JUKOMTEK
(Jurnal Komputer Dan Teknologi)
e-ISSN 2961-9009
p-ISSN 2963-1289

ANALISIS PENGARUH SELEKSI FITUR TERHADAP KINERJA RANDOM FOREST PADA KLASIFIKASI KUALITAS UDARA

Onesimus Harefa[1], Septian Trio Sitohang[2], Glenn Desmon Sirait[3], Rado Rama Jaya Manurung[4], Jaya Tata Hardinata[5]

[1,2,3,4,5] *Universitas HKBP nommensen Pematangsiantar, Jl. Sangnawaluh No.4. Pematang Siantar, 21136*
** Email Korespondensi: onesimusharefa14@gmail.com*

INFO ARTIKEL

Sejarah Artikel:
Diterima Tgl. 14/07/2026
Diperbaiki Tgl. 23/07/2026
Disetujui Tgl. 25/07/2026
Tersedia daring Tgl. 25/07/2026

e-ISSN 2961-9009
p-ISSN 2963-1289

DOI:
https://doi.org/10.64626/jukomtek.v5i2.717

Abstract: *To reduce the impact of pollution on human health and the environment, air quality is one of the important indicators that must be monitored. The purpose of this study was to evaluate the effect of feature selection on the performance of the Random Forest algorithm on air quality classification. The data is processed through the stages of preprocessing, feature selection, model formation, and evaluation using 10-Fold Cross Validation. This data is used as an experimental method with a quantitative approach using Orange Data Mining. The results of the feature selection resulted in five main attributes: PM10, CO, NO_2, SO_2, and O_3. The Random Forest model yielded an accuracy of 76.6%, an accuracy of 0.764, a recall of 0.766, an F1 score of 0.764, a Matthews correlation coefficient of 0.689, and an area under the curve of 0.948. The results show that although feature selection succeeded in simplifying the model by reducing the number of attributes, the use of all attributes has not been able to improve the performance of Random Forest.*

Keywords:
air quality, machine learning, random forest, feature selection, classification

Abstrak: Untuk mengurangi dampak pencemaran terhadap kesehatan manusia dan lingkungan, kualitas udara adalah salah satu indikator penting yang harus dipantau. Tujuan dari penelitian ini adalah untuk mengevaluasi pengaruh seleksi fitur terhadap kinerja algoritma Random Forest pada klasifikasi kualitas udara. Data diolah melalui tahapan preprocessing, seleksi fitur, pembentukan model, dan evaluasi menggunakan 10-Fold Cross Validation. Data ini digunakan sebagai metode eksperimen dengan pendekatan kuantitatif menggunakan Orange Data Mining. Hasil seleksi fitur menghasilkan lima atribut utama: PM10, CO, NO_2, SO_2, dan O_3. Model Random Forest menghasilkan akurasi 76,6%, ketepatan 0,764, recall 0,766, skor F1 0,764, koefisien korelasi Matthews 0,689, dan area di bawah kurva 0,948. Hasil penelitian menunjukkan bahwa meskipun seleksi fitur berhasil menyederhanakan model dengan mengurangi jumlah atribut, penggunaan seluruh atribut belum mampu meningkatkan kinerja Random Forest.

Kata Kunci:
kualitas udara, machine learning, random forest, seleksi fitur, klasifikasi

E-ISSN : 2776-9666
P-ISSN : 2776-1789

PENDAHULUAN

Salah satu masalah lingkungan yang terus menjadi perhatian adalah pencemaran udara karena berdampak pada kesehatan masyarakat, aktivitas ekonomi, dan keberlanjutan lingkungan. Dibutuhkan sistem yang dapat mengidentifikasi kondisi kualitas udara secara cepat dan akurat karena peningkatan konsentrasi polutan seperti PM10, CO, NO_2, SO_2, dan O_3 dapat menurunkan kualitas udara. Machine learning memungkinkan pembuatan model klasifikasi yang dapat mengidentifikasi pola kualitas udara menggunakan data pengamatan polutan (Anisa Ma'u Luthfi & Fatkhurokhman Fauzi, 2024; Ikhsan & Musiafa, 2025; Wong et al., 2025).

Random Forest adalah salah satu algoritma yang paling banyak digunakan dalam klasifikasi karena memiliki kemampuan untuk menangani hubungan nonlinier antarvariabel (Kadir, 2026) dan (Bachtiar et al., 2026), dan tahan terhadap overfitting. Namun, menggunakan semua atribut dalam proses pembelajaran belum tentu menghasilkan model yang ideal, beberapa atribut dapat memberikan kontribusi yang kurang, sehingga menambah kompleksitas model tanpa meningkatkan performa yang signifikan. Oleh karena itu, sebelum memulai proses klasifikasi (Aditya et al., 2025; Fang et al., 2022) proses seleksi fitur merupakan langkah penting.

Penelitian sebelumnya telah menggunakan metode seleksi fitur dan Random Forest untuk mengklasifikasikan kualitas udara, dan mereka memiliki hasil yang beragam. Studi tertentu menemukan bahwa pengaturan fitur dapat meningkatkan efisiensi dan akurasi model, tetapi studi lain menemukan bahwa penurunan atribut tertentu justru menyebabkan penurunan performa karena kehilangan informasi penting (Kristiyanto & Lubis, 2026; Nur Fachmi et al., 2026; Rezk et al., 2024; Salsabila & Tessy Octavia Mukhti, 2026). Hasil yang berbeda menggambarkan bahwa faktor seleksi fitur sangat dipengaruhi oleh karakteristik dataset yang digunakan. Oleh karena itu, penelitian lebih lanjut harus dilakukan dengan dataset yang berbeda.

penelitian ini bertujuan untuk mempelajari pengaruh seleksi fitur terhadap kinerja algoritma Random Forest pada klasifikasi kualitas udara menggunakan Beijing Multi-Site Air Quality Dataset. Proses seleksi fitur dilakukan menggunakan widget Rank pada Orange Data Mining, sedangkan evaluasi model dilakukan melalui 10-Fold Cross Validation. Accuracy, Precision, Recall, F1-score, Matthews Correlation Coefficient (MCC), Area Under Curve

E-ISSN : 2776-9666
P-ISSN : 2776-1789

(AUC), dan Confusion Matrix adalah metrik yang digunakan untuk melakukan evaluasi model. Diharapkan bahwa hasil penelitian akan memberikan gambaran tentang seberapa efektif penggunaan fitur dalam meningkatkan efisiensi dan kinerja model klasifikasi kualitas udara.

LANDASAN TEORI

Kualitas udara adalah kondisi atmosfer yang dipengaruhi oleh konsentrasi berbagai polutan, termasuk PM10, CO, NO_2, SO_2, dan O_3. Jika konsentrasi polutan melebihi batas aman, itu dapat berdampak buruk baik pada lingkungan maupun kesehatan manusia. Akibatnya, klasifikasi kualitas udara merupakan salah satu langkah penting dalam proses pengambilan keputusan tentang bagaimana mengendalikan pencemaran udara. Dengan pengembangan teknologi pembelajaran mesin, proses klasifikasi dapat dilakukan secara otomatis berdasarkan pola data historis, yang memungkinkan hasil data yang lebih cepat dan akurat dibandingkan dengan metode konvensional (Anisa Ma'u Luthfi & Fatkhurokhman Fauzi, 2024; Bachtiar et al., 2026; Ikhsan & Musiafa, 2025).

Salah satu pendekatan klasifikasi yang bergantung pada pembelajaran kelompok, Algoritma Hutan Acak membangun sejumlah pohon keputusan (decision tree) secara acak, kemudian menentukan hasil klasifikasi berdasarkan suara mayoritas (majority voting). Metode ini memiliki keunggulan dalam mengurangi kemungkinan overfitting, memiliki kemampuan untuk menangani data berukuran besar, dan memberikan kinerja yang stabil untuk berbagai masalah klasifikasi. Namun, atribut yang tidak relevan dapat meningkatkan kompleksitas model, sehingga proses seleksi fitur diperlukan untuk mempertahankan atribut yang paling berkontribusi pada proses pembelajaran (Aditya et al., 2025; Kadir, 2026).

Tujuan dari seleksi fitur adalah untuk menemukan atribut yang paling memengaruhi variabel target, sehingga model menjadi lebih sederhana dan efektif tanpa menghilangkan informasi penting. Penelitian ini menggunakan widget Rank pada Orange Data Mining untuk melakukan proses seleksi fitur. Metode Gain Ratio dan Gini Index juga digunakan. Selanjutnya, metode validasi cross-fold 10-fold digunakan untuk mengevaluasi kinerja model untuk mengetahui kemampuan model untuk mengklasifikasikan setiap kategori kualitas udara (Izabi et al., 2026; Liu et al., 2024; Rezk et al., 2024). Metode validasi cross-fold ini menggunakan metrik ketepatan, akurasi, precision, recall, skor F1, Matthews Correlation Coefficient (MCC), Area Under Curve (AUC), dan Confusion Matrix.

Tabel 1. Ringkasan Penelitian Terdahulu

Peneliti	Metode	Hasil Penelitian
Ikhsan & Musiafa (2025)	Random Forest	Random Forest mampu mengklasifikasikan kualitas udara dengan tingkat akurasi yang baik.

E-ISSN : 2776-9666
P-ISSN : 2776-1789

Salsabila & Tessy Octavia Mukhti (2026)	Random Forest	Random Forest efektif digunakan pada klasifikasi ISPU berbasis data polutan.
Nur Fachmi et al. (2026)	Random Forest	Model berhasil mengidentifikasi kategori kualitas udara berdasarkan data ISPU.
Dwi Aprilyanto et al. (2025	Random Forest & Gradient Boosting	Random Forest memberikan performa yang kompetitif pada prediksi kualitas udara.
Kadir (2026)	Random Forest & CatBoost	Perbandingan menunjukkan Random Forest memiliki performa yang stabil pada klasifikasi kualitas udara.

Berdasarkan penelitian sebelumnya, algoritma Random Forest telah terbukti memiliki kinerja yang baik dalam proses klasifikasi dan prediksi kualitas udara pada berbagai jenis data. Namun, sebagian besar penelitian masih berfokus pada evaluasi kinerja algoritma atau perbandingan dengan metode lain, sementara banyak penelitian masih membahas pengaruh seleksi fitur terhadap perubahan kinerja algoritma. Oleh karena itu, penelitian ini bertujuan untuk memeriksa pengaruh penerapan seleksi fitur terhadap kinerja algoritma Random Forest pada klasifikasi kualitas udara. Dengan menggunakan sampel data kualitas udara Beijing Multi-Site, hasilnya diharapkan dapat memberikan informasi tentang seberapa efektif penerapan seleksi fitur dalam menyederhanakan model sekaligus mempertahankan kinerja klasifikasi sesuai dengan karakteristik data yang digunakan.

METODE PENELITIAN

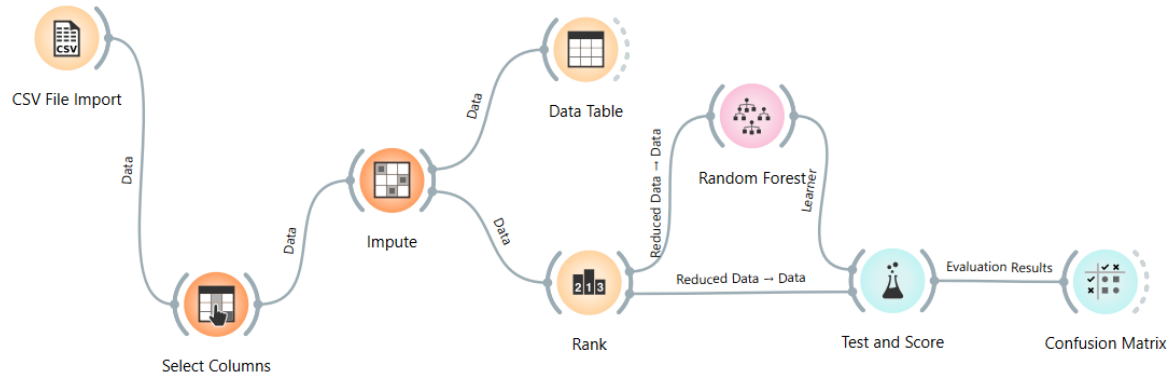
Penelitian ini menggunakan metode eksperimen dan pendekatan kuantitatif untuk mengevaluasi pengaruh seleksi fitur terhadap kinerja algoritma Random Forest pada klasifikasi kualitas udara. Beijing Multi-Site Air Quality Dataset digunakan dalam penelitian ini. Perangkat lunak Orange Data Mining digunakan untuk menyelesaikan seluruh proses analisis. Program ini menyediakan lingkungan visual untuk melakukan preprocessing, memilih fitur, membuat model, dan menilai hasil klasifikasi (Ikhsan & Musiafa, 2025; Irya et al., 2026).

Penelitian dimulai dengan mengimpor dataset menggunakan widget CSV Import File. Kemudian, menggunakan widget Select Columns, variabel target adalah Kategori_AQI. Ada lima atribut hasil seleksi: PM10, CO, NO_2, SO_2, dan O_3. Atribut identitas, seperti stasiun, digunakan sebagai meta atribut. Atribut lain yang tidak digunakan dipindahkan ke bagian Ignored. Untuk membuat data siap untuk digunakan pada tahap analisis berikutnya, widget Impute digunakan untuk menangani nilai yang hilang. Sebelum proses seleksi fitur (Irya et al., 2026) dilakukan, hasil preprocessing ditampilkan menggunakan Data Table untuk memastikan kualitas data.

E-ISSN : 2776-9666
P-ISSN : 2776-1789

Widget Rank digunakan untuk memilih fitur. Metode Gain Ratio dan Index Gini digunakan untuk menentukan atribut mana yang paling penting untuk proses klasifikasi. Dengan menggunakan hasil seleksi, lima atribut utama—PM10, CO, NO_2, SO_2, dan O_3—ditambahkan ke dalam algoritma Random Forest. Tujuan penggunaan seleksi fitur adalah untuk mengurangi atribut yang tidak relevan, sehingga model menjadi lebih sederhana dan proses pembelajaran menjadi lebih efisien tanpa menghilangkan informasi penting yang diperlukan untuk proses klasifikasi (Fang et al., 2022; Rezk et al., 2024).

Model klasifikasi dibangun menggunakan algoritma Random Forest, dan metode 10-Fold Cross Validation digunakan untuk mengevaluasi widget Test and Score. Nilai Accuracy (CA), Precision, Recall, F1-score, Matthews Correlation Coefficient (MCC), dan Area Under Curve (AUC) digunakan untuk mengukur kinerja model. Selanjutnya, data aktual dan hasil prediksi (Izabi et al., 2026; Kadir, 2026) digunakan untuk menganalisis hasil prediksi menggunakan Confusion Matrix. Ini dilakukan untuk mengetahui kemampuan model untuk mengkategorikan setiap kategori kualitas udara berdasarkan data aktual.



Gambar 1. Workflow penelitian menggunakan Orange Data Mining

HASIL DAN PEMBAHASAN

Dengan menggunakan dataset Kualitas Udara Multi-Site Beijing, penelitian ini mengkaji pengaruh seleksi fitur terhadap kinerja algoritma Random Forest dalam klasifikasi kualitas udara. Orange Data Mining, perangkat lunak yang digunakan untuk melakukan seluruh proses analisis, dilakukan sesuai dengan tahapan penelitian yang tercantum dalam metodologi. Hasil penelitian mencakup proses seleksi fitur, evaluasi model klasifikasi, analisis matriks kekacauan, dan perbandingan kinerja model sebelum dan sesudah penerapan seleksi fitur.

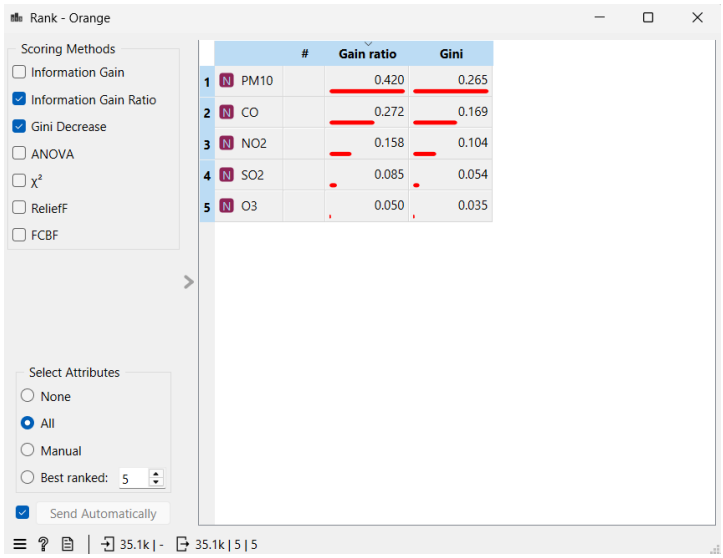
Untuk memilih fitur, widget Rank digunakan dan metode Gain Ratio dan Gini Index digunakan. Hasil seleksi menunjukkan bahwa lima atribut terpenting adalah PM10, CO, NO_2,

E-ISSN : 2776-9666
P-ISSN : 2776-1789

SO_2, dan O_3. Nilai Gain Ratio dan Gini Index untuk masing-masing atribut ditampilkan pada Tabel 2, dan hasil seleksi fitur menggunakan Orange Data Mining ditampilkan pada Gambar 2.

Tabel 2. Hasil Seleksi Fitur Menggunakan Widget Rank

Fitur	Gain Ratio	Gini
PM10	0,420	0,265
CO	0,272	0,169
NO_2	0,158	0,104
SO_2	0,085	0,054
O_3	0,050	0,035



Gambar 2. Hasil seleksi fitur menggunakan widget Rank berdasarkan Gain Ratio dan Gini Index.

Dalam proses klasifikasi kualitas udara, atribut PM10 memiliki nilai gain ratio tertinggi sebesar 0,420 dan nilai Gini tertinggi sebesar 0,265. Selanjutnya adalah sifat CO, NO_2, SO_2, dan O_3. Hasil ini menunjukkan bahwa kontribusi terhadap proses pembelajaran model berbeda untuk setiap atribut dalam dataset. Dalam proses seleksi fitur, atribut dengan kontribusi rendah dieliminasi, sehingga hanya atribut yang paling relevan digunakan untuk klasifikasi. Hasil ini sejalan dengan penelitian (Rezk et al., 2024) yang menyatakan bahwa seleksi fitur memiliki kemampuan untuk menentukan atribut penting, sehingga model menjadi lebih mudah dan efektif.

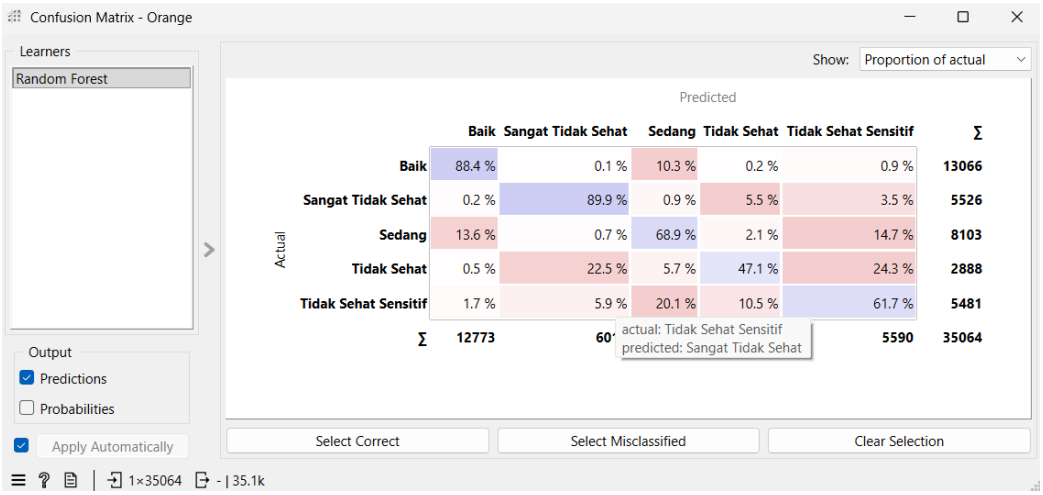
E-ISSN : 2776-9666
P-ISSN : 2776-1789

Selanjutnya, lima atribut hasil seleksi digunakan untuk membangun model Random Forest, dan hasil evaluasinya disajikan dalam Tabel 3. Hasil evaluasi model ditunjukkan dalam metode 10-Fold Cross Validation.

Tabel 3. Hasil Evaluasi Model Random Forest Setelah Seleksi Fitur

Parameter	Nilai
Accuracy	0,766
Precision	0,764
Recall	0,766
F1-score	0,764
MCC	0,689
AUC	0,948

Setelah menerapkan seleksi fitur, model Random Forest menghasilkan nilai Accuracy sebesar 76,6%, Precision sebesar 76,4%, Recall sebesar 76,6%, F1-score sebesar 76,4%, MCC sebesar 0,689, dan AUC sebesar 0,948. Nilai AUC yang mendekati satu menunjukkan bahwa model memiliki kemampuan yang luar biasa untuk membedakan setiap kategori kualitas udara. Selain itu, nilai Precision, Recall, dan F1-score yang relatif seimbang menunjukkan bahwa model memiliki kemampuan. Gambar 3 menunjukkan proses analisis yang dilakukan dengan menggunakan Confusion Matrix untuk menentukan kemampuan model untuk mengkategorikan setiap kategori kualitas udara. Nilai AUC yang mendekati satu menunjukkan untuk mengetahui distribusi hasil prediksi pada masing-masing kelas, dilakukan analisis menggunakan Confusion Matrix.



Gambar 3. Confusion Matrix hasil klasifikasi Random Forest setelah seleksi fitur

E-ISSN : 2776-9666
P-ISSN : 2776-1789

Analisis Gambar 3 menggambarkan bahwa model berhasil mengklasifikasikan kategori Baik dengan tingkat ketepatan 88,4%, kategori Sangat Tidak Sehat dengan tingkat ketepatan tertinggi 89,9%, kategori Sedang dengan tingkat ketepatan 68,9%, kategori Tidak Sehat Sensitif dengan tingkat ketepatan 61,7%, dan kategori Tidak Sehat dengan tingkat ketepatan hanya 47,1%. Dengan demikian, kategori ini menjadi kelas dengan tingkat kesalahan klasifikasi tertinggi, menurut Gambar 3. Kondisi tersebut memperlihatkan bahwa karakteristik data dalam kategori Tidak Sehat serupa dengan karakteristik kelas lain, sehingga model masih kesulitan membedakan beberapa sampel dalam kategori tersebut. Namun, secara keseluruhan, model masih dapat mengklasifikasikan sebagian besar data dengan baik. Untuk mengetahui pengaruh seleksi fitur terhadap kinerja Random Forest, dilakukan perbandingan antara model yang menggunakan seluruh atribut dan model yang menggunakan hasil seleksi fitur. Hasil perbandingan ditunjukkan pada Tabel 4.

Tabel 4. Perbandingan Kinerja Model Sebelum dan Sesudah Seleksi Fitur

Parameter	Sebelum Seleksi	Setelah Seleksi
Jumlah Fitur	11	5
Accuracy	0,824	0,766
Precision	0,825	0,764
Recall	0,824	0,766
F1-score	0,824	0,764
MCC	0,767	0,689
AUC	0,969	0,948

Hasil pengujian pada tabel 4, model menjadi lebih sederhana dan efisien karena seleksi fitur berhasil mengurangi jumlah atribut dari sebelas atribut menjadi lima. Namun, penurunan metrik evaluasi secara keseluruhan diikuti setelah penyederhanaan. Precision turun 5,8 poin persentase (dari 82,4% menjadi 76,6%), dan AUC turun 0,021 (dari 0,969 menjadi 0,948). Selain itu, terjadi penurunan pada Precision, Recall, F1-score, dan MCC. Hasil ini mengindikasikan bahwa atribut yang tidak terpilih masih menyimpan informasi yang membantu proses klasifikasi. Dengan demikian, menghapus atribut ini menurunkan kemampuan model untuk mengidentifikasi pola data.

E-ISSN : 2776-9666
P-ISSN : 2776-1789

Hasil penelitian ini menunjukkan bahwa algoritma Random Forest tidak selalu bekerja lebih baik dengan seleksi fitur. Pada set data kualitas udara Beijing Multi-Site, penggunaan seluruh atribut lebih baik daripada penggunaan lima hasil seleksi. Meskipun demikian, seleksi fitur tetap memberikan manfaat dalam menyederhanakan model melalui pengurangan jumlah atribut yang digunakan selama proses pembelajaran. Oleh karena itu, karakteristik dataset yang digunakan sangat memengaruhi efektivitas seleksi fitur. Hasil penelitian mendukung temuan penelitian sebelumnya (Dwi Aprilyanto et al., 2025; Fang et al., 2022; Kadir, 2026; Salsabila & Tessy Octavia Mukhti, 2026) yang menyatakan bahwa pengaruh seleksi fitur terhadap performa model bersifat kontekstual dan bergantung pada informasi yang ada pada setiap fitur.

KESIMPULAN

Dengan menggunakan dataset Kualitas Udara Multi-Site Beijing, penelitian ini berhasil menganalisis pengaruh seleksi fitur terhadap kinerja algoritma Random Forest pada klasifikasi kualitas udara. Hasil penelitian menunjukkan bahwa seleksi fitur mampu mengidentifikasi lima atribut utama: PM10, CO, NO_2, SO_2, dan O_3. Ini membuat model lebih sederhana karena jumlah atribut telah dikurangi dari sebelas menjadi lima. Meskipun demikian, nilai akurasi, ketepatan, pengenalan, skor F1, koefisien korelasi Matthews (MCC), dan Area Under Curve (AUC) menurun dibandingkan model yang menggunakan semua atribut setelah penyederhanaan. Penemuan ini mempertegas bahwa, pada dataset yang digunakan, atribut yang dieliminasi masih berkontribusi pada proses klasifikasi, sehingga penggunaan atribut secara keseluruhan menghasilkan kinerja yang lebih baik. Dengan demikian, penelitian ini menunjukkan bahwa efisiensi seleksi fitur dalam klasifikasi kualitas udara tidak hanya ditentukan oleh kemampuan untuk mengurangi jumlah atribut, tetapi juga oleh karakteristik data yang digunakan, sehingga hasil penelitian ini dapat menjadi pertimbangan dalam menentukan penggunaan metode seleksi fitur pada pengembangan model klasifikasi kualitas udara di masa mendatang.

DAFTAR PUSTAKA

Aditya, A., Umbara, F. R., & Sabrina, P. N. (2025). Klasifikasi Indeks Standar Pencemaran Udara Menggunakan Algoritma Catboost Dengan Teknik Balancing Data Random UnderSampling. *Jurnal Algoritma*, *22*(2). https://doi.org/10.33364/algoritma/v.22-2.2971

Anisa Ma'u Luthfi, & Fatkhurokhman Fauzi. (2024). Perbandingan Klasifikasi Random Forest, Support Vector Machines, dan LGBM Pada Klasifikasi Kualitas Udara di Jakarta. *JUSTINDO (Jurnal Sistem Dan Teknologi Informasi Indonesia)*, *9*(2), 99–108. https://doi.org/10.32528/justindo.v9i2.1912

E-ISSN : 2776-9666
P-ISSN : 2776-1789

Bachtiar, V. S., Purnawan, P., Raudina, A., & Habibah, H. R. I. (2026). Performance and application of air quality models in Indonesia: A systematic review of progress, challenges, and future directions. In *Atmospheric Environment: X* (Vol. 30). Elsevier Ltd. https://doi.org/10.1016/j.aeaoa.2026.100457

Dwi Aprilyanto, R., Gustian, R., Hendra Hernawan, M., & Surya Budiman, A. (2025). *Jurnal Sains Informatika Terapan (JSIT) ANALISIS PREDIKSI KUALITAS UDARA DENGAN METODE RANDOM FOREST BERDASARKAN DATA CUACA*. https://doi.org/10.62357/jsit.v4i3.670

Fang, L., Jin, J., Segers, A., Lin, H. X., Pang, M., Xiao, C., Deng, T., & Liao, H. (2022). Development of a regional feature selection-based machine learning system (RFSML v1.0) for air pollution forecasting over China. *Geoscientific Model Development*, *15*(20), 7791–7807. https://doi.org/10.5194/gmd-15-7791-2022

Ikhsan, N., & Musiafa, Z. (2025). *Analisis dan Klasifikasi Kualitas Udara dengan Metode Random Forest* (Vol. 10, Number 02).

Irya, A., Syukron, S., Kiswanto, D., Hafiz, A., Harahap, S. N., William, J., Ps, I. V, Baru, K., Percut, K., Tuan, S., Serdang, D., & Utara, S. (2026). Sistem Monitoring Kualitas Udara dan Peringatan Dini Berbasis IoT dengan Prediksi Polusi Menggunakan Random Forest Regression. *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, *9*(1). https://doi.org/10.32672/jnkti.v9i1.10243

Izabi, Muh. B., Annas, S., & Ahmar, A. S. (2026). Evaluating Random Forest Regression for Air Quality Prediction. *Jurnal Varian*, *9*(1), 77–84. https://doi.org/10.30812/varian.v9i1.6046

Kadir, N. (2026). Comparative Analysis of the Performance of Random Forest and CatBoost for Air Quality Prediction Based on Meteorological Factor. *Jurnal Teknik Informatika (JUTIF)*, *7*(3), 2723–3863. https://doi.org/10.52436/1.jutif.2026.7.3.5412

Kristiyanto, A., & Lubis, H. (2026). *Development of an Air Quality Classification System Using SMOTE-Based Random Forest and XAI Analysis*.

Liu, Q., Cui, B., & Liu, Z. (2024). Air Quality Class Prediction Using Machine Learning Methods Based on Monitoring Data and Secondary Modeling. *Atmosphere*, *15*(5). https://doi.org/10.3390/atmos15050553

E-ISSN : 2776-9666
P-ISSN : 2776-1789

Nur Fachmi, A., -, F. N. H., -, A. P. S., -, F. R. R., -, R. D. P., -, S. P., & -, S. P. (2026). ANALISIS KLASIFIKASI INDEKS STANDAR PENCEMARAN UDARA JAKARTA TAHUN 2025 MENGGUNAKAN ALGORITMA RANDOM FOREST. *Jurnal Informatika Dan Teknik Elektro Terapan*, *14*(1). https://doi.org/10.23960/jitet.v14i1.8477

Rezk, N. G., Alshathri, S., Sayed, A., El-Din Hemdan, E., & El-Behery, H. (2024). Sustainable Air Quality Detection Using Sequential Forward Selection-Based ML Algorithms. *Sustainability (Switzerland)*, *16*(24). https://doi.org/10.3390/su162410835

Salsabila, K., & Tessy Octavia Mukhti. (2026). Random Forest Algorithm Implementation for Air Quality Classification in DKI Jakarta Based on ISPU. *UNP Journal of Statistics and Data Science*, *4*(2), 141–149. https://doi.org/10.24036/ujsds/vol4-iss2/474

Wong, P. Y., Zeng, Y. T., Su, H. J., Lung, S. C. C., Chen, Y. C., Chen, P. C., Hsiao, T. C., Adamkiewicz, G., & Wu, C. Da. (2025). Effects of feature selection methods in estimating SO2 concentration variations using machine learning and stacking ensemble approach. *Environmental Technology and Innovation*, *37*. https://doi.org/10.1016/j.eti.2024.103996