Jakarta dengan metode linier berupa *Seasonal Autoregressive Integrated Moving Average* (SARIMA) (Amelia et al, 2022) . Berdasarkan hasilnya, diperoleh akurasi forecasting yang baik dengan *mean absolute error percentage* (MAPE) 11,52%. Meninjau dari metode nonlinier, terdapat penelitian yang mengestimasi berbagai polutan di DKI Jakarta dengan menggunakan metode *Artificial Neural Network* (ANN) (Sari, 2019). Hasilnya cukup beragam tergantung zat polutannya. Prediksi mengenai PM terkait (PM₁₀) menghasilkan MAPE sebesar 12,32% untuk data testing. Namun, penelitian ini tidak melibatkan konsentrasi PM_{2.5} di dalamnya.

Data konsentrasi PM_{2.5} bersifat sangat acak dan mengandung sifat linier maupun nonlinier (Hajirahimi, 2019). Oleh karena itu, pemodelan secara *hybrid* dilakukan untuk meningkatkan performa dan akurasi dari sistem (Alrweili, 2022). Berdasarkan penjabaran tersebut, maka penelitian ini berfokus pada prediksi konsentrasi PM_{2.5} harian di DKI Jakarta dengan inovasi berupa metode yang digunakan adalah *hybrid* ARIMA-ANN. Dengan demikian, diharapkan penelitian ini dapat menyumbangkan metode prediksi konsentrasi PM_{2.5} yang tepat. Serta, membantu pemerintah dalam merencanakan kebijakan pencegahan pencemaran udara untuk masyarakat DKI Jakarta, bahkan untuk masyarakat Indonesia.

METODE PENELITIAN

Sumber Data dan Variabel Data

Data yang digunakan dalam penelitian berupa data sekunder mengenai konsentrasi PM_{2.5} harian di DKI Jakarta yang diperoleh dari situs internasional airnow.gov. Situs tersebut merupakan situs buatan organisasi lingkungan internasional Environmental Protection Agency (EPA) dan National Oceanic and Atmospheric Administration (NOAA) yang menyediakan informasi dari kualitas udara lokal seluruh wilayah dunia. Pada data tersebut, stasiun pemantauan udara yang digunakan adalah stasiun pemantauan udara di DKI Jakarta Pusat yang terletak di Bundaran HI, Jalan MH Thamrin, Jakarta Pusat dengan periode harian dimulai dari tanggal 1 Januari 2022 sampai 17 September 2023 sebagai data *train* dan 18 September – 1 Oktober 2023 sebagai data *test*. Taraf signifikansi penelitian ditetapkan sebesar 5%, serta satuan pengukuran konsentrasi PM_{2.5} menggunakan $\mu\text{g}/\text{m}^3$.

Autoregressive Integrated Moving Average (ARIMA)

Model yang menggabungkan AR dan MA pada satu kesatuan disebut sebagai model stasioner *Autoregressive Moving Average* (ARMA) dengan memiliki 2 orde berupa p dan q . Model ARIMA merupakan model runtun waktu yang mengatasi nonstasioneritas data dengan *differencing*. Istilah “*Integrated*” pada ARIMA merupakan orde dari *differencing* dengan simbol orde d . Secara umum model ARIMA orde p, d , dan q dapat dituliskan pada Persamaan (1) sebagai berikut.

$$\phi_p(B)(1-B)^d Z_t = \theta_0 + \theta_q(B)a_t \quad (1)$$

dengan

Z_t : Data Z_t pada periode $t, t = 1, 2, \dots, n$

a_t : *Error* pada periode t

$\theta_q(B)$: Operasi *backshift* untuk parameter MA

$\phi_p(B)$: Operasi *backshift* untuk parameter AR

θ_0 : Parameter deterministik $\mu(1 - \phi_1 - \phi_2 - \dots - \phi_p)$

Namun, suatu runtun waktu terkadang mengandung fenomena *seasonal* atau musiman yang berulang pada periode waktu tertentu (Wei, 2016). Secara umum, model ARIMA musiman dinyatakan sebagai model $\text{ARIMA}(p, d, q)(P, D, Q)^S$ dengan, D merupakan orde *differencing* musiman, S merupakan periode musiman, P merupakan orde AR musiman, dan Q merupakan orde MA musiman. Bentuk umum ARIMA musiman dinyatakan pada Persamaan (2) sebagai berikut.

$$\phi_p(B) \Phi_P(B^S)(1-B)^d(1-B^S)^D Z_t = \theta_q(B) \Theta_Q(B^S) a_t \quad (2)$$

dengan

$\phi_p(B)$: $(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ yang merupakan polinomial AR(p) nonmusiman

$\theta_q(B)$: $(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$ yang merupakan polinomial MA(q) nonmusiman

$\Phi_P(B^S)$: $(1 - \Phi_1 B^S - \Phi_2 B^{2S} \dots - \Phi_P B^{PS})$ yang merupakan polinomial AR(P) musiman

$\Theta_Q(B^S)$: $(1 - \Theta_1 B^S - \Theta_2 B^{2S} \dots - \Theta_Q B^{QS})$ yang merupakan polinomial MA(Q) musiman

Artificial Neural Network (ANN)

Artificial neural network (ANN) atau jaringan syaraf tiruan merupakan teknik kecerdasan buatan *supervised learning* yang meniru sifat – sifat jaringan syaraf manusia melalui operasi matematika. ANN terdiri dari lapisan – lapisan buatan yang membentuknya menjadi suatu arsitektur – arsitektur tertentu (Sakinah *et al*, 2018). Arsitektur ANN adalah susunan komponen

lapisan dan *neuron* pada *input*, *hidden*, dan *output* yang terhubung dengan bobot, fungsi aktivasi, dan fungsi pembelajaran (*learning function*). Bobot yang terdapat pada masing-masing *neuron* diinisialisasikan secara acak dan dilatih dengan algoritma *backpropagation* yang bertujuan untuk meminimalkan nilai *loss* dari *loss function* pada data validasi (Satria, 2018). Model ANN *backpropagation* dituliskan pada Persamaan (3) sebagai berikut.

$$y = g \left(\sum_{j=1}^p w_0 + w_j f \left(v_{0j} + \sum_{i=1}^n x_i v_{ij} \right) \right) \quad (3)$$

dengan

w_0, w_j	: bobot bias pada unit <i>output</i>	$f()$	: fungsi aktivasi pada unit <i>input</i>
y ;			ke tersembunyi;
v_{0j}	: bobot bias pada unit	x_i	: unit <i>input</i> ke- <i>i</i> ;
tersembunyi z_j ;		z_j	: unit tersembunyi ke- <i>j</i> ;
v_{ij}	: bobot garis dari unit x_i ke	y	: unit <i>output</i> ;
unit tersembunyi z_j ;		n	: banyaknya unit <i>input</i> ;
$g()$	: fungsi aktivasi pada unit	p	: banyaknya unit tersembunyi;
tersembunyi ke <i>output</i> ;		l	: banyaknya unit <i>output</i> .

Proses dalam algoritma *backpropagation* pada ANN dibedakan menjadi 2 tahap utama, yaitu *feed forward* dan *backward*. Pada proses ini juga dihitung nilai galat yang merepresentasikan tingkat kesalahan dari *backpropagation* (Yuberta, 2022).

Hybrid ARIMA-ANN

Model *hybrid* merupakan suatu metode kombinasi dari satu atau lebih dari dua model dalam fungsi suatu sistem. Model ARIMA atau ANN merupakan model untuk mengatasi masalah linier atau nonlinier. Secara fakta, jarang ditemukan kejadian *time series* yang murni linier maupun murni nonlinier maka model *hybrid* yang mempunyai kemampuan memodelkan linier dan nonlinier dapat dijadikan strategi yang praktis. Secara umum, model *hybrid* yang pertama kali diperkenalkan adalah model aditif seperti yang ada pada Persamaan (4) berikut (Zhang, 2003).

$$y = Z_t + N_t \quad (4)$$

dengan

Z_t : Model linier (Model ARIMA)

N_t : Model nonlinier (Model Residual ARIMA)

Melalui $\hat{Z}_t$, maka diperoleh komponen yang nonlinier residual dari ARIMA diperoleh sebagai berikut.

$$e_t = y_t - \hat{Z}_t \quad (5)$$

Maka, estimator untuk $\hat{N}_t$ dengan *artificial neural network* dituliskan sebagai berikut.

$$\hat{N}_t = \hat{e}_t = g(e_{t-1}, e_{t-2}, \dots, e_{t-n}) + v_t \quad (6)$$

Fungsi $g(e_{t-1}, e_{t-2}, \dots, e_{t-n})$ akan melalui proses pendekatan dengan ANN, serta v_t merupakan galat yang bersifat *white noise* (Alsuwaylimi, 2023).

Langkah – langkah Analisis Data

Adapun, langkah – langkah analisis pada penelitian ini ditulis secara singkat adalah sebagai berikut.

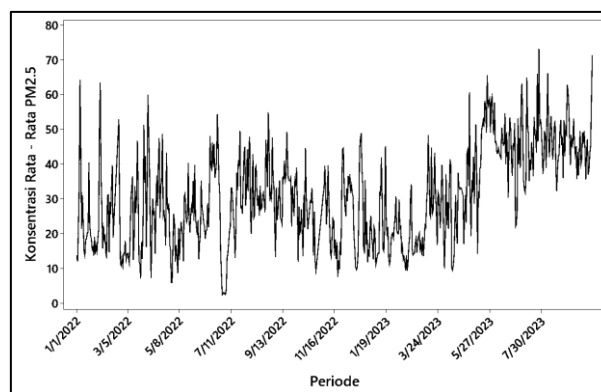
1. Mendeskripsikan data konsentrasi harian PM_{2.5} DKI Jakarta dari 1 Januari 2022 sampai 1 Oktober 2023 dengan metode statistika deskriptif dan visualisasi data.

2. Membagi *dataset* menjadi 2, yaitu data *train* dari 1 Januari 2022 sampai 17 September 2023 dan data *test* dari 18 September sampai 1 Oktober 2023.
3. Melakukan analisis pemodelan ARIMA pada data konsentrasi PM_{2.5} DKI Jakarta dan menyimpan residual model ARIMA
4. Melakukan analisis ANN pada data konsentrasi PM_{2.5} dan residual model ARIMA.
5. Mengimplementasikan estimasi *hybrid* dengan menjumlahkan konsentrasi PM_{2.5} yang telah diestimasi ARIMA dengan residualnya yang telah diestimasi oleh *neural network*.
6. Membandingkan hasil dari ARIMA, ANN, dan *hybrid* ARIMA-ANN dengan RMSE dan MAPE terkecil.

HASIL DAN PEMBAHASAN

Statistika Deskriptif

Kajian ini membahas bagaimana karakteristik rata – rata PM_{2.5} harian di DKI Jakarta. Visualisasi runtun waktu data konsentrasi PM_{2.5} secara harian yang dimulai dari 1 Januari 2022 sampai 1 Oktober 2023 ditunjukkan pada Gambar 1 sebagai berikut.



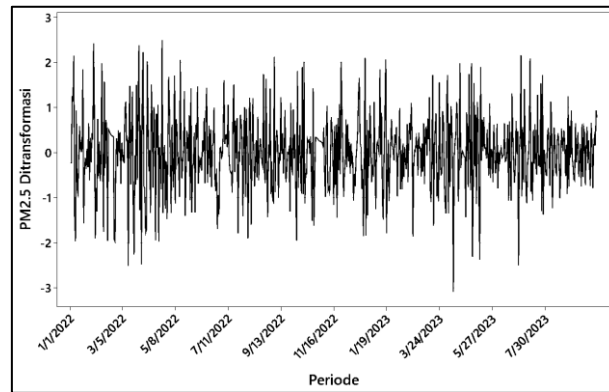
Gambar 1. *Plot Runtun Waktu PM_{2.5} di DKI Jakarta*

Pada gambar tersebut terlihat jelas bahwa konsentrasi PM_{2.5} di DKI Jakarta menunjukkan periode yang fluktuatif dan acak. Data yang fluktuatif memerlukan metode pemodelan yang tepat karena sangat rumit untuk memprediksi ataupun mengestimasi nilainya. Selama periode tersebut rata – rata keseluruhan konsentrasi berada pada 30,672 $\mu\text{g}/\text{m}^3$ dengan standar deviasi yang tinggi sebesar 13,717. Secara histori, konsentrasi PM_{2.5} tertinggi mencapai 73,1 $\mu\text{g}/\text{m}^3$ yang terjadi pada tanggal 27 Juli 2023 dan termasuk kategori “Tidak Sehat” (Larasati, 2021).

Analisis Seasonal ARIMA

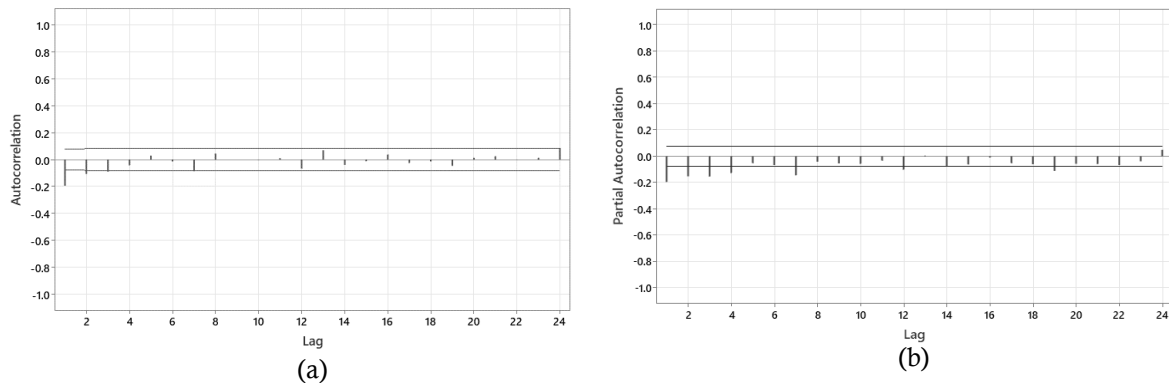
ARIMA merupakan model runtun waktu linier yang memerlukan beberapa syarat dalam pemodelannya. Syarat – syarat tersebut berupa data yang wajib stasioner dalam rata – rata dan varians. Serta galat model berdistribusi normal dengan rata – rata 0 dan varians yang konstan (Abebe, 2020). Sehingga sebagai langkah awal, maka pada analisis ini akan dilakukan pengecekan stasioneritas terlebih dahulu.

Pengecekan stasioneritas varians data menggunakan transformasi λ Tukey. Pada awalnya nilai λ menunjukkan 0,5 yang berarti data sebaiknya ditransformasi dalam bentuk akar ($\sqrt{Z_t}$) (Nugroho, 2023). Kemudian, uji *Augmented Dickey Fuller* (ADF) menunjukkan *p-value* sebesar 0,069 dengan keputusan gagal menolak H_0 . Hal tersebut menunjukkan bahwa data tidak stasioner juga dalam rata – rata sehingga perlu transformasi *differencing* setelah transformasi Tukey (Xue, 2016). Hasil transformasi ditunjukkan pada Gambar 2 sebagai berikut.



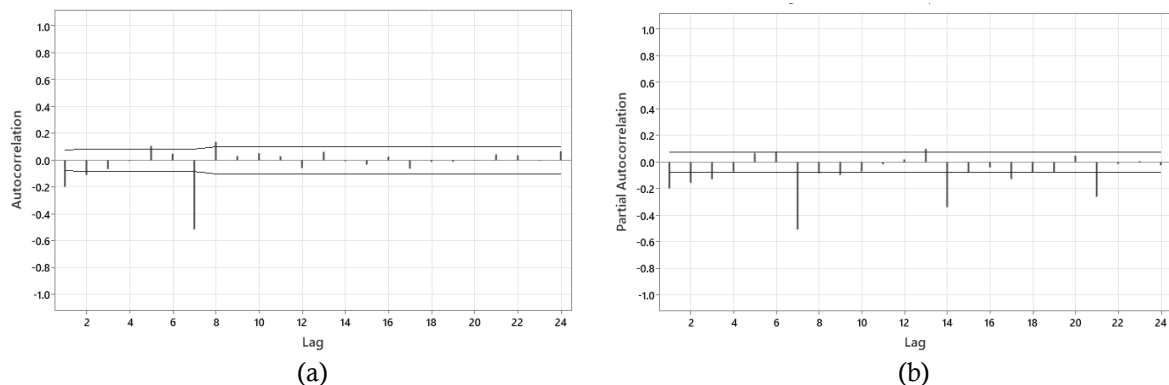
Gambar 2. Plot Runtun Waktu Transformasi

Setelah data konsentrasi $PM_{2.5}$ harian stasioner dalam rata – rata dan varians, maka analisis selanjutnya adalah meninjau *plot autocorrelation function* (ACF) untuk menentukan orde q dan *plot partial autocorrelation function* (PACF) untuk menentukan orde p (Wei, 2016). Hasilnya ditunjukkan pada Gambar 3 berikut.



Gambar 3. (a) Plot ACF (b) Plot PACF

Namun, pada Gambar 3 teridentifikasi pada PACF terdapat *lag* musiman yang keluar, yaitu *lag* 7. Sehingga, ada kemungkinan indikasi bahwa data konsentrasi $PM_{2.5}$ di DKI Jakarta memiliki pola musiman (Lubis, 2022). Sehingga, perlu dilakukan *differencing* musiman, yaitu *differencing lag* 7. Hasil *plot* ACF dan PACF ditunjukkan kembali pada Gambar 4 berikut.



Gambar 4. (a) Plot ACF Lag Musiman (b) Plot PACF Lag Musiman

Gambar 4 lebih memperjelas bahwa *lag* 7 yang telah keluar batas signifikansi menunjukkan adanya pola musiman pada data konsentrasi PM_{2.5}. Selanjutnya, pemodelan *seasonal* ARIMA dapat dilakukan dengan menentukan orde *p*, *q*, *Q*, dan *P*.

ACF yang terputus setelah *lag* 2 menunjukkan orde *q* sebanyak 1 sampai 2 dan PACF yang terputus setelah *lag* 3 menunjukkan orde *p* sebanyak 1 sampai 3. Mengenai orde *P* dan *Q* dijadikan 1 pada kasus ini untuk simplifikasi model (Kabbilawsh, 2020). Pada Tabel 1 ditunjukkan hasil estimasi parameter ARIMA dan dipilih ARIMA yang memiliki pengaruh signifikan pada seluruh estimator, serta residual yang normal dan *white noise*.

Tabel 1. Hasil Estimasi Model *Seasonal* ARIMA

Model ARIMA	Sig. Parameter	<i>P-value</i> Normalitas	<i>P-value</i> White Noise	Model ARIMA	Sig. Parameter	<i>P-value</i> Normalitas	<i>P-value</i> White Noise
(1,1,1)(1,1,1) ⁷	Estimator $\hat{\Phi}_1$ dan $\hat{\Theta}_1$ tidak signifikan	0,59	0,94	(2,1,2)(1,1,1) ⁷	Seluruh estimator signifikan	0,56	0,98
(1,1,2)(1,1,1) ⁷	Estimator $\hat{\theta}_1$ dan $\hat{\Theta}_1$ tidak signifikan	0,59	0,96	(3,1,1)(1,1,1) ⁷	Estimator $\hat{\phi}_2, \hat{\phi}_3, \hat{\Phi}_1$ dan $\hat{\Theta}_1$ tidak signifikan	0,60	0,99
(2,1,1)(1,1,1) ⁷	Estimator $\hat{\phi}_2, \hat{\Phi}_1$ dan $\hat{\Theta}_1$ tidak signifikan	0,61	0,99	(3,1,2)(1,1,1) ⁷	Estimator $\hat{\phi}_3$ tidak signifikan	0,56	0,96

Pada tabel tersebut, yang memenuhi semua kondisi pemodelan ARIMA adalah model *seasonal* dengan orde (2,1,2)(1,1,1)⁷. Persamaan estimasi ARIMA (2,1,2)(1,1,1)⁷ ditunjukkan pada Persamaan (7).

$$\hat{Z}_t = \left[\frac{(1 + B^7)(1 + 2,557B - 1,579B^2) a_t}{(1 + 0,089B^7)(1 - 1,267B + 0,394B^2)(1 - B)(1 - B^7)} \right]^2 \quad (7)$$

Hasilnya untuk data *train* menghasilkan R² sebesar 0,537, *Mean Absolute Percentage Error* (MAPE) sebesar 30,065% (berada di kategori cukup), dan *Root Mean Squared Error* (RMSE) sebesar 9,241. Adapun, untuk data *test* menghasilkan MAPE sebesar 13,105% (berada di kategori baik) dan RMSE sebesar 8,801. Semakin tinggi nilai R² akan membuat hasil estimasi *train* semakin sesuai dengan data aslinya, sedangkan MAPE dan RMSE yang semakin kecil akan membuat hasil estimasi ataupun prediksi memiliki *error* yang relatif kecil (Zhai, 2021).

Analisis *Artificial Neural Network* (ANN)

Estimasi konsentrasi PM_{2.5} dengan *neural network* berfokus pada arsitektur *neural network*, jumlah *neuron* pada *hidden layers*, fungsi aktivasi *hidden layer*, optimasi yang digunakan, *learning rate*, dan banyak iterasi (*epoch*). Pada penelitian ini, parameter yang ditetapkan adalah arsitektur *neural network* menggunakan 3 *hidden layer* dengan setiap lapisan secara berurutan memiliki 24 *neuron*, 12 *neuron*, dan 8 *neuron*. Optimasi yang digunakan adalah optimasi Adam dengan *learning rate* awal sebesar 0,01 dan iterasi atau *epochs* sebanyak 5000. Sebagai penghematan waktu komputasi, diterapkan fungsi *callback* yang mampu mengembalikan hasil *neural network* terbaik. Selain itu, diterapkan juga fungsi *earlystopping* pada 200 *epochs* dimana jika pada 200 *epochs* tidak ada improvisasi fungsi *loss validation* maka proses dihentikan.

Pada analisis ANN, data tidak perlu ditransformasi seperti pada analisis ARIMA sebelumnya. Parameter yang dicoba melalui *hyperparameter tuning* adalah fungsi aktivasi *hidden layer* berupa *Rectified Linear Unit* (ReLU), *Exponential Linear Unit* (ELU), dan *Sigmoid* dengan *lag* (fitur dari *neural network*) dicoba dari *lag* 7 – *lag* 14. Pemilihan arsitektur *neural network* terbaik

didasarkan dengan RMSE dan MAPE. Hasil estimasi dengan *neural network* ditunjukkan dalam Tabel 2 sebagai berikut.

Tabel 2. Hasil Estimasi *Neural Network*

Jumlah <i>lag</i> fitur	Fungsi Aktivasi	Metrik Evaluasi		Jumlah <i>lag</i> fitur	Fungsi Aktivasi	Metrik Evaluasi	
		RMSE	MAPE			RMSE	MAPE
7	ReLU	9,110	33,804%	11	ReLU	9,379	33,390%
	ELU	9,093	33,892%		ELU	8,987	33,184%
	<i>Sigmoid</i>	9,470	33,699%		<i>Sigmoid</i>	12,310	49,468%
8	ReLU	9,351	36,229%	12	ReLU	8,839	31,855%
	ELU	9,499	37,368%		ELU	8,980	32,027%
	<i>Sigmoid</i>	9,006	29,323%		<i>Sigmoid</i>	9,177	31,201%
9	ReLU	7,618	26,368%	13	ReLU	9,129	34,367%
	ELU	9,048	33,744%		ELU	8,980	33,116%
	<i>Sigmoid</i>	12,403	48,954%		<i>Sigmoid</i>	13,565	58,664%
10	ReLU	7,537	26,870%	14	ReLU	8,597	32,197%
	ELU	6,913	23,531%		ELU	8,839	32,604%
	<i>Sigmoid</i>	9,122	30,108%		<i>Sigmoid</i>	13,582	59,254%

Berdasarkan tabel tersebut, *neural network* dengan fitur 10 *lag* dari data konsentrasi PM_{2.5} dan fungsi aktivasi ELU yang dipilih untuk mengestimasi data konsentrasi PM_{2.5}. Estimasi ini menghasilkan nilai R^2 sebesar 0,739 yang menunjukkan peningkatan dibandingkan model ARIMA(2,1,2)(1,1,1)⁷.

Analisis Model *Hybrid* dan Komparasi Tiga Model

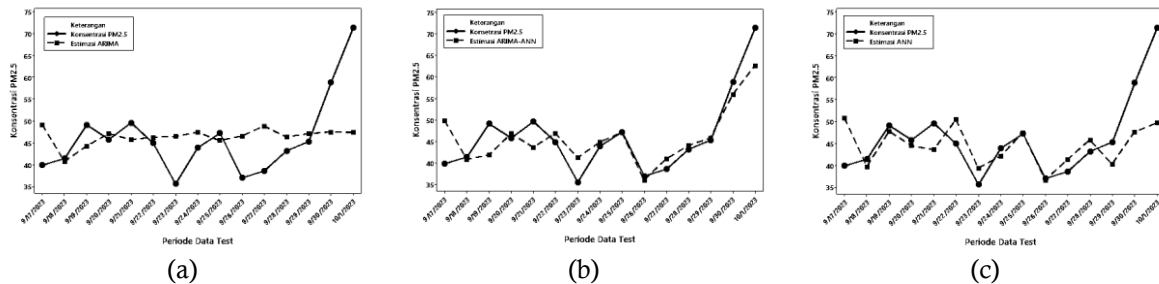
Analisis model *hybrid* memiliki langkah yang sama dengan *neural network*, tetapi perbedaannya adalah model *hybrid* memasukkan residual ARIMA yang diperoleh ke dalam proses *neural network*. Residual yang memiliki nilai negatif kurang cocok menggunakan MAPE sebagai evaluasi metriknya, sehingga *Mean Absolute Error* (MAE) digunakan pada kasus ini (Davydenko, 2016). Hasil estimasi residual ARIMA dengan *neural network* ditunjukkan pada Tabel 3 sebagai berikut.

Tabel 3. Hasil Estimasi *Neural Network* (Residual ARIMA(2,1,2)(1,1,1)⁷)

Jumlah <i>lag</i> fitur	Fungsi Aktivasi	Evaluasi Metrik		Jumlah <i>lag</i> fitur	Fungsi Aktivasi	Evaluasi Metrik	
		RMSE	MAE			RMSE	MAE
7	ReLU	8,559	6,634	11	ReLU	6,001	4,532
	ELU	5,672	4,433		ELU	8,942	6,850
	<i>Sigmoid</i>	9,181	7,017		<i>Sigmoid</i>	9,067	6,974
8	ReLU	6,162	4,679	12	ReLU	4,463	3,256
	ELU	8,885	6,843		ELU	7,547	5,693
	<i>Sigmoid</i>	9,076	6,965		<i>Sigmoid</i>	9,051	6,954
9	ReLU	8,755	6,737	13	ReLU	4,061	3,127
	ELU	7,527	5,799		ELU	6,409	4,761
	<i>Sigmoid</i>	9,056	6,949		<i>Sigmoid</i>	5,555	4,153
10	ReLU	6,179	7,949	14	ReLU	5,122	4,016
	ELU	6,910	8,960		ELU	5,689	4,286
	<i>Sigmoid</i>	6,971	9,058		<i>Sigmoid</i>	9,032	6,924

Pada tabel tersebut, estimasi residual terbaik berada pada *neural network* dengan fungsi aktivasi ReLU dan fitur *lag* sebanyak 13 *lag*. Hasil estimasi data *train* pada Tabel 3 diaplikasikan ke Persamaan (4) dengan $\hat{Z}_t$ adalah estimasi konsentrasi PM_{2.5} melalui ARIMA(2,1,2)(1,1,1)⁷ dan $\hat{N}_t$ adalah estimasi residual ARIMA melalui *neural network*.

Hasilnya sangat memuaskan pada data *train* menghasilkan R^2 sebesar 0,9012, MAPE sebesar 13,603%, dan RMSE sebesar 4,061. Sedangkan, pada data *test* menghasilkan MAPE sebesar 7,622% dan RMSE sebesar 5,961. Model *Hybrid* ARIMA-ANN memberikan peningkatan yang signifikan secara akurasi dan memiliki metrik *error* yang lebih kecil dibandingkan model ARIMA dan *neural network* secara terpisah. Visualisasi hasil prediksi (data *test*) secara komparasi dari ketiga model ditunjukkan pada Gambar 5 berikut.



Gambar 5. Komparasi Hasil pada Prediksi Data Test : (a) ARIMA (b) ARIMA-ANN (c) ANN

Pada grafik di atas nilai asli dari *data test* diwakili dengan garis menyambung dan pola lingkaran, sedangkan hasil dari prediksi diwakili garis putus – putus dan pola kotak. Terlihat bahwa ARIMA tidak baik untuk memprediksi jangka menengah karena setelah beberapa periode, menunjukkan tren konstan. Hasil *neural network* pada Gambar 5 (c) mampu menangkap pola *data test*, tetapi sedikit gagal dalam mengikuti tren peningkatan $PM_{2.5}$ di akhir periode. Hasil *Hybrid* ARIMA-ANN pada Gambar 4 (b) menunjukkan performa yang sangat baik memprediksi data *test* konsentrasi $PM_{2.5}$. Hasil prediksi *hybrid* ARIMA-ANN beserta statusnya ditunjukkan pada Tabel 4 sebagai berikut (Larasati, 2021).

Tabel 4. Hasil Prediksi *Hybrid* ARIMA-ANN

Tanggal	Konsentrasi $PM_{2.5}$ (Status)	Prediksi ARIMA-ANN (Status)	Tanggal	Konsentrasi $PM_{2.5}$ (Status)	Prediksi ARIMA-ANN (Status)
9/18/2023	41,4 (Cukup Tidak Sehat)	49,8 (Cukup Tidak Sehat)	9/25/2023	47,2 (Cukup Tidak Sehat)	47,1 (Cukup Tidak Sehat)
9/19/2023	49,1 (Cukup Tidak Sehat)	40,8 (Cukup Tidak Sehat)	9/26/2023	37,0 (Cukup Tidak Sehat)	36,0 (Cukup Tidak Sehat)
9/20/2023	45,8 (Cukup Tidak Sehat)	41,9 (Cukup Tidak Sehat)	9/27/2023	38,6 (Cukup Tidak Sehat)	40,9 (Cukup Tidak Sehat)
9/21/2023	49,6 (Cukup Tidak Sehat)	46,8 (Cukup Tidak Sehat)	9/28/2023	43,2 (Cukup Tidak Sehat)	44,0 (Cukup Tidak Sehat)
9/22/2023	44,9 (Cukup Tidak Sehat)	43,6 (Cukup Tidak Sehat)	9/29/2023	45,3 (Cukup Tidak Sehat)	45,8 (Cukup Tidak Sehat)
9/23/2023	35,6 (Cukup Tidak Sehat)	46,9 (Cukup Tidak Sehat)	9/30/2023	58,8 (Tidak Sehat)	55,8 (Tidak Sehat)
9/24/2023	43,9 (Cukup Tidak Sehat)	41,2 (Cukup Tidak Sehat)	10/1/2023	71,3 (Tidak Sehat)	62,5 (Tidak Sehat)

KESIMPULAN DAN SARAN

Kesimpulan

Model *Hybrid* ARIMA-ANN menunjukkan performa yang luar biasa dibandingkan ARIMA dan ANN secara terpisah. Hasil penggabungan keduanya meningkatkan R^2 menjadi 0,9012, MAPE sebesar 13,603% dan RMSE sebesar 4,061 pada data *train*. Adapun metrik evaluasi data *test* menghasilkan RMSE sebesar 5,961 dan MAPE sebesar 7,622% . Hal ini mengindikasikan bahwa kemampuan model *hybrid* sangat baik sekali untuk memprediksi konsentrasi $PM_{2.5}$ selama

14 periode ke depan. Hasil penelitian ini dapat dijadikan acuan oleh pemerintah dalam memprediksi konsentrasi PM_{2.5} di DKI Jakarta selama beberapa hari ke depan dan meningkatkan kualitas kebijakan terhadap masalah polusi udara di DKI Jakarta. Dengan demikian, penelitian ini juga dapat mendorong tercapainya poin SDG's poin ke-3 mengenai kehidupan sehat dan sejahtera, poin ke-13 penanganan perubahan iklim, dan poin ke-15 mengenai ekosistem daratan.

Saran

Saran dalam penelitian ini berupa kajian model *hybrid* yang dapat diperluas lagi. Ada beberapa model *hybrid* seperti, model multiplikatif dan model Khansei-Bijari yang dapat digunakan untuk membandingkan akurasi dengan model aditif. Serta, pengembangan dapat juga berupa penambahan faktor eksternal ke dalam model *seasonal* ARIMA sehingga menjadi model *seasonal* ARIMAX.

REFERENSI

- Abebe, T. H. (2020). Time Series Analysis of Monthly Average Temperature and Rainfall Using Seasonal ARIMA Model (in Case of Ambo Area, Ethiopia). *International Journal of Theoretical and Applied Mathematics*, 6(5), 76-87.
- Ahdiat, A. (2023, 29 Maret). *Kualitas Udara di Kota Besar Indonesia Buruk, Jauh dari Standar WHO*. Diakses dari <https://databoks.katadata.co.id/datapublish/2023/03/29/kualitas-udara-di-kota-besar-indonesia-buruk-jauh-dari-standar-who>.
- Alrweili, H., & Fawzy, H. (2022). Forecasting Crude Oil Prices Using an ARIMA-ANN Hybrid Model. *Journal of Statistics Applications & Probability*, 11(3), 845-855.
- Alsuwaylimi, A. (2023). Comparison of ARIMA, ANN and Hybrid ARIMA-ANN Models for Time Series Forecasting. *Information Science Letters An International Journal*, 12(2), 1003-1016.
- Amelia, R., Mahardika, R. G., Niani, C. R., & Lewaherilla, N. (2022). Predicting Particulate Matter PM_{2.5} Using The Exponential Smoothing and Seasonal ARIMA with R studio. *IOP Conference Series: Earth and Environmental Science*, 1108(1).
- Badan Meteorologi Klimatologi dan Geofisika. Informasi Konsentrasi Partikulat (PM_{2.5}). Diakses pada Desember 1, 2023, dari <https://www.bmkg.go.id/kualitas-udara/informasi-partikulat-pm25.bmkg>
- Burki, T. (2021). WHO Introduces Ambitious New Air Quality Guidelines. *The Lancet*, 398(10306), 1117.
- Davydenko, A., & Fildes, R. (2016). Forecast Error Measures: Critical Review and Practical Recommendations. *Business forecasting: Practical problems and solutions*, 34.
- Hajirahimi, Z., & Khashei, M. (2019). Hybrid Structures in Time Series Modeling and Forecasting: A Review. *Engineering Applications of Artificial Intelligence*, 86, 83-106.
- IQAir. Diakses pada November 30, 2023, dari <https://www.iqair.com/id/indonesia/jakarta>
- Jia, J., Yuan, X., Peng, X., & Yan, B. (2019). Cr (VI)/Pb2+ Are Responsible for PM_{2.5}-Induced Cytotoxicity in A549 Cells While Pulmonary Surfactant Alleviates Such Toxicity. *Ecotoxicology and Environmental Safety*, 172, 152-158.
- Kabbilawsh, P., Sathish Kumar, D., & Chithra, N. R. (2020). Trend Analysis and SARIMA Forecasting of Mean Maximum and Mean Minimum Monthly Temperature for The State of Kerala, India. *Acta Geophysica*, 68(4), 1161-1174.
- Larasati, D. N., Faris, M., & Panuntun, S. B. (2021). *Kajian Big Data Sinyal Pemulihan Indonesia dari Pandemi Covid – 19 : Menilik Kualitas Udara di Tengah Pandemi Covid-19*. Indonesia : Badan Pusat Statistik
- Lubis, M. F. D. I. (2020). Analisis Kinerja Model Prophet untuk Peramalan Kualitas Udara DKI Jakarta. Bachelor thesis, UIN Syarif Hidayatullah Jakarta.
- Manisalidis I., Stavropoulou E., Stavropoulos A., Bezirtzoglou E. (2020). Environmental and Health Impacts of Air Pollution: A Review. *Frontiers in Public Health*, 8(2296).

- Nugroho, D. B., Wicaksono, B. A., & Larwuy, L. (2023). GARCH-X (1, 1) Model Allowing A Non-Linear Function of The Variance to Follow An AR (1) Process. *Communications for Statistical Applications and Methods*, 30(2), 163-178.
- Putra, R. E., Kalista, M., & Setianingsih, C. (2023). Klasifikasi Prediksi Kualitas Udara Menggunakan Metode Support Vector Machine (SVM). *eProceedings of Engineering*, 10(4).
- Rodríguez-Urrego, D., & Rodríguez-Urrego, L. (2020). Air Quality During The COVID-19: PM_{2.5} Analysis in The 50 Most Polluted Capital Cities in The World. *Environmental Pollution*, 266, 115042.
- Safira, M. D., Syafiuddin, A., Fasya, A. H. Z., & Setianto, B. (2022). Literature Review: Kualitas Udara di Kawasan Industri di Berbagai Lokasi di Indonesia. *Jurnal Public Health*, 9(2), 38-47.
- Sakinah, N. P., Cholissodin, I., & Widodo, A. W. (2017). Prediksi Jumlah Permintaan Koran Menggunakan Metode Jaringan Syaraf Tiruan Backpropagation. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(7), 2612–2618.
- Sari, T. I. (2019). Penerapan Artificial Neural Network Dalam Prediksi Pencemaran Udara Studi Kasus Kota Jakarta. Bachelor Thesis, Universitas Pelita Bangsa. Diakses dari <https://repo.pelitabangsa.ac.id>.
- Satria, B. (2018). Prediksi Volume Penggunaan Air PDAM Menggunakan Metode Jaringan Syaraf Tiruan Backpropagation. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 2(3), 674-684.
- Wei, W., Jiang, J., Liang, H., Gao, L., Liang, B., Huang, J., & Chen, H. (2016). Application of A Combined Model with Autoregressive Integrated Moving Average (ARIMA) and Generalized Regression Neural Network (GRNN) in Forecasting Hepatitis Incidence in Heng County, China. *PloS one*, 11(6), e0156768.
- Xue, D. M., & Hua, Z. Q. (2016). ARIMA Based Time Series Forecasting Model. *Recent Advances in Electrical & Electronic Engineering (Formerly Recent Patents on Electrical & Electronic Engineering)*, 9(2), 93-98.
- Yuberta, A. (2022). Jaringan Syaraf Tiruan dengan Algoritma Backpropagation dalam Memprediksi Hasil Asesmen Nasional Berbasis Komputer (ANBK) SMP Se Kota Sawahlunto. *Jurnal Informasi dan Teknologi*, 200–205.
- Zhai, M., Li, W., Tie, P., Wang, X., Xie, T., Ren, H., ... & Qiu, L. (2021). Research on The Predictive Effect of A Combined Model of ARIMA and Neural Networks on Human Brucellosis in Shanxi Province, China: A Time Series Predictive Analysis. *BMC Infectious Diseases*, 21(1), 1-12.
- Zhang, G. P. (2003). Time Series Forecasting Using A Hybrid ARIMA and Neural Network Model. *Neurocomputing*, 50, 159-175.