

A TREE-BASED STACKED ENSEMBLE WITH RIDGE META-LEARNER FOR NOX CONCENTRATION FORECASTING FROM IOT SENSORS

ENSEMBEL BERTUMPUK BERBASIS POHON DENGAN *META-LEARNER* RIDGE UNTUK PERAMALAN KONSENTRASI NOX DARI SENSOR IOT

Afiq Dzakwan Anasti^{1‡}, Argado Samuel Panogu¹, Alden
Nabil Wibowo Effendy¹, Berliana Destari Hermansyah¹,
Sabitala Dinika¹, and Cici Suhaeni¹

¹Program Studi Statistika dan Sains Data, SSML, IPB University

[‡]corresponding author: afiqdzwnanasti@apps.ipb.ac.id

Copyright © 2026 Afiq Dzakwan Anasti, Argado Samuel Panogu, Alden Nabil Wibowo Effendy, Berliana Destari Hermansyah, Sabitala Dinika, and Cici Suhaeni. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

The increase in transportation and industrial activities has made nitrogen oxides (NO_x) concentration in the air more difficult to predict because its pattern is dynamic, nonlinear, and influenced by environmental conditions. This study aims to apply a stacked regression model to predict NO_x concentration based on air quality sensor data from the UCI Machine Learning Repository. The research stages include data preprocessing, missing value conversion, data leakage prevention, temporal feature engineering, time-series-based data splitting, model training, performance evaluation, and model interpretation using SHAP. The proposed model consists of Multi-Layer Perceptron (MLP), Extra Trees, and XGBoost as base-learners, with RidgeCV as the meta-learner. Temporal features, including lag features and rolling mean, are added to capture the historical pattern of NO_x concentration. The results show that MLP is the best-performing base-learner, with an RMSE of 52.52, an MAE of 36.19, and an R^2 of 0.9233. The stacked regression model achieves better performance, with an RMSE of 50.73, an MAE of 33.90, and an R^2 of 0.9285. These results indicate that the stacking approach improves prediction stability and accuracy compared with a single model. The SHAP analysis shows that PT08.S2(NMHC), PT08.S4(NO₂), and NO_x(GT)_lag1 are the most influential variables in the prediction. The stacked regression model has potential as an air quality prediction approach based on IoT sensors.

Keywords: Air Quality, IoT sensors, NOx, SHAP, Stacked Regression.

Abstrak

Peningkatan aktivitas transportasi dan industri menyebabkan konsentrasi nitrogen oksida (NOx) di udara semakin sulit diprediksi karena pola perubahannya bersifat dinamis, nonlinier, dan dipengaruhi oleh kondisi lingkungan. Penelitian ini bertujuan menerapkan model *stacked regression* untuk memprediksi konsentrasi NOx berbasis data sensor kualitas udara dari UCI *Machine Learning Repository*. Tahapan penelitian meliputi prapemrosesan data, konversi nilai *missing value*, pencegahan *data leakage*, rekayasa peubah temporal, pembagian data berbasis deret waktu, pelatihan model, evaluasi performa, serta interpretasi model menggunakan SHAP. Model yang digunakan terdiri atas *Multi-Layer Perceptron* (MLP), *Extra Trees*, dan XGBoost sebagai *base-learner*, dengan RidgeCV sebagai *meta-learner*. Fitur temporal berupa *lag feature* dan *rolling mean* ditambahkan untuk menangkap pola historis konsentrasi NOx. Hasil penelitian menunjukkan bahwa MLP menjadi *base-learner* terbaik dengan RMSE sebesar 52,52, MAE sebesar 36,19, dan R^2 sebesar 0,9233. Model *stacked regression* menghasilkan performa yang lebih baik dengan RMSE sebesar 50,73, MAE sebesar 33,90, dan R^2 sebesar 0,9285. Hasil ini menunjukkan bahwa pendekatan *stacking* mampu meningkatkan stabilitas dan akurasi prediksi dibandingkan model tunggal. Analisis SHAP menunjukkan bahwa PT08.S2(NMHC), PT08.S4(NO2), dan NOx(GT)_lag1 merupakan peubah yang paling berkontribusi terhadap prediksi. Model *stacked regression* berpotensi digunakan sebagai pendekatan prediksi kualitas udara berbasis sensor IoT.

Kata Kunci: Kualitas Udara, NOx, Sensor IoT, SHAP, *Stacked Regression*.

1. Pendahuluan

Peningkatan aktivitas transportasi dan industri menyebabkan emisi nitrogen oksida (NO_x) semakin sulit dikendalikan, sehingga menurunkan kualitas udara dan membentuk ozon troposferik. Untuk mengatasinya, teknologi *Internet of Things* (IoT) digunakan karena mampu menyediakan data kualitas udara secara *real-time* dan kontinu dengan biaya rendah. Namun, data sensor IoT sangat kompleks karena memiliki *noise* tinggi, *missing value*, serta fluktuasi yang dipengaruhi faktor lingkungan. Pola konsentrasi NO_x juga bersifat non-linear dan dinamis, sehingga sulit dimodelkan dengan pendekatan statistik konvensional (Wood, 2023).

Model regresi tunggal umumnya gagal menangkap kompleksitas data deret waktu lingkungan. Oleh karena itu, algoritma *machine learning* berbasis *tree* dan *ensemble* banyak digunakan karena mampu menangani hubungan non-linear dan data heterogen dengan performa tinggi

(Prokhorenkova *et al.*, 2018). Penggunaan berbagai *tree-based models* sebagai *base-learners* bertujuan mengeksplorasi variasi pola prediksi demi meningkatkan akurasi dan kemampuan generalisasi model.

Sebagai solusi atas keterbatasan model tunggal, strategi *stacking regression* diterapkan dengan menggabungkan *output* dari berbagai *base-learners* menjadi *input* bagi *meta-learner* (Ahrens, Hansen & Schaffer, 2023). Pendekatan ini bertujuan menangkap dinamika temporal data IoT demi hasil prediksi yang lebih stabil dan akurat. *Ridge Regression* dipilih sebagai *meta-learner* karena regularisasi L2-nya mampu mengurangi dampak multikolinearitas (Duzan & Shariff, 2015).

Penelitian sebelumnya membuktikan bahwa pendekatan ensemble meningkatkan performa *forecasting* kualitas udara dibandingkan model tunggal (Ahn, Kim & Kim, 2023). Namun, penelitian yang mengombinasikan berbagai *tree-based models* dengan *Ridge Regression* sebagai *meta-learner* untuk *forecasting* NO_x berbasis IoT masih terbatas. Oleh karena itu, penelitian ini bertujuan menerapkan model *stacked ensemble* pada *forecasting* konsentrasi NO_x berbasis data sensor IoT, sekaligus menguji efektivitas *Ridge Regression* sebagai *meta-learner* dan mengevaluasi performanya.

2. Metodologi

2.1 Bahan dan Data

Penelitian ini menggunakan *Air Quality Dataset* yang diperoleh dari repositori UCI Machine Learning Repository. Dataset tersebut berisi data kualitas udara hasil pengamatan sensor gas di wilayah perkotaan yang direkam dengan interval satu jam. Total observasi awal yang digunakan adalah 7718 data. Untuk mencegah kebocoran informasi temporal (*look-ahead bias*), data deret waktu ini dibagi secara kronologis tanpa pengacakan menjadi 80% data pelatihan dan 20% data pengujian.

Peubah target dalam penelitian ini adalah konsentrasi Nitrogen Oksida ($NO_x(GT)$), sedangkan peubah prediktor dipaparkan pada Tabel 1.

Tabel 1: Peubah yang digunakan.

Peubah	Keterangan
PT08.S1 (CO)	Rata-rata sensor gas per jam (bertarget utama CO).
PT08.S2 (NMHC)	Rata-rata sensor gas per jam (bertarget utama NMHC).
PT08.S3 (NO _x)	Rata-rata sensor gas per jam (bertarget utama NO _x).
PT08.S4 (NO ₂)	Rata-rata sensor gas per jam (bertarget utama NO ₂).
PT08.S5 (O ₃)	Rata-rata sensor gas per jam (bertarget utama O ₃).
Suhu (°C) (T)	Suhu udara aktual dalam satuan derajat Celsius.
Kelembapan Relatif (%) (RH)	Kelembapan udara relatif dalam satuan persen.
Kelembapan Absolut (AH)	Kelembapan udara absolut.

2.2 Prapemrosesan Data

Rangkaian prapemrosesan dilakukan secara terintegrasi di dalam pipeline untuk menjamin kualitas data sebelum proses pelatihan model. Tahapan awal meliputi konversi atribut waktu menjadi format datetime sebagai indeks temporal, dilanjutkan dengan penyeleksian peubah numerik. Untuk menghindari bias, dilakukan penyaringan atribut dengan mengeliminasi peubah yang memiliki persentase *missing value* di atas 15%. Selain itu, pencegahan kebocoran informasi dilakukan dengan menghapus seluruh atribut prediktor lain yang mengandung akhiran *ground truth* (GT). Rekayasa peubah temporal dan normalisasi data diuraikan dalam subbab berikut.

2.2.1 Penanganan *Missing Value*

Nilai hilang berupa kode khusus -200 dikonversi menjadi *Not a Number* (NaN), lalu diimputasi menggunakan metode interpolasi linear. Pendekatan interpolasi linear dinilai efektif untuk mempertahankan kestabilan pola data deret waktu lingkungan karena tidak menghasilkan perubahan distribusi yang terlalu ekstrem dibanding metode imputasi sederhana (Flores *et al.*, 2024).

2.2.2 Standardisasi Data

Standardisasi Z-score dilakukan agar peubah memiliki rata-rata $\mu = 0$ dan standar deviasi $\sigma = 1$. Langkah ini krusial untuk menjaga stabilitas konvergensi algoritma berbasis jaringan saraf dan regresi, sekaligus mencegah dominasi peubah berskala besar terhadap proses optimasi (Xu *et al.*, 2025).

2.3 Rekayasa Peubah Temporal

Guna meningkatkan kemampuan model dalam menangkap pola temporal, dibentuk fitur turunan berbasis waktu berupa *lag feature* orde 1 dan 2, serta *rolling mean* dengan jendela 6 periode. Peubah *lag* digunakan untuk merepresentasikan pengaruh nilai historis terhadap kondisi saat ini melalui persamaan $X_t^{(lag\ k)} = X_{t-k}$. Penggunaan *lag feature* efektif dalam pemodelan kualitas udara karena konsentrasi polutan sangat dipengaruhi oleh kondisi historis sebelumnya (Xu *et al.*, 2025). Selanjutnya, peubah *rolling mean* dengan ukuran jendela 6 jam ($n = 6$) digunakan untuk menangkap tren lokal sekaligus mereduksi fluktuasi *noise* jangka pendek.

2.4 Kerangka Kerja Pemodelan (*Stacked Ensemble*)

Arsitektur *stacked ensemble regression* diterapkan menggunakan MLP, *Extra Trees*, dan XGBoost sebagai *base-learner* Level 0, dengan *Ridge Regression* sebagai *meta-learner* Level 1 untuk meminimalkan multikolinearitas (Cankurt & Subasi, 2022). Untuk mencegah *overfitting*, MLP dikonfigurasi dengan dua *hidden layer* (50, 50) dan maks 200 iterasi, *Extra Trees* dengan 100 estimator, serta XGBoost menggunakan 100 estimator, *learning rate* 0.05, dan *max depth* 6. Strategi ini memanfaatkan keragaman karakteristik ketiga model untuk menghasilkan estimasi akhir yang lebih stabil, akurat, dan memiliki generalisasi superior.

2.4.1 *Multi-Layer Perceptron* (MLP)

MLP adalah jaringan saraf tiruan *feedforward* yang efektif untuk memodelkan hubungan nonlinier kompleks pada data lingkungan multivariat (Birinci *et al.*, 2026). Model ini menggunakan dua lapisan

tersembunyi berkonfigurasi (50, 50), dengan proses pembelajaran berbasis minimisasi fungsi kerugian via *backpropagation*.

2.4.2 Extra Trees Regressor

Extra Trees membangun ensembel pohon keputusan melalui skema *randomized splitting* pada titik pecah pohon untuk menekan varians model. Model ini efektif untuk menangani hubungan nonlinier dan interaksi kompleks antarpeubah tanpa memerlukan asumsi distribusi data tertentu (Birinci *et al.*, 2026). Prediksi akhir diperoleh melalui rata-rata hitung seluruh M pohon.

2.4.3 Extreme Gradient Boosting (XGBoost)

XGBoost dipilih karena unggul dalam menangani data tabular nonlinier dan mereduksi *overfitting* melalui regularisasi, dengan efektivitas tinggi pada pemodelan kualitas udara (Birinci *et al.*, 2026). Dengan kompleksitas pohon dikontrol melalui fungsi penalti, algoritma ini mengoptimalkan pohon secara sekuensial lewat minimisasi fungsi objektif pada iterasi ke- t .

2.4.4 Stacking dan Ridge Meta-Learner

Pendekatan *stacking* digunakan untuk meningkatkan kapabilitas generalisasi melalui kombinasi model berkarakteristik berbeda, yang terbukti efektif dalam peramalan deret waktu lingkungan (Cankurt & Subasi, 2022). Estimasi final dirumuskan sebagai $\hat{y}_{final} = g(f_{MLP(x)}, f_{ET(x)}, f_{XGB(x)})$, di mana *meta-learner* (g) berbasis *Ridge Regression* mereduksi dampak multikolinearitas melalui minimisasi fungsi kerugian.

2.5 Validasi dan Evaluasi Model

Penyetelan hyperparameter pada objek RidgeCV diselesaikan melalui skema *5-fold TimeSeriesSplit* berbasis *expanding window* untuk menjaga struktur kronologis data pelatihan. Validitas prediksi akhir diuji menggunakan tiga metrik standar: MAE untuk mengukur rata-rata kesalahan absolut, RMSE untuk memberikan penalti lebih besar pada

kesalahan ekstrem, dan R^2 guna mengevaluasi kemampuan model dalam menjelaskan variasi data (Prasetyo *et al.*, 2025).

2.6 Interpretabilitas Model dengan *SHapley Additive exPlanations* (SHAP)

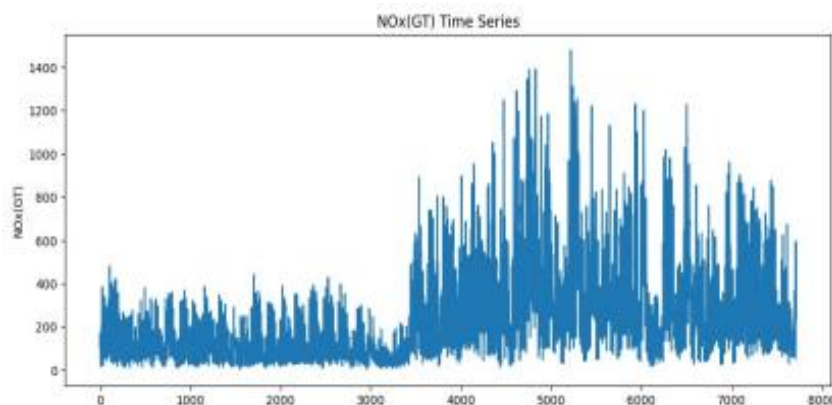
Transparansi keputusan model *stacked ensemble* dianalisis secara global dan lokal menggunakan SHAP untuk memahami kontribusi setiap peubah terhadap prediksi kualitas udara. Pendekatan ini banyak diterapkan dalam penelitian modern berbasis *environmental forecasting* dan *explainable AI* (Clement *et al.*, 2024).

3. Hasil dan Pembahasan

Bagian ini menyajikan hasil implementasi pemodelan *stacked ensemble* untuk prediksi konsentrasi NO_x , pembobotan optimal *meta-learner*, serta analisis fisis kontribusi peubah berbasis nilai SHAP.

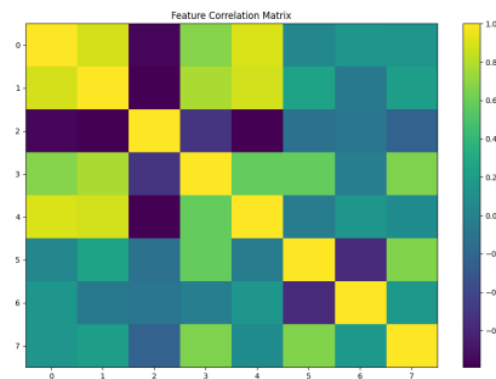
3.1 Eksplorasi Data dan Prapemrosesan

Tahap awal meliputi eksplorasi dan prapemrosesan data untuk memastikan kualitas sebelum pemodelan. Dalam dataset, nilai -200 dikonversi menjadi NaN karena merupakan kode khusus untuk *missing value*. Setelah penanganan nilai hilang selesai, seluruh kolom informasi *ground truth* dari sensor lain yang ditandai dengan akhiran "(GT)" dihapus guna mencegah terjadinya *data leakage*.



Gambar 1: Deret waktu konsentrasi $\text{NO}_x(\text{GT})$.

Visualisasi deret waktu $\text{NO}_x(\text{GT})$ pada Gambar 1 menunjukkan pola fluktuasi yang sangat dinamis. Pada paruh pertama periode pengamatan, konsentrasi NO_x umumnya berada pada kisaran rendah, yaitu 0–400 $\mu\text{g}/\text{m}^3$. Selanjutnya, konsentrasi meningkat secara drastis dan mencapai puncak sebesar 1479 $\mu\text{g}/\text{m}^3$ pada sekitar observasi ke-4000 hingga ke-5000. Pola tersebut mengindikasikan adanya komponen musiman dan autokorelasi yang kuat, sehingga pendekatan pemodelan berbasis deret waktu dinilai relevan untuk digunakan.



Gambar 2: Matriks korelasi antarpeubah prediktor.

Matriks korelasi pada Gambar 2 menunjukkan hubungan antarsensor IoT yang beragam dan non-linear, sehingga mendukung penggunaan model *machine learning* berbasis *ensemble*. Untuk menangkap pola autokorelasi temporal, dilakukan rekayasa fitur dengan menambahkan *lag* orde 1 dan 2, serta *rolling mean* jendela 6 jam. Penambahan fitur turunan ini didasarkan pada sifat konsentrasi NO_x yang persisten secara temporal, di mana pemanfaatan informasi historis terbukti relevan untuk meningkatkan akurasi prediksi.

3.2 Evaluasi Performa *Base-Learner* dan Model *Stacked Regression*

Tiga model *base-learner* dilatih menggunakan data pelatihan dan dievaluasi pada data uji. Evaluasi dilakukan menggunakan tiga metrik utama: RMSE, MAE, dan R^2 .

Tabel 2: Perbandingan performa *base-learner*.

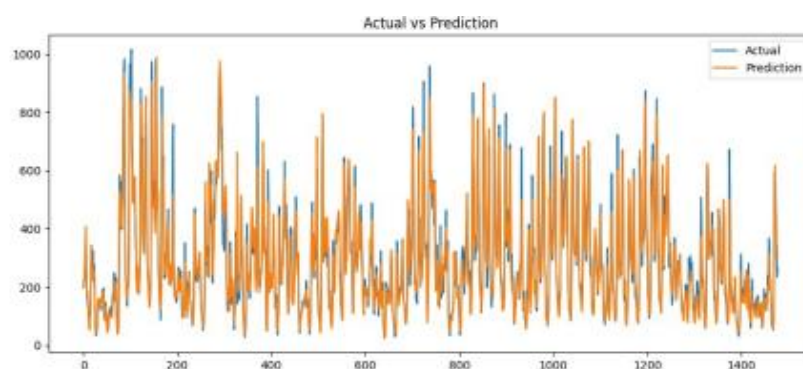
Model	RMSE	MAE	R^2
MLP	52.52	36.19	0.9233
Extra Trees	56.17	39.44	0.9123
XGBoost	56.51	39.90	0.9113

Berdasarkan Tabel 2, model MLP menunjukkan kinerja terbaik di antara ketiga *base-learner*. Secara keseluruhan, ketiga model dasar tersebut menghasilkan nilai R^2 di atas 0.91, yang mengindikasikan kemampuan sangat baik dalam menjelaskan variabilitas data uji. Keunggulan MLP menunjukkan bahwa pendekatan jaringan saraf lebih efektif dalam menangkap hubungan non-linear kompleks antara variabel sensor IoT dan konsentrasi NO_x . Karakteristik pembelajaran yang berbeda dari ketiga model ini dimanfaatkan dalam arsitektur *stacking* untuk meningkatkan generalisasi melalui keberagaman representasi pola data deret waktu kualitas udara yang kompleks.

Tabel 3: Performa final model *stacked regression* pada data uji.

Metrik Evaluasi	Nilai
RMSE	50.73
MAE	33.90
R^2	0.9285

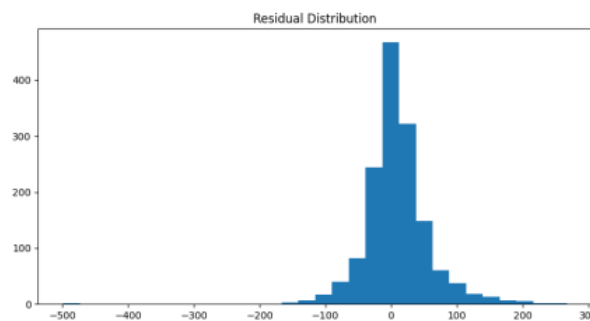
Model *stacked regression* mencapai $\text{RMSE} = 50.73$, $\text{MAE} = 33.90$, dan $R^2 = 0.9285$. Dibandingkan dengan *base-learner* terbaik (MLP, $\text{RMSE} = 52.52$), model *stacking* mampu menurunkan RMSE sebesar 1.79 poin (3.4%) dan meningkatkan R^2 dari 0.9233 menjadi 0.9285. Peningkatan ini menunjukkan bahwa *meta-learner* mampu mengoreksi pola kesalahan residual dari masing-masing *base-learner* sehingga menghasilkan prediksi akhir yang lebih stabil.



Gambar 3: Grafik perbandingan nilai aktual dan prediksi model.

Visualisasi pada Gambar 3 menunjukkan perbandingan nilai aktual dan prediksi sepanjang 1479 observasi data uji. Secara keseluruhan, kurva prediksi mampu mengikuti pola fluktuasi nilai aktual dengan baik. Meskipun terdapat kecenderungan *underprediction* pada beberapa puncak

konsentrasi ekstrem, hal tersebut merupakan karakteristik umum model *ensemble* saat menangani data yang sangat berfluktuasi.



Gambar 4: Distribusi residual model *stacked regression*.

Gambar 4 menyajikan evaluasi diagnostik distribusi residual untuk menilai pola kesalahan model *stacked regression*. Hasilnya menunjukkan sebaran mendekati normal dan terpusat di sekitar nol, dengan mayoritas pada rentang -100 hingga $+100$, yang mengindikasikan bahwa prediksi model secara umum stabil dan tidak bias sistematis. Meskipun demikian, terdapat ekor kiri yang lebih panjang akibat adanya beberapa lonjakan polusi ekstrem yang masih sulit ditangkap oleh model.

3.3 Spesifikasi *Hyperparameter* dan Pembobotan *Meta-Learner*

Penyetelan *hyperparameter* pada tingkat *base-learner* dikonfigurasi secara optimal untuk memaksimalkan kapasitas generalisasi model serta mencegah risiko *overfitting*. Spesifikasi parameter komputasi dari ketiga algoritma heterogen yang digunakan disajikan pada Tabel 4.

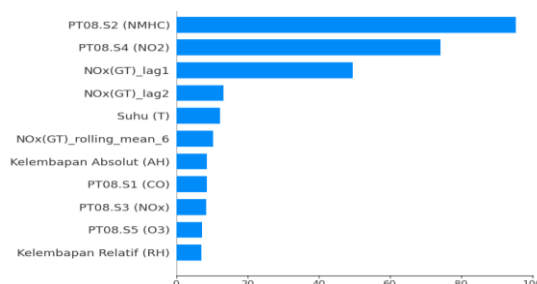
Distribusi bobot pada Tabel 4 menunjukkan dominasi signifikan dari MLP dengan kontribusi sebesar 65.3% (koefisien 0.626). Hal ini mengindikasikan bahwa aproksimasi non-linear kompleks dari jaringan saraf tiruan memiliki relevansi tertinggi menurut evaluasi linear *meta-learner*. Sebaliknya, *Extra Trees* dan XGBoost berperan sebagai pengoreksi residu dan varians tebakan. Interaksi asimetris ini menegaskan efisiensi arsitektur *stacked regression* dalam mengeksplorasi diversitas hipotesis model untuk menghasilkan prediksi akhir yang stabil dengan bias minimal.

Tabel 4: Spesifikasi *Base-Learner* dan Koefisien Bobot *Meta-Learner*.

Model Dasar	Parameter	Nilai	Representasi Fungsional	Koefisien Bobot pada <i>Meta-Learner</i>
MLP	<i>hidden_layer_sizes</i>	(50,50)	Jaringan saraf tiruan dengan dua lapisan tersembunyi.	0.625964
	<i>max_iter</i>	200	Batas maksimal iterasi konvergensi optimasi bobot.	
Extra Trees	<i>n_estimators</i>	100	Jumlah pohon keputusan penaksir dalam ansambel.	0.220146
	<i>n_estimators</i>	100	Jumlah siklus tahapan <i>boosting</i> secara sekuensial.	
XGBoost	<i>learning_rate</i>	0.05	Faktor penurunan langkah setiap pembaruan.	0.112619
	<i>max_depth</i>	6	Batas kedalaman maksimum pada struktur pohon.	

3.4 Interpretasi Model melalui Pendekatan SHAP

Hasil analisis pada Gambar 5 dengan metode SHAP menunjukkan bahwa respons sensor kimia primer, khususnya PT08.S4(NMHC) dan PT08.S5(NO₂) yang berbanding lurus dengan estimasi target, serta fitur temporal *lag* 1 yang mencerminkan persistensi dan autokorelasi NO_x, menjadi penentu paling dominan terhadap prediksi. Kombinasi fitur historis dan respons dinamis sensor ini membuktikan bahwa arsitektur *stacked regression* mampu menangkap dinamika dispersi emisi secara logis, sehingga menghasilkan prediksi valid yang dapat dipertanggungjawabkan berdasarkan prinsip kimia lingkungan.

Gambar 5: Kontribusi Peubah terhadap Model Prediksi Konsentrasi NO_x.

4. Simpulan dan Saran

Hasil penelitian menunjukkan bahwa model *stacked regression* mampu memprediksi konsentrasi NO_x berbasis sensor IoT dengan baik berkat tahapan prapemrosesan krusial seperti penanganan missing value, pencegahan *data leakage*, standardisasi, serta pembentukan fitur temporal. Di antara tiga *base-learner* yang diuji, MLP memberikan performa tunggal terbaik (RMSE = 52.52, MAE = 36.19, R^2 = 0.9233), namun penggabungannya menggunakan RidgeCV sebagai *meta-learner* berhasil meningkatkan akurasi dan stabilitas prediksi (RMSE = 50.73, MAE = 33.90, R^2 = 0.9285). Secara visual, model mampu mengikuti pola fluktuasi aktual dengan baik meskipun masih terdapat *underprediction* pada lonjakan ekstrem. Terakhir, analisis SHAP menunjukkan bahwa peubah PT08.S2(NMHC), PT08.S4(NO₂), dan NO_x(GT)_{lag1} paling berpengaruh, membuktikan bahwa prediksi sangat bergantung pada respons sensor sekaligus informasi historis data.

Pengembangan selanjutnya disarankan memperluas fitur temporal (*lag* panjang, *rolling* SD, jam/hari/musim), memperdalam tuning hiperparameter dengan validasi deret waktu, serta membandingkan performanya dengan model dependensi jangka panjang seperti *Long Short-Term Memory* (LSTM), *Gated Recurrent Unit* (GRU), atau hibrida. Terakhir, pengujian pada dataset dari lokasi dan periode berbeda juga dianjurkan guna mengevaluasi generalisasi model secara komprehensif.

Daftar Pustaka

- Ahrens, A., Hansen, C.B. & Schaffer, M.E. (2023). pystacked: Stacking generalization and machine learning in Stata. *The Stata Journal*, 23(4).
- Birinci, E., Ekmekcioğlu, Ö., Ozdemir, H., & Deniz, A. (2026). Interpretable machine learning framework for air quality prediction in Istanbul using Shapley additive explanations (SHAP). *Stochastic Environmental Research and Risk Assessment*, 40(2). <https://doi.org/10.1007/s00477-026-03168-4>
- Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. *John Wiley & Sons*.
- Cankurt, S., & Subasi, A. (2022). Tourism demand forecasting using stacking ensemble model with adaptive fuzzy combiner. *Soft Computing*, 26(7), 3455–3467. <https://doi.org/10.1007/s00500-021-06695-0>

- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference*.
- Clement, T., Nguyen, H. T. T., Kemmerzell, N., Abdelaal, M., & Stjelja, D. (2024). *Beyond explaining: XAI-based Adaptive Learning with SHAP Clustering for Energy Consumption Prediction*. <http://arxiv.org/abs/2402.04982>
- De Vito, S., Massera, E., Piga, M., Martinotto, L., & Di Francia, G. (2008). On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. *Sensors and Actuators B: Chemical*, 129(2), 750-757.
- Duzan, H. & Shariff, N.S.M. (2015). Ridge Regression for Solving the Multicollinearity Problem: Review of Methods and Models. *Journal of Applied Sciences*, 15(3), 392–404.
- Flores, A., Tito-Chura, H., Cuentas-Toledo, O., Yana-Mamani, V., & Centty-Villafuerte, D. (2024). PM2.5 Time Series Imputation with Moving Averages, Smoothing, and Linear Interpolation. *Computers*, 13(12). <https://doi.org/10.3390/computers13120312>
- Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, 63(1), 3-42.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. *Springer Science & Business Media*.
- Kaufman, S., Rosset, S., Perlich, C., & Stitelman, O. (2012). Leakage in data mining: Formulation, detection, and avoidance. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 6(4), 1-21.
- Lundberg, S. M., & Lee, Su-In. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
- Prasetyo, D. I., Rachmawardani, A., Satrio, B., & Adinara, D. B. (2025). Comparative Analysis of Machine Learning and Deep Learning Models for PM2.5 and PM10 Time Series Forecasting Using the Sisanapas Web Platform. *Teknika*, 14(3), 377–386. <https://doi.org/10.34148/teknika.v14i3.1335>
- Prokhorenkova, L. et al. (2018). CatBoost: unbiased boosting with categorical features. *NeurIPS*.
- Wolpert, D. H. (1992). Stacked generalization. *Neural networks*, 5(2), 241-259.
- Wood, D.A. (2023). Machine Learning for Hours-Ahead Forecasts of Urban Air Concentrations of Oxides of Nitrogen. *Environmental Science: Advances*, 2, 1505–1526.

Xu, Z., Zhang, H., Zhai, A., Kong, C., & Zhang, J. (2025). Stacking Ensemble Learning and SHAP-Based Insights for Urban Air Quality Forecasting: Evidence from Shenyang and Global Implications. *Atmosphere*, 16(7). <https://doi.org/10.3390/atmos16070776>