

# Pemanfaatan Pemodelan Machine Learning dalam Memprediksi Parameter Kualitas Udara Nitrogen Dioksida (NO<sub>2</sub>) Berdasarkan Algoritma Extra Trees Regression di DKI Jakarta

Muhammad Aulia Zikri<sup>1\*</sup>, I Wayan Jyesta Jaya Taruna<sup>1</sup>, Juang Merdeka<sup>1</sup> & Agung Hari Saputra<sup>1</sup>

<sup>1</sup>Program Studi Meteorologi, Sekolah Tinggi Meteorologi Klimatologi dan Geofisika, Indonesia

Corresponding author: [auliazikri1809@gmail.com](mailto:auliazikri1809@gmail.com)

## Article history

Received: 15 December 2024

Received in revised form: 25  
December 2024

Accepted: 27 December 2024

DOI:

10.17977/um0260v7i22023p031

## Kata-kata kunci:

Kualitas Udara

Nitrogen Dioksida (NO<sub>2</sub>)

Extra Trees Regression

Time Series

## Abstrak

Penelitian ini mengkaji kualitas udara di DKI Jakarta, terutama pada parameter Nitrogen Dioksida (NO<sub>2</sub>) dengan memanfaatkan model Extra Trees Regression untuk memprediksi indeks NO<sub>2</sub>. Penelitian menggunakan data time series NO<sub>2</sub> tahun 2022, yang menunjukkan tidak adanya tren jangka panjang yang signifikan serta mengindikasikan data bersifat stasioner dan acak. Analisis periodogram, histogram, dan plot Q-Q menunjukkan distribusi normal dengan penyimpangan minor. Tidak ditemukan autokorelasi yang signifikan antara data aktual NO<sub>2</sub> dan data model, menandakan kemungkinan adanya white noise. Evaluasi model dengan parameter seperti MASE, MAE, RMSE, MAPE, SMAPE, dan R<sup>2</sup> menunjukkan kinerja model yang baik. Nilai R<sup>2</sup> yang mencapai 73.14% menandakan kemampuan model dalam menjelaskan variabilitas data aktual. Meskipun model Extra Trees Regression mengikuti pola musiman, terdapat ketidaksesuaian antara nilai aktual dan prediksi di beberapa titik. Hal ini menandakan adanya potensi overfitting atau kesulitan dalam menangkap pola data secara spesifik. Penelitian ini memberikan informasi pemodelan yang cocok untuk memprediksi kualitas udara di DKI Jakarta.

## Abstract

The study examined the air quality in DKI Jakarta, mainly on the nitrogen dioxide (NO<sub>2</sub>) parameters, using the Extra Trees Regression model to predict the NO<sub>2</sub> index. The study used the 2022 NO<sub>2</sub> time series data, which showed no significant long-term trends and indicated stationary and random data. Periodogram, histogram, and Q-Q plot analysis showed normal distribution with minor deviations. No significant autocorrelation was found between actual NO<sub>2</sub> data and model data, indicating the possibility of white noise. Evaluation of models with parameters such as MASE, MAE, RMSE, MAPE, SMAPE, and R<sup>2</sup> showed good model performance. An R<sup>2</sup> value of 73.14% indicates a model's ability to explain actual data variability. Although the Extra Trees Regression model follows seasonal patterns, there is an inconsistency between actual values and predictions at some points. This indicates a potential overfitting or difficulty in capturing specific data patterns. This research provides modeling information that is suitable for predicting air quality in DKI Jakarta.

## PENDAHULUAN

Kualitas udara merupakan salah satu penyokong kehidupan yang sangat penting untuk dimonitor karena mampu memberikan dampak langsung terhadap kesehatan makhluk hidup [1]. Penurunan kualitas udara menjadi penyebab tingginya angka kematian, menurunnya kualitas kesehatan masyarakat, serta merugikan pekerjaan manusia [2], [3]. Bahkan, penurunan kualitas udara ini menyebabkan peningkatan kemungkinan kanker

paru-paru dan penyakit kardiopulmoner sekitar 4-8% (Ameer et al., 2019; Bedekar et al., 2021). Menurut Blacksmith Institute, ada dua masalah utama terkait kualitas udara di dunia, yaitu kualitas udara luar perkotaan dan polusi udara dalam ruangan [6]. Salah satu penyebab menurunnya kualitas udara adalah meningkatnya penggunaan kendaraan bermotor yang berakibat tingginya polutan yang dilepaskan kendaraan ke udara [7].

Proses penurunan kualitas udara terjadi akibat adanya penambahan zat lain di udara, seperti Partikulat Matter (PM10 dan PM2.5), Nitrogen Dioksida (NO<sub>2</sub>), Sulfur Dioksida (SO<sub>2</sub>), Ozon (O<sub>3</sub>), dan Karbon Dioksida (CO) [8]. NO<sub>2</sub> merupakan satu dari banyaknya jenis oksida nitrogen (NO<sub>x</sub>) yang dihasilkan oleh pembakaran tidak sempurna mesin kendaraan [9]. Polutan ini dapat bercampur dengan gas lainnya di atmosfer dan menyebabkan penyakit pernapasan seperti penyakit paru obstruktif kronis (COPD) yang mengakibatkan emphysema dan bronkitis kronis [10].

Penurunan kualitas udara ini lebih dikenal dengan istilah pencemaran udara. Hal ini menjadi salah satu masalah vital yang dihadapi masyarakat di berbagai negara [11]. Salah satu ibukota negara yang menarik untuk dibahas terkait kualitas udara adalah DKI Jakarta. Dengan julukan Kota Metropolitan Terbesar di Indonesia [12], DKI Jakarta memiliki data-data indeks standar pencemaran udara (ISPU) yang diobservasi dan diolah oleh Dinas Lingkungan Hidup [7]. Hingga saat ini, Dinas Lingkungan Hidup DKI Jakarta telah memantau dan mengolah data dari lima (5) Stasiun Pemantau Kualitas Udara (SPKU) yang terletak di Bundaran Hotel Indonesia (HI), Kelapa Gading, Jagaraksa, Lubang Buaya, dan Kebon Jeruk [7].

Kualitas udara di DKI Jakarta menjadi masalah serius yang harus dihadapi akibat aktivitas manusia yang semakin kompleks [13]. Selain itu, diperlukan juga kesadaran masyarakat tentang dampak yang akan ditimbulkan akibat penurunan kualitas udara [14]. Lembaga yang mengelola informasi kualitas udara memerlukan pengolahan data dan memprakirakan kualitas udara di masa yang akan datang sehingga mampu menghasilkan informasi untuk pengambilan kebijakan [15].

Ada berbagai metode untuk mengukur tingkat kualitas udara, namun beberapa metode tidak selalu tepat dan memerlukan perhitungan matematis yang kompleks [16], [17]. Sistem machine learning merupakan sistem yang banyak digunakan dan memainkan peran penting dalam penilaian dan prediksi kualitas udara [18], [19]. Algoritma machine learning telah digunakan dalam berbagai penelitian untuk mengukur indeks kualitas udara [20], [21]. Salah satu metode prediksi yang digunakan dalam meramalkan indeks kualitas udara adalah Extra Trees Regression. Pemodelan ini diperkenalkan pertama kali oleh [22]. Ini

merupakan salah satu tree-based dari pembelajaran supervised machine learning yang bisa digunakan untuk klasifikasi dan regresi. Pemodelan ini membangun pohon dengan semua sampel dan memilih point perpotongan acak untuk setiap fitur yang dipertimbangkan dibandingkan menghitung optimalisasi secara lokal [23]. Selain itu, dalam pemodelan ini seluruh sampel pelatihan digunakan untuk mengembangkan data dalam model extra trees dibandingkan replika bootstrap dalam model random forest [24].

Penelitian ini berfokus pada penilaian keakuratan model dalam memprediksi kualitas udara khususnya parameter NO<sub>2</sub> di DKI Jakarta. Selain itu penelitian ini diharapkan mampu menginterpretasikan kualitas udara NO<sub>2</sub> DKI Jakarta tahun 2022 dalam bentuk time series. Selanjutnya adalah untuk menganalisis potensi dan akurasi pemodelan Extra Trees dalam memprediksi indeks NO<sub>2</sub> DKI Jakarta. Dalam penelitian ini juga akan diidentifikasi rasio dari nilai prediksi dan nilai aktual yang dihasilkan dari model dengan dataset yang telah dibuat sebelumnya yang disajikan dalam bentuk time series.

Melalui evaluasi dan studi komprehensif dalam penelitian ini diharapkan mampu memberikan wawasan penting bagi pemangku kepentingan untuk merancang langkah strategis terkait kualitas udara khususnya di ibukota Indonesia [25].

## METODE

### Data dan Lokasi

Pada penelitian ini, wilayah DKI Jakarta yang digunakan merupakan wilayah kota administrasi tanpa adanya Kepulauan Seribu. Pembagian wilayahnya yaitu kota administrasi Jakarta Pusat dengan luas 48,13 km<sup>2</sup>, Jakarta Utara dengan luas 146,66 km<sup>2</sup>, Jakarta Barat dengan luas 129,54 km<sup>2</sup>, Jakarta Selatan dengan luas 141,37 km<sup>2</sup>, Jakarta Timur dengan luas 188,03 km<sup>2</sup>. (Jakarta, 2023)

Data yang digunakan dalam penelitian ini yaitu data yang diperoleh dari Dinas Lingkungan Hidup DKI Jakarta melalui website resminya yaitu data.jakarta.go.id yang melakukan pengukuran dan menyediakan data kualitas udara dengan menggunakan Indeks Standar Pencemaran Udara (ISPU).

Dataset ini mengambil sampel 5 SPKU DKI Jakarta dengan atribut yang dicatat adalah tanggal pengambilan data, nama stasiun pengambilan data, partikulat salah satu parameter yang diukur (PM10), sulfida dalam bentuk SO<sub>2</sub>, karbon monoksida, ozon (O<sub>3</sub>), nitrogen dioksida (NO<sub>2</sub>), nilai ukur paling tinggi dari seluruh parameter yang diukur dalam waktu yang sama (*max*), parameter yang hasil pengukurannya paling tinggi (*critical*), dan kategori [26]. NO<sub>2</sub> dipilih sebagai parameter untuk penelitian analisis model prediksi menggunakan pemodelan *extra trees regression* dalam menentukan tingkat keakuratan *forecast* model.

### Preprocessing Data

Data yang diperoleh dari Dinas Lingkungan Hidup Jakarta berbentuk csv. yang kemudian diubah menjadi bentuk excel untuk memudahkan dalam proses visualisasi data. Setelah itu data yang tidak diperlukan dalam pemodelan dihapus, yaitu nilai max, nilai kritis, kategori, dan lokasi SPKU. Data ini kemudian dibaca dalam bentuk data frame dengan menggunakan library python pandas tools. Proses running model diawali dengan menginstall package pycaret yang akan digunakan dan mengimport building package untuk running model seperti pandas, pycaret, matplotlib untuk memvisualisasikan data dan pickle library untuk menyimpan model training dalam melakukan prediksi [27].

### Analisis Data

Sebelum proses peramalan, perlu adanya konfigurasi hyperparameter pada Pycaret untuk memudahkan model dalam menentukan metode terbaik untuk data pada penelitian ini.

Tabel 1. Konfigurasi model prediksi

No	Hyperparameter	Nilai
1	Target	sNO <sub>2</sub>
2	fh	30
3	session_id	123
4	fold	3
5	verbose	true
6	remove_harmonics	false
7	seasonal_period	7
8	fold_strategy	sliding
9	max_sp_to_consider	60
10	harmonic_order_methode	harmonic_max
11	num_sps_to_use	1

Keterangan:

1. *target* merupakan variabel kualitas udara yang

akan diprediksi.

2. *fh* merupakan nilai *hyperparameter* untuk pemodelan *pycaret*.
3. *fold* merupakan teknik validasi silang (*cross validation*) melalui pembagian data menjadi sejumlah *fold* untuk mengevaluasi kinerja model.
4. *verbose* merupakan *hyperparameter* yang mengatur *log* yang dihasilkan selama proses pemodelan berlangsung. Jika *true*, maka output akan dijelaskan secara mendetail.
5. *remove\_harmonics* merujuk pada penghapusan atau tidaknya harmonisa data sebelum dianalisis.
6. *seasonal\_period* merujuk pada periode musiman dalam data.
7. *fold\_strategy* merujuk pada metode yang digunakan untuk membagi data menjadi *lipatan* dalam validasi silang.
8. *max\_sp\_to\_consider* berkaitan dengan jumlah titik data maksimum yang akan dipertimbangkan dalam analisis data.
9. *harmonic\_order\_methode* merujuk pada metode untuk menentukan urutan harmonik yang akan digunakan dalam analisis data.
10. *num\_sps\_to\_use* merujuk pada jumlah titik data yang akan dianalisis.

Penelitian ini menggunakan metode *extra trees* (ET) atau pohon sangat teracak yang merupakan metode dalam pembelajaran ensemble yang memanfaatkan pohon keputusan sebagai model dasarnya, mirip dengan algoritma *random forest* (RF). Namun, pemodelan ET memiliki ciri khas yang membedakannya dari RF, terutama dalam proses *bagging* dan seleksi fitur.

Pada proses *bagging*, ET menerapkan metode *bootstrap aggregating* seperti halnya RF. Ini melibatkan pembuatan beberapa pohon keputusan yang dilatih pada subset acak dari data pelatihan, dihasilkan melalui pengambilan sampel dengan penggantian dari kumpulan data pelatihan. Dalam hal seleksi fitur, ET memperlihatkan ciri khususnya dengan melakukan pemilihan fitur secara acak untuk setiap *split node* di setiap pohon. Berbeda dengan RT yang mencoba berbagai fitur dan memilih yang terbaik, ET tidak mengikutinya [28].

Setelah konfigurasi pada tabel 1 dijalankan, hasil prediksi dari masing-masing pohon diambil dan kemudian diintegrasikan melalui proses

optimisasi. Optimisasi dapat berupa penggunaan rerata prediksi untuk regresi atau voting mayoritas untuk klasifikasi dari semua pohon yang terlibat. Salah satu kelebihan ET adalah kemampuannya mengurangi overfitting pada model. Pendekatan acak dalam pemilihan fitur dan pembentukan pohon membantu mengatasi overfitting, terutama ketika data pelatihan mengandung noise atau terdapat fitur yang tidak relevan [29].

Meskipun terlihat kasar karena pemilihan fitur yang acak, ET dapat menunjukkan kekuatan prediktif yang baik. Beberapa pengujian empiris bahkan menunjukkan bahwa dalam beberapa situasi, ET dapat bersaing bahkan melampaui performa RT atau model ensemble lainnya. akurat dan relevan dengan perubahan yang terjadi pada deret waktu tersebut.

Model yang digunakan perlu melalui tahap evaluasi untuk mendapatkan tingkat keakuratan dari prediksi dan data aktual yang terjadi di lapangan dengan beberapa teknik yaitu MASE (Mean Absolute Scaled Error), MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), MAPE (Mean Absolute Percentage Error), SMAPE (Symmetric Mean Absolute Percentage Error),  $R^2$  (R square) yang dirumuskan pada persamaan berikut: [25], [30], [31]

$$MASE = \frac{1}{n} \sum_{i=1}^n \left( \frac{|y_i - y_l|}{\frac{1}{n-1} \sum_{i=2}^n |y_i - y_{l-1}|} \right) \quad (1)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_l - y_i| \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_l - y_i)^2} \quad (3)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - y_l|}{y_i} \times 100\% \quad (4)$$

$$SMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - y_l|}{\left(\frac{y_i + y_l}{2}\right)} \times 100\% \quad (5)$$

$$R^2 = \frac{\frac{1}{n} \sum_{i=1}^n (y_l - y_i)^2}{\frac{1}{n} \sum_{i=1}^n (y_i - y_l)^2} \quad (6)$$

Keterangan:

n = jumlah sampel data

$y_i$  = nilai sebenarnya pada sampel ke-i

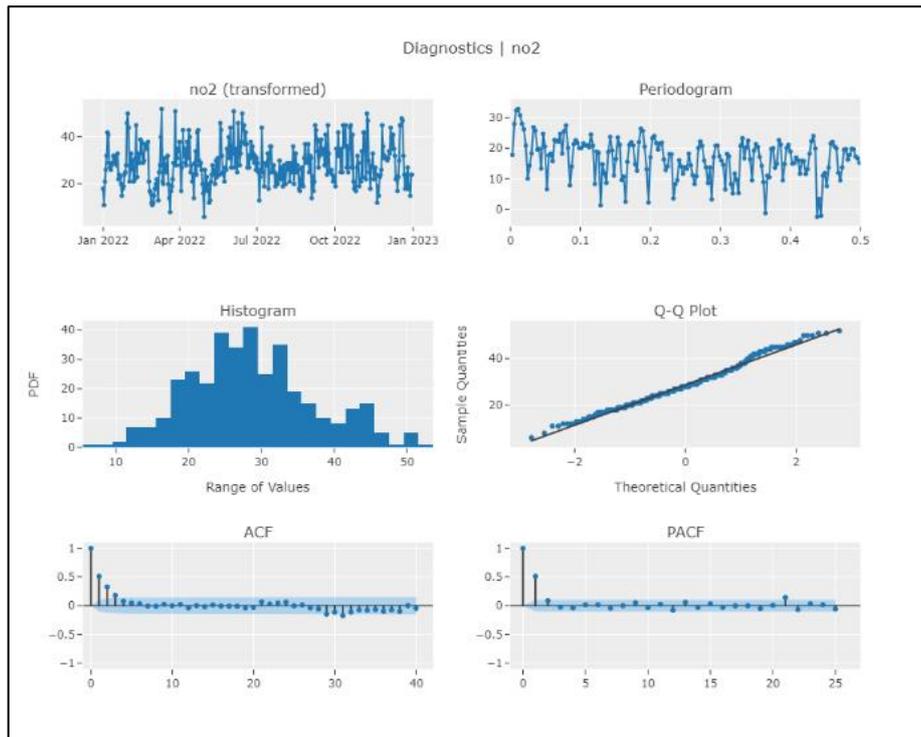
$y_l$  = nilai prediksi pada sampel ke-i

## HASIL DAN PEMBAHASAN

Pada gambar 1, plot time series yang ditransformasi ( $NO_2$  transformed) menunjukkan

konsentrasi  $NO_2$  sepanjang tahun 2022 dengan variasi data acak mengikuti suatu pola. Transformasi dilakukan untuk stabilisasi varian atau membuat data lebih mendekati distribusi normal. Pada grafik yang ditunjukkan pada gambar 1a tidak memiliki adanya tren naik atau turun jangka panjang yang jelas menginformasikan data cenderung stasioner dan bersifat acak (white noise) pada Januari 2022 hingga Desember 2022. Selain itu, grafik periodogram (gambar 1b) menunjukkan tidak adanya frekuensi dominan yang jelas. Hal ini menunjukkan bahwa tidak ada siklus musiman yang kuat dalam data. Namun, ada beberapa puncak yang menandakan frekuensi yang relevan, namun tidak menonjol secara signifikan. Berdasarkan grafik histogram (gambar 1c), data menunjukkan distribusi frekuensi nilai-nilai  $NO_2$ . Hal ini selaras dengan penelitian Kanezar, 2023 yang menyebutkan adanya beberapa puncak data yang menandakan frekuensi yang relevan. Residual data menyebar mengikuti distribusi normal yang mana terlihat bentuk histogram berbentuk seperti segitiga atau piramida tetapi dengan ekor di sisi kanan (positive skew). Hal ini dapat diamati dari adanya beberapa bar yang lebih jauh ke kanan dari modus. Grafik Q-Q plot (gambar 1d) membandingkan kuantil data yang dimiliki dengan kuantil dari distribusi normal teoritis. Karena grafik menunjukkan titik pada plot berada di sekitar garis yang menunjukkan bahwa residual data terdistribusi normal. Namun, ada beberapa penyimpangan di bagian ekor, terutama di ekor kanan (nilai positif lebih tinggi) yang mengindikasikan adanya outlier atau skewness.

Pada grafik ACF (Autocorrelation Function) dapat diamati korelasi data yang menurun cukup cepat ke arah 0 dan tetap di dalam batas kepercayaan. Hal ini mengindikasikan tidak adanya autokorelasi yang signifikan di lag-lag yang lebih tinggi dan bisa saja adanya white noise dalam data. Sedangkan pada grafik PACF (Partial Autocorrelation Function) menunjukkan adanya nilai-nilai yang signifikan pada lag awal dan kemudian menurun menjadi tidak signifikan. Berdasarkan gambar 1 dapat disimpulkan bahwa data time series  $NO_2$  tidak memiliki pola musiman yang kuat dan cenderung terdistribusi normal dengan beberapa penyimpangan. Korelasi waktu tidak signifikan dan dapat menunjukkan bahwa data cukup acak.



**Gambar 1.** Plot diagnostik data time series NO<sub>2</sub>: Gambar ini menampilkan transformasi plot time series NO<sub>2</sub> dan menampilkan data sesuai deret waktunya (kiri atas), periodogram untuk menampilkan elemen frekuensi (kanan atas), histogram menunjukkan persebaran nilai NO<sub>2</sub> (kiri tengah), plot Q-Q untuk melihat rasio persebaran nilai NO<sub>2</sub> dengan persebaran normal (kanan tengah), fungsi autokorelasi (ACF) dan *partial correlation* (PACF) untuk mendeteksi korelasi data.

Pada gambar 2a, grafik nilai aktual, menunjukkan time series asli dari konsentrasi NO<sub>2</sub> yang telah ditransformasi. Grafik ini adalah titik awal analisis dan menampilkan data yang telah diobservasi sepanjang waktu dari Januari 2022 hingga Desember 2022. Pada grafik *Seasonal Component* (gambar 2b), dapat diamati pola berulang pada interval tetap. Dalam hal ini, dinyatakan bahwa periode musimannya adalah 2 yang berarti ada pola berulang setiap dua unit waktu. Plot menunjukkan variasi musiman yang terlihat konstan sepanjang waktu dan merupakan karakteristik dekomposisi aditif. Pada grafik *trend* (gambar 2c), menunjukkan adanya tren jangka panjang dalam data dengan mengabaikan fluktuasi musiman dan acak. Dari grafik ini, ada beberapa fluktuasi dalam level NO<sub>2</sub> yang ditransformasi namun tidak menunjukkan adanya tren naik atau turun dengan jelas. Gambar paling bawah merupakan komponen residual. Hal ini mencakup fluktuasi acak dan informasi yang tidak dijelaskan oleh dua komponen lainnya. Pada grafik ini, residu berpusat di sekitar nilai 0 dengan beberapa outlier. Hal ini menunjukkan

bahwa dekomposisi telah menangkap sebagian besar struktur dalam data.

Pada tahap evaluasi model, dilakukan perhitungan beberapa parameter yang meliputi MASE, MAE, RMSE, MAPE, SMAPE, R<sup>2</sup>.

**Tabel 2.** Nilai evaluasi model

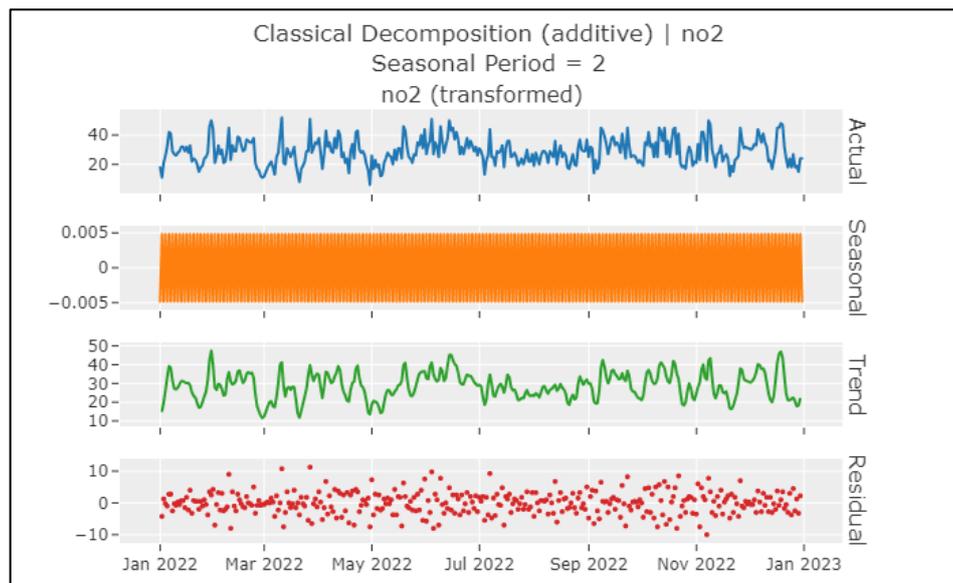
No	Parameter Evaluasi	Nilai
1	MASE	0.4855
2	MAE	3.7292
3	RMSE	4.5117
4	MAPE	0.1324
5	SMAPE	0.1305
6	R <sup>2</sup>	0.7314

Pada tabel 2, nilai MASE mengukur tingkat keakuratan prediksi relatif terhadap model naif yang hanya memprediksi nilai terakhir. Nilai di bawah 1 menunjukkan bahwa model lebih baik dibandingkan model naif. Pada pemodelan ini, nilai MASE sekitar 0.4855 menunjukkan bahwa model memiliki kinerja yang lebih baik dibandingkan model naif pada rata-rata. Nilai MAE digunakan untuk mengukur rata-rata kesalahan absolut dari

prediksi. Besaran MAE pada evaluasi model ini adalah 3.7292 yang menunjukkan tingkat kesalahan rata-rata. Nilai RMSE mencatat nilai 4.5117 yang menginterpretasikan tingkat variabilitas prediksi yang terkait dengan nilai aktual data. Ini menandakan secara rata-rata prediksi memiliki kesalahan sebesar 4.5117 unit dari nilai aktual. Nilai MAPE sebesar 13.24% menunjukkan bahwa rata-rata prediksi model menyimpang dari nilai aktual sebesar 0.1324. MAPE memberikan persepsi tentang kesalahan relatif terhadap magnitudo nilai sebenarnya. Nilai SMAPE 0.1305 mengindikasikan

kesalahan rata-rata kesalahan prediksi sekitar 13.05% dari nilai aktual. Pada pemodelan ini, nilai  $R^2$  sangat tinggi, yaitu 73.14%. Hal ini menunjukkan model telah memiliki kemampuan yang sangat baik dalam menjelaskan variabilitas data aktual.

Secara umum, model memiliki kinerja yang baik dan stabil yang dibuktikan dengan nilai parameter evaluasi model yang diharapkan.  $R^2$  yang relatif tinggi dan kesalahan yang terukur dalam kisaran wajar.

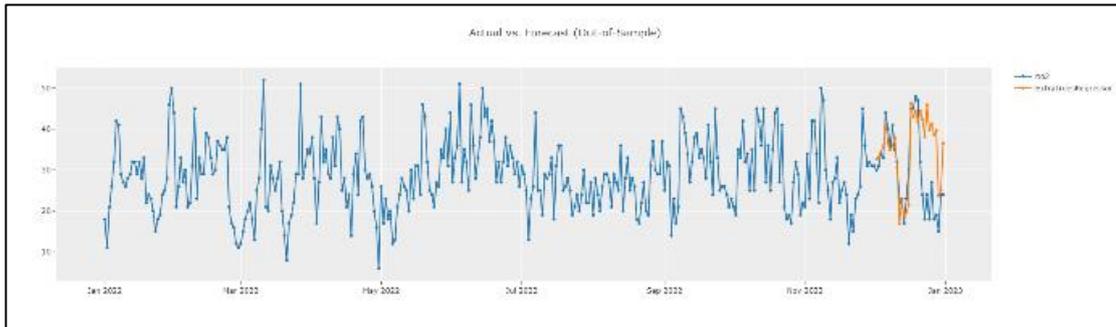


**Gambar 2.** Plot dekomposisi klasik *time series* yang terbagi menjadi 4, yaitu nilai aktual, musiman, tren, dan grafik persebaran nilai residu untuk mengidentifikasi pola yang tidak dijelaskan oleh musiman dan tren pada periode Januari 2022 hingga Desember 2022

Pada gambar 3, kemampuan model extra trees regression untuk mengikuti tren naik dan turun  $\text{NO}_2$  belum cukup baik yang dilihat dari ketidaksesuaian garis biru (nilai aktual) dan garis jingga (nilai forecast). Namun secara umum, model sudah mengikuti pola musiman yang ditandakan kesesuaian di beberapa titik nilai aktual dan nilai prediksi. Hal ini terjadi karena adanya beberapa periode waktu dimana model bisa saja mengalami

*overfitting* atau kesulitan dalam menangkap pola data secara spesifik.

Selain itu, pada awalnya grafik cukup mengikuti tren nilai aktual dengan beberapa variasi. Namun, terdapat beberapa titik dimana prediksi tidak cocok dengan data aktual yang dilihat dari posisi puncak dan lembah. Hal ini bisa terjadi karena model kesulitan dalam menangkap semua pola data.



**Gambar 3.** Perbandingan nilai aktual dan nilai *forecast* dengan pemodelan extra trees regression untuk NO<sub>2</sub> selama tahun 2022.

## KESIMPULAN

Penelitian ini menganalisis data konsentrasi NO<sub>2</sub> selama 2022 menggunakan metode time series, yang tidak menemukan tren jangka panjang signifikan, mengindikasikan data stasioner dan acak tanpa pola musiman yang kuat. Analisis statistik menunjukkan distribusi normal dengan sedikit penyimpangan dan tidak ada autokorelasi yang signifikan. Evaluasi model menggunakan parameter seperti MASE, MAE, RMSE, dan R<sup>2</sup> menunjukkan kinerja model yang baik dengan tingkat kesalahan wajar dan R<sup>2</sup> tinggi (73.14%). Pemodelan dengan extra trees regression untuk prediksi NO<sub>2</sub> menunjukkan kinerja umum yang baik meskipun ada potensi overfitting dan kesulitan dalam menangkap tren spesifik.

## DAFTAR RUJUKAN

- [1] Y. S. & F. Adnan, "Model Driving Force, Pressure, State, Impact, Response (Dpsir) Dalam Menilai Kualitas Udara," *Tekno. Lingkung.*, pp. 31–36, 2019.
- [2] H. N. Mahendra *et al.*, "Assessment and Prediction of Air Quality Level Using ARIMA Model: A Case Study of Surat City, Gujarat State, India," *Nat. Environ. Pollut. Technol.*, vol. 22, no. 1, pp. 199–210, 2023, doi: 10.46488/NEPT.2023.V22I01.018.
- [3] L. Al-Awadi, "Assessment of indoor levels of volatile organic compounds and carbon dioxide in schools in Kuwait," *J. Air Waste Manag. Assoc.*, vol. 68, no. 1, pp. 54–72, 2018, doi: 10.1080/10962247.2017.1365781.
- [4] G. R. H. Bedekar, Gayatri., Patil, R.S., Tergundi, Parimal., "An Efficient Implementation of ARIMA Technique for
- [5] S. Ameer *et al.*, "Comparative Analysis of Machine Learning Techniques for Predicting Air Quality in Smart Cities," *IEEE Access*, vol. 7, pp. 128325–128338, 2019, doi: 10.1109/ACCESS.2019.2925082.
- [6] A. Bhosale, K. Chaudhari, K. Andhe, F. Naik, and S. Chaudhari, "Analysis Of Air Pollutants Affecting The Air Quality Using ARIMA," *Int. Res. J. Eng. Technol.*, 2023, [Online]. Available: www.irjet.net.
- [7] A. Toha, P. Purwono, and W. Gata, "Model Prediksi Kualitas Udara dengan Support Vector Machines dengan Optimasi Hyperparameter GridSearch CV," *Bul. Ilm. Sarj. Tek. Elektro*, vol. 4, no. 1, pp. 12–21, 2022, doi: 10.12928/biste.v4i1.6079.
- [8] A. Wicahyo, A. Pudoli, D. Kusumaningsih, and N. Noripansyah, "Penggunaan Algoritma Naive Bayes dalam klasifikasi Pengaruh Pencemaran Udara," *ICT (Information Commun. Technol.*, vol. 20, no. 1, pp. 103–108, 2021.
- [9] M. A. S. Jawwad and Z. Oktaviani, "Rencana Pengelolaan dan Pemantauan Kualitas Udara Terhadap Rencana Kegiatan Pembangunan Permukiman," *Pros. ESEC*, vol. 4, no. 1, pp. 154–159, 2023.
- [10] I. Area, K. C. Jusoh, N. Zila, A. Hamid, and S. Side, "Peramalan Melalui Pendekatan Kalut ke atas Siri Masa Karbon Monoksida di Kawasan Perindustrian," vol. 9, no. i, pp. 55–62, 2021.
- [11] H. Zheng, Y. Cheng, and H. Li, "Investigation of model ensemble for fine-grained air quality prediction," *China*

Air Quality Prediction," *SSRN*, 2021.

- Commun.*, vol. 17, no. 7, pp. 207–223, 2020, doi: 10.23919/J.CC.2020.07.015.
- [12] D. A. Christian, A. Bachtiar, and C. Candi, “Urban Health for the Development of Healthy Cities in Indonesia,” *JKMP (Jurnal Kebijak. dan Manaj. Publik)*, vol. 11, no. 2, pp. 138–146, 2023, doi: 10.21070/jkmp.v11i2.1759.
- [13] U. F. Rizzi, Suradi, Sunaryo, and A. Agus, “Analisis Dampak Diterapkannya Kebijakan Working From Home Saat Pandemi Covid-19 Terhadap Kondisi Kualitas Udara Di Jakarta,” *J. Meteorol. ...*, vol. 6, no. 3, pp. 6–14, 2019.
- [14] D. Carruthers *et al.*, “Urban emission inventory optimisation using sensor data, an urban air quality model and inversion techniques,” *Int. J. Environ. Pollut.*, vol. 66, no. 4, pp. 252–266, 2019, doi: 10.1504/IJEP.2019.104878.
- [15] Nurjanah Siti, Siregar Amril Mutoi, and Kusumaningrum Dwi Sulistya, “Penerapan Algoritma K – Nearest Neighbor (KNN) Untuk Klasifikasi Pencemaran Udara Di Kota Jakarta,” *Sci. Student J. Information, Technol. Sci.*, vol. 1, no. 2, pp. 71–76, 2020, [Online]. Available: <http://journal.uobpkarawang.ac.id/mahasiswa/index.php/ssj/article/view/14>.
- [16] S. Ferrari, M. G. Badas, M. Garau, A. Seoni, and G. Querzoli, “The air quality in narrow two-dimensional urban canyons with pitched and flat roof buildings,” *Int. J. Environ. Pollut.*, vol. 62, no. 2–4, pp. 347–368, 2017, doi: 10.1504/IJEP.2017.089419.
- [17] J. Huang *et al.*, “A crowdsourcing-based sensing system for monitoring fine-grained air quality in urban environments,” *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3240–3247, 2019, doi: 10.1109/JIOT.2018.2881240.
- [18] T. S. Kumar, H. S. Das, U. Choudhary, P. E. Dutta, D. Guha, and Y. Laskar, “Analysis and Prediction of Air Pollution in Assam Using ARIMA/SARIMA and Machine Learning,” no. May, pp. 317–330, 2021, doi: 10.1007/978-981-16-1119-3\_28.
- [19] V. R. Pasupuleti, Uhasri, P. Kalyan, Srikanth, and H. K. Reddy, “Air Quality Prediction of Data Log by Machine Learning,” *2020 6th Int. Conf. Adv. Comput. Commun. Syst. ICACCS 2020*, pp. 1395–1399, 2020, doi: 10.1109/ICACCS48705.2020.9074431.
- [20] K. Nandini and G. Fathima, “Urban Air Quality Analysis and Prediction Using Machine Learning,” *1st Int. Conf. Adv. Technol. Intell. Control. Environ. Comput. Commun. Eng. ICATIECE 2019*, pp. 98–102, 2019, doi: 10.1109/ICATIECE45860.2019.9063845.
- [21] T. Madan, S. Sagar, and D. Virmani, “Air Quality Prediction using Machine Learning Algorithms-A Review,” *Proc. - IEEE 2020 2nd Int. Conf. Adv. Comput. Commun. Control Networking, ICACCCN 2020*, no. December 2020, pp. 140–145, 2020, doi: 10.1109/ICACCCN51052.2020.9362912.
- [22] P. Geurts, D. Ernst, and L. Wehenkel, “Extremely randomized trees,” *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, 2006, doi: 10.1007/s10994-006-6226-1.
- [23] C. Alkahfi, Z. R. Santoso, A. Fitrianto, and S. D. Oktarina, “Jurnal Sains dan Informatika,” *J. Sains dan Inform.*, vol. 4, no. 1, pp. 77–85, 2018, doi: 10.22216/jsi.v4i1.
- [24] Y. Li, T. Bao, J. Gong, X. Shu, and K. Zhang, “The Prediction of Dam Displacement Time Series Using STL, Extra-Trees, and Stacked LSTM Neural Network,” *IEEE Access*, vol. 8, pp. 94440–94452, 2020, doi: 10.1109/ACCESS.2020.2995592.
- [25] A. K. MS, M. A. Sasmita, and A. H. Saputra, “Prediksi Particulate Matter (PM 2.5) di DKI Jakarta Menggunakan XGBoost,” *J. Apl. Meteorol.*, vol. 2, no. 1, pp. 1–9, 2023, doi: 10.36754/jam.v2i1.355.
- [26] D. D. Purwanto and E. S. Honggara, “Klasifikasi Kategori Hasil Perhitungan Indeks Standar Pencemaran Udara dengan Gaussian Naïve Bayes (Studi Kasus: ISPU DKI Jakarta 2020),” *J. Intell. Syst. Comput.*, vol. 4, no. 2, pp. 102–108, 2022, doi: 10.52985/insyst.v4i2.259.

- [27] C. Sihombing, A. H. Saputra, F. P. Sari, and A. Mulya, "Prediksi Curah Hujan di Wilayah DKI Jakarta dengan Model NeuralProphet," *J. Apl. Meteorol.*, vol. 1, no. 2, 2022.
- [28] M. W. Ahmad, J. Reynolds, and Y. Rezgui, "Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees," *J. Clean. Prod.*, vol. 203, pp. 810–821, 2018, doi: 10.1016/j.jclepro.2018.08.207.
- [29] M. M. Hameed, M. K. Alomar, F. Khaleel, and N. Al-Ansari, "An Extra Tree Regression Model for Discharge Coefficient Prediction: Novel, Practical Applications in the Hydraulic Sector and Future Research Directions," *Math. Probl. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/7001710.
- [30] J. B. Habyarimana, "Forecasting Crop Production: A Seasonal Regression Model Decomposition of MAPE and SMAPE," *J. Stat. Sci. Appl.*, vol. 2, pp. 203–212, 2014.
- [31] G. Perone, "Using the SARIMA Model to Forecast the Fourth Global Wave of Cumulative Deaths from COVID-19: Evidence from 12 Hard-Hit Big Countries," *Econometrics*, vol. 10, no. 2, 2022, doi: 10.3390/econometrics10020018.