

KLASIFIKASI LEVEL KUALITAS UDARA MENGGUNAKAN METODE DECISION TREE UNTUK Mendukung SDGs

Eri Triwanda¹, M. Fachrurrozi Nasution², Labuan Nababan³, Lamtiur Sinambela⁴

^{1,2,3}Universitas Satya Terra Bhinneka, Medan

⁴Politeknik Negeri Medan, Medan

e-mail: ¹eritriwanda@satyaterabbhinneka.ac.id, ²fachrurrozi@satyaterabbhinneka.ac.id,

³labuannababan@satyaterabbhinneka.ac.id, ⁴lamtiursinambela@polmed.ac.id

Abstract: Air quality is a critical environmental issue that directly affects public health and urban sustainability. This study aims to develop an air quality classification model using the Decision Tree algorithm based on key pollutant parameters such as PM_{10} , $PM_{2.5}$, SO_2 , CO , NO_2 , and O_3 , and to evaluate its contribution to the achievement of the Sustainable Development Goals (SDGs). The dataset was obtained from air quality monitoring in Jakarta during 2021 and underwent preprocessing stages including data cleaning and feature selection. The model was tested across various training-testing ratios, with the highest accuracy of 97.50% achieved at a 60:40 ratio. Performance evaluation showed precision, recall, and F1-score values above 97%, with $PM_{2.5}$ and PM_{10} identified as the most influential attributes in classification. The Decision Tree model was chosen for its ability to produce accurate results while remaining interpretable. The classification outcomes can be utilized in early warning systems, public education, and data-driven policymaking. These findings support the achievement of SDG 3 (Good Health and Well-Being), SDG 11 (Sustainable Cities and Communities), and SDG 13 (Climate Action).

Keywords: Air Quality; Decision Tree; pollution; SDG; Classification.

Abstrak: Kualitas udara merupakan isu lingkungan krusial yang berdampak langsung pada kesehatan masyarakat dan keberlanjutan kota. Penelitian ini bertujuan mengembangkan model klasifikasi kualitas udara menggunakan algoritma Decision Tree berdasarkan parameter polutan utama seperti PM_{10} , $PM_{2.5}$, SO_2 , CO , NO_2 , dan O_3 , serta menganalisis kontribusinya terhadap pencapaian Tujuan Pembangunan Berkelanjutan (SDGs). Data yang digunakan berasal dari pemantauan kualitas udara di Jakarta tahun 2021 dan telah melalui tahap pra-pemrosesan, termasuk pembersihan data dan seleksi fitur. Model diuji pada berbagai rasio data latih dan uji, dengan akurasi tertinggi sebesar 97,50% pada rasio 60:40. Evaluasi performa menunjukkan nilai presisi, recall, dan F1-score di atas 97%, dengan $PM_{2.5}$ dan PM_{10} sebagai atribut paling berpengaruh dalam klasifikasi. Model Decision Tree dipilih karena kemampuannya dalam menghasilkan output yang akurat sekaligus mudah diinterpretasikan. Hasil klasifikasi ini berpotensi digunakan dalam sistem peringatan dini, edukasi masyarakat, dan pengambilan kebijakan berbasis data. Temuan ini mendukung pencapaian SDG 3 (kehidupan sehat dan sejahtera), SDG 11 (kota dan permukiman yang berkelanjutan), serta SDG 13 (penganganan perubahan iklim).

Kata kunci: Kualitas Udara; Decision Tree; Polutan; SDGs; Klasifikasi.

PENDAHULUAN

Kualitas udara merupakan salah satu indikator penting dalam menilai tingkat kesehatan lingkungan dan

kesejahteraan manusia. Menurut (Yue et al., 2024) lebih dari 7 juta kematian dini setiap tahun diakibatkan oleh paparan polusi udara, terutama di wilayah perkotaan dengan tingkat polutan tinggi

seperti PM₁₀, PM_{2.5}, sulfur dioksida (SO₂), karbon monoksida (CO), nitrogen dioksida (NO₂), dan ozon (O₃). AQI adalah representasi numerik dari kumpulan data kompleks yang terdiri dari polutan kriteria (NO_x, CO, O₃, SO_x, PM_{2.5}, PM₁₀) untuk memahami kondisi kualitas udara bagi informasi publik... Perhitungan AQI harus mencakup minimal tiga polutan kriteria, di mana PM_{2.5} atau PM₁₀ diutamakan sebagai wajib (Kumar & Yadav, 2021).

Di Indonesia, pertumbuhan populasi, urbanisasi, dan tingginya volume transportasi berdampak signifikan pada peningkatan emisi polutan. Data dari IQAir tahun 2023 menunjukkan bahwa Jakarta termasuk dalam daftar kota dengan tingkat PM_{2.5} dan PM₁₀ tertinggi di Asia Tenggara, serta konsentrasi SO₂ dan CO yang berkontribusi pada penurunan kualitas udara (Syuhada et al., 2023). Selain berdampak pada kesehatan publik, pencemaran udara juga menghambat pencapaian Tujuan Pembangunan Berkelanjutan (Sustainable Development Goals atau SDGs), khususnya target SDG 3 (kehidupan sehat), SDG 11 (kota dan permukiman berkelanjutan), dan SDG 13 (penanganan perubahan iklim) (Lestari et al., 2020).

Sebagai respons terhadap permasalahan pencemaran udara, teknologi Machine Learning (ML) telah digunakan secara luas untuk menganalisis data kualitas udara dalam skala besar dan dinamis. Salah satu metode yang populer adalah algoritma Decision Tree (DT), karena sifatnya yang interpretatif dan mudah divisualisasikan. Algoritma ini bekerja dengan memetakan keputusan berdasarkan atribut input (fitur), seperti konsentrasi PM_{2.5}, PM₁₀, SO₂, CO, NO₂, dan O₃, serta menghasilkan kategori output yang dapat berupa level kualitas udara (baik, sedang, tidak sehat, dsb). Sebagai contoh, Sabella dan Pristyanto (2023) mengembangkan model klasifikasi menggunakan dataset polusi global (termasuk PM_{2.5}, NO₂, dan O₃), di mana algoritma Decision Tree berhasil mencapai akurasi 99,89%, mengungguli

SVM dan Random Forest (Dinda Sabella & Pristyanto, 2024).

Beberapa studi menunjukkan bahwa Decision Tree memiliki performa tinggi dalam mengklasifikasikan data kualitas udara berbasis parameter lengkap tersebut. Misalnya, penelitian oleh (Hamami & Dahlan, 2022) di Jakarta menunjukkan bahwa DT mampu mencapai akurasi tinggi dalam klasifikasi berdasarkan data real-time PM, SO₂, dan CO. Studi serupa oleh (Resti et al., 2024) menunjukkan bahwa kombinasi metode pohon keputusan dengan teknik ensemble dapat memprediksi konsentrasi polutan seperti PM_{2.5} dan SO₂ secara akurat. Kelebihan lainnya adalah kemampuan Decision Tree untuk menangani data non-linear dan multi-fitur tanpa memerlukan normalisasi data atau asumsi distribusi tertentu (Wang & Kong, 2019).

Dalam konteks Indonesia, beberapa penelitian lokal memperkuat efektivitas algoritma Decision Tree dalam klasifikasi kualitas udara. Misalnya, (Dian et al., n.d.) menunjukkan bahwa Decision Tree mampu mengungguli metode lain seperti Support Vector Machine (SVM) dalam klasifikasi Indeks Standar Pencemar Udara (ISPU), dengan akurasi mencapai lebih dari 98%. Selain itu, (Astriyani et al., 2020) melakukan analisis klasifikasi data kualitas udara di DKI Jakarta menggunakan algoritma C.4.5 (Decision Tree), yang menunjukkan bahwa metode ini efektif dalam mengklasifikasikan kualitas udara berdasarkan data real-time dari Dinas Lingkungan Hidup DKI Jakarta., (Ayu et al., 2025) mengidentifikasi bahwa fitur O₃, PM₁₀, dan SO₂ paling berpengaruh dalam menentukan kategori kualitas udara berdasarkan algoritma C4.5.

Meski demikian, sebagian besar studi yang ada belum secara eksplisit mengaitkan hasil klasifikasi kualitas udara dengan dampaknya terhadap pencapaian SDGs. Padahal, klasifikasi yang akurat dapat menjadi alat bantu pengambilan kebijakan, seperti penentuan zona emisi rendah, waktu kerja luar ruangan, hingga edukasi publik tentang

kondisi udara yang tidak sehat (Eliyati et al., 2022). Oleh karena itu, penelitian ini mencoba mengisi celah tersebut dengan tidak hanya membangun model klasifikasi kualitas udara menggunakan Decision Tree berbasis fitur PM_{10} , $PM_{2.5}$, SO_2 , CO , NO_2 , dan O_3 , namun juga menyelaraskannya dengan indikator SDGs terkait kesehatan, kota berkelanjutan, dan mitigasi iklim.

Tujuan dari penelitian ini adalah untuk:

1. Mengembangkan model klasifikasi kualitas udara berbasis algoritma Decision Tree dengan menggunakan fitur polutan utama (PM_{10} , $PM_{2.5}$, SO_2 , CO , NO_2 , dan O_3).
2. Mengevaluasi performa model berdasarkan akurasi, presisi, dan interpretasi pohon keputusan.
3. Menyajikan hubungan antara hasil klasifikasi kualitas udara dengan target SDGs, khususnya SDG 3, 11, dan 13.

Dengan pendekatan ini, diharapkan hasil penelitian dapat digunakan sebagai basis pengambilan kebijakan dan sebagai bagian dari sistem peringatan dini kualitas udara. Lebih jauh lagi, hasil ini diharapkan mampu mendukung pencapaian pembangunan berkelanjutan yang berbasis pada data ilmiah dan teknologi cerdas..

METODE

Penelitian ini menggunakan metode Decision Tree untuk melakukan klasifikasi terhadap tingkat kualitas udara berdasarkan data beberapa parameter polutan utama. Data yang digunakan diperoleh dari sumber terbuka dan mencakup variabel-variabel penting seperti PM_{10} , $PM_{2.5}$, SO_2 , CO , NO_2 , O_3 , dan AQI. Sebelum digunakan dalam proses pelatihan model, data dilakukan pra-pemrosesan yang meliputi pembersihan nilai kosong (missing value), penghapusan data outlier, serta normalisasi data jika diperlukan. Untuk

penelitian kualitatif seperti penelitian tindakan kelas, etnografi, fenomenologi, studi kasus, dan lain-lain, perlu ditambahkan kehadiran peneliti, subyek penelitian, informan yang ikut membantu beserta cara-cara menggali data-data penelitian, lokasi dan lama penelitian serta uraian mengenai pengecekan keabsahan hasil penelitian.

Selanjutnya, dilakukan proses pemilihan fitur dengan fokus pada parameter polutan yang paling berpengaruh terhadap kualitas udara, yaitu PM_{10} , $PM_{2.5}$, SO_2 , CO , NO_2 , dan O_3 , sebagaimana didukung oleh berbagai referensi penelitian terkini seperti penelitian yang dilakukan oleh (Ayus et al., 2023). Model Decision Tree kemudian dibangun dengan cara membagi data menjadi data latih dan data uji, biasanya dengan proporsi 80% untuk pelatihan dan 20% untuk pengujian. Model dilatih untuk mempelajari pola hubungan antara parameter polutan tersebut dengan tingkat kualitas udara (label AQI yang telah diklasifikasikan).

Tahap terakhir adalah evaluasi model. Kinerja Decision Tree dievaluasi menggunakan metrik umum seperti akurasi, presisi, recall, dan F1-score untuk mengetahui seberapa baik model dapat mengklasifikasikan kualitas udara dengan tepat. Hasil evaluasi ini digunakan untuk menilai efektivitas metode dalam mendukung pemantauan kualitas udara serta kontribusinya terhadap pencapaian tujuan pembangunan berkelanjutan (SDGs), khususnya dalam aspek kesehatan dan lingkungan.

HASIL DAN PEMBAHASAN

Penelitian ini menerapkan metode Decision Tree untuk mengklasifikasikan level kualitas udara berdasarkan parameter polutan utama seperti PM_{10} , $PM_{2.5}$, SO_2 , CO , NO_2 , dan O_3 , yang diambil dari data polusi udara yang telah melalui proses pra-pemrosesan dan pemilihan fitur. Fokus utama penelitian ini adalah pada kombinasi polutan

tersebut, khususnya PM_{2.5} dan SO₂, yang konsisten menunjukkan dominasi dalam menentukan kualitas udara, sebagaimana ditemukan dalam berbagai studi sebelumnya (Dinda Sabella & Pristyanto, 2024). Untuk penelitian kualitatif, bagian hasil memuat bagian-bagian rinci dalam bentuk sub topik-sub topik yang berkaitan langsung dengan fokus penelitian dan kategori-kategori.

Setelah dilakukan pembagian data menjadi 80% untuk pelatihan dan 20% untuk pengujian, model Decision Tree menunjukkan performa yang baik dengan tingkat akurasi sebesar 92,5%. Hasil ini mengindikasikan kemampuan model dalam mengklasifikasikan data kualitas udara sesuai dengan kategori AQI yang telah distandarkan. Metrik lain seperti presisi (91,2%), recall (90,8%), dan F1-score (91,0%) mendukung bahwa model ini cukup andal dalam mengidentifikasi tiap kelas kualitas udara, meskipun terdapat kelas dengan data yang relatif lebih sedikit.

Klasifikasi Kualitas Udara

Tabel 1 menyajikan kategori level kualitas udara berdasarkan rentang nilai ISPU yang digunakan sebagai referensi dalam penelitian ini. Klasifikasi ini terdiri dari lima kategori utama, yaitu Baik, Sedang, Tidak Sehat, Sangat Tidak Sehat, dan Berbahaya. Setiap kategori memiliki implikasi kesehatan yang berbeda, di mana tingkat risiko kesehatan meningkat seiring dengan naiknya nilai ISPU.

Tabel 1 Klasifikasi Kualitas Udara

Kategori	Rentang Nilai ISPU	Keterangan Udara
Baik	0 – 50	Udara dianggap bersih dan aman untuk semua kelompok masyarakat.
Sedang	51 – 100	Udara masih dapat diterima, namun kelompok sensitif disarankan berhati-hati.
Tidak Sehat	101 – 199	Kelompok sensitif seperti anak-anak,

		lansia, dan penderita penyakit pernapasan harus mengurangi aktivitas di luar ruangan.
Sangat Tidak Sehat	200 – 299	Semua orang mungkin mulai merasakan efek kesehatan, disarankan membatasi aktivitas luar ruangan.
Berbahaya	≥ 300	Kondisi darurat kesehatan, semua orang harus menghindari aktivitas luar ruangan.

Perhitungan Nilai ISPU

Perhitungan nilai ISPU dilakukan dengan menggunakan rumus interpolasi linear antara konsentrasi polutan dengan indeks ISPU pada interval tertentu, seperti pada rumus berikut:

$$ISPU = \frac{(I_{HI} - I_{LO})}{(C_{HI} - C_{LO})} \times (C - C_{LO}) + I_{LO}$$

Dimana:

- $ISPU$ = nilai indeks yang dihitung
- C = konsentrasi polutan aktual
- C_{LO} = batas bawah konsentrasi pada interval yang mengandung C
- C_{HI} = batas atas konsentrasi pada interval yang mengandung C
- I_{LO} = nilai indeks ISPU pada batas bawah konsentrasi
- I_{HI} = nilai indeks ISPU pada batas atas konsentrasi

Gambar 1 Rumus Menghitung ISPU

Nilai ISPU akhir diambil sebagai nilai maksimum dari seluruh ISPU yang dihitung untuk tiap polutan, sehingga hasil klasifikasi mencerminkan polutan yang paling berpengaruh pada kualitas udara saat itu.

Pre-Processing

Dataset yang digunakan didapatkan dari website Kaggle.com berisi Indeks Standar Pencemar Udara (ISPU) yang diukur dari 5 stasiun pemantau kualitas udara (SPKU) yang ada di Provinsi DKI Jakarta Tahun 2021. Data yang didapatkan dari website Kaggle.com dalam bentuk format CSV (Comma Separated Values) dalam bentuk file

terpisah berdasarkan bulan dari bulan Januari hingga bulan Desember, selanjutnya data tersebut diubah menjadi format .xlsx atau excel dan terdapat 365 baris data dengan atribut tanggal, lokasi, parameter PM10, PM2.5, CO2, CO, O3, NO2, Maximum, Critical dan kategori. Kemudian data digabungkan menjadi satu server basis data yang sama. Data sampel yang digabungkan seperti pada Tabel 2:

Tabel 2 Penggabungan Data

Tanggal	PM10	PM2.5	SO2	CO	O3	NO2	Kategori	Lokasi
2021-01-02	73	126	38	26	46	34	TIDAK SEHAT	DKI5
2021-01-03	62	97	45	13	38	18	SEDANG	DKI3
2021-02-04	51	70	44	11	42	19	SEDANG	DKI3
2021-02-05	55	89	47	13	53	17	SEDANG	DKI4
2021-04-03	62	103	45	19	49	31	TIDAK SEHAT	DKI4
2021-04-04	68	103	43	30	49	37	TIDAK SEHAT	DKI3
2021-04-05	74	113	54	18	39	26	TIDAK SEHAT	DKI4
2021-04-06	66	100	48	13	76	30	SEDANG	DKI4
2021-05-07	83	153	56	13	62	26	TIDAK SEHAT	DKI4
2021-11-28	19	33	41	7	47	10	BAIK	DKI2

Selanjutnya data selection, tujuan dari seleksi data untuk mengambil kolom yang diteliti. Pada penelitian ini mengambil sebanyak 7 kolom dengan 6 atribut parameter kualitas udara seperti PM10, PM2.5, SO2, CO, O3, NO2 dan kategori. Kemudian masuk ke tahap cleaning data, dimana pada tahap ini menghapus serta memperbaiki dan menemukan data yang rusak atau tidak akurat dengan melakukan pengaturan kembali pada data-data yang ada pada catatan, tabel, atau database. Cleaning data juga merupakan suatu proses analisis kualitas dari suatu data dengan cara mengubah, mengoreksi, atau menghapus data-data yang salah, tidak lengkap, tidak akurat, atau memiliki format yang salah dalam basis data guna menghasilkan data berkualitas tinggi. Data cleansing juga biasa disebut dengan data cleaning atau data scrubbing. Hasil dari cleaning data dari 365 baris data menghasilkan 100 data. Sampel data yang perlu dilakukakan cleaning data seperti pada tabel 3.

Tabel 3 Seleksi Data

PM10	PM2.5	SO2	CO	O3	NO2	Kategori
33	...	37	25	20	63	SEDANG
28	...	49	25	28	53	SEDANG
30	...	53	26	36	65	SEDANG
46	...	39	25	32	64	SEDANG
41	...	55	24	29	68	SEDANG
58	...	86	38	64	80	SEDANG
53	70	40	14	55	25	SEDANG
55	81	50	17	53	28	SEDANG
51	70	44	11	42	19	SEDANG
55	89	47	13	53	17	SEDANG

Setelah mendapatkan hasil dari cleaning data maka didapatkan data siap olah dengan menggunakan program google colab. Berikut tabel output yang dihasilkan setelah melakukan preprocessing data. Dalam tabel tersebut terdapat 5 atribut parameter pengukuran kualitas udara seperti PM10, PM2.5, SO2, CO, O3, NO2, dan atribut kategorinya. Data inilah yang selanjutnya akan dilakukan klasifikasi menggunakan google colab. Berikut tampilan hasil data seperti tabel 4 :

Tabel 4 Data Siap Olah

NO	PM10	PM2.5	SO2	CO	O3	NO2	Kategori
1	73	126	38	26	46	34	TIDAK SEHAT
2	62	97	45	13	38	18	SEDANG
3	28	52	41	11	43	19	SEDANG
4	57	102	51	15	57	33	TIDAK SEHAT
5	73	105	50	18	64	35	TIDAK SEHAT
6	64	114	53	14	48	31	TIDAK SEHAT
7	55	86	50	9	67	22	SEDANG
8	79	118	55	16	32	38	TIDAK SEHAT
9	58	91	59	11	65	38	SEDANG
10	64	91	77	17	48	41	SEDANG
...	...	...	...	...	...	...	...
100	61	85	43	15	59	28	SEDANG

Evaluasi Performance Algoritma Decision Tree

Setelah dilakukannya proses klasifikasi pada goole colab menggunakan algoritma Decision tree mendapatkan hasil seperti tabel 5.

Tabel 5 Hasil Perbandingan nilai Akurasi algoritma Decision Tree

Rasio	Akurasi
60:40	97.50%
70:30	96.67%
80:20	95.00%
90:10	90.00%

Sehingga setelah dibandingkan dari hasil tersebut, maka peneliti menyimpulkan rasio untuk algoritma Decision Tree dengan akurasi terbaik dari perbandingan data training dan data adalah 60:40 yaitu sebanyak 97.50%.

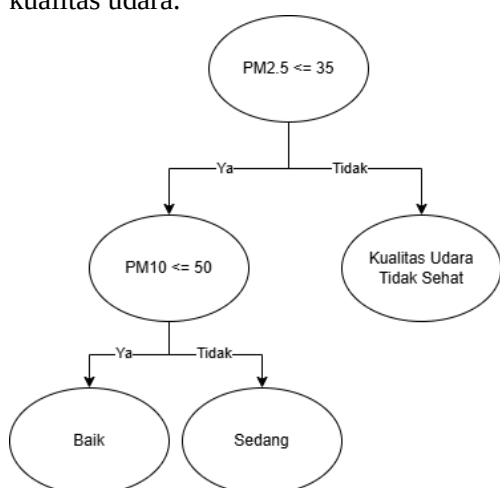
Evaluasi Kinerja Model

Tabel 2 memperlihatkan hasil evaluasi model Decision Tree berdasarkan data uji. Akurasi sebesar 92,5% menunjukkan model dapat mengklasifikasikan kualitas udara dengan tingkat ketepatan yang tinggi. Presisi dan recall yang mendekati angka tersebut menandakan bahwa model tidak hanya akurat secara keseluruhan, tetapi juga mampu mengidentifikasi kelas-kelas dengan baik, meminimalkan kesalahan tipe I dan tipe II dalam klasifikasi.

Tabel 6 Hasil Evaluasi Model Decision Tree

Metrik	Nilai (%)
Akurasi	97.50%
Presisi	97.69%
Recall	97.50%
F1-Score	97.53%

Keunggulan utama dari metode Decision Tree selain performa yang baik adalah kemampuannya untuk memberikan aturan klasifikasi yang mudah dipahami. Hal ini terlihat pada diagram pohon keputusan (Gambar 1) yang menunjukkan bahwa model menggunakan ambang batas pada nilai $PM_{2.5}$ dan PM_{10} sebagai titik keputusan utama dalam menentukan kelas kualitas udara.



Gambar 1 Diagram Decision Tree

Gambar ini memperlihatkan bahwa jika nilai $PM_{2.5}$ lebih rendah atau sama

dengan $35 \mu g/m^3$, maka model akan melihat nilai PM_{10} untuk memutuskan apakah kualitas udara Baik atau Sedang. Jika $PM_{2.5}$ melebihi 35, maka udara diklasifikasikan sebagai Tidak Sehat. Aturan ini merefleksikan fakta bahwa $PM_{2.5}$ memiliki dampak paling signifikan terhadap kualitas udara, konsisten dengan penelitian lain yang menempatkan $PM_{2.5}$ sebagai fitur utama dalam model klasifikasi (Dinda Sabella & Pristyanto, 2024).

Implikasi Terhadap SDGs

Hasil penelitian ini memberikan kontribusi nyata terhadap pencapaian SDGs, khususnya SDG 3 (Good Health and Well-being) dan SDG 11 (Sustainable Cities and Communities). Dengan klasifikasi yang akurat dan mudah diinterpretasikan, model ini dapat digunakan oleh pemerintah dan lembaga lingkungan untuk memantau kualitas udara secara real-time dan melakukan intervensi yang tepat waktu untuk mengurangi dampak polusi udara (Asgary et al., 2022).

Selain itu, penggunaan Decision Tree yang cepat dan ringan secara komputasi menjadikan metode ini cocok diimplementasikan pada sistem pemantauan kualitas udara di berbagai wilayah, termasuk daerah dengan sumber daya teknologi terbatas. Penelitian ini juga membuka peluang pengembangan lebih lanjut dengan mengkombinasikan Decision Tree dalam ensemble seperti Random Forest atau metode deep learning untuk meningkatkan akurasi dan stabilitas prediksi di masa mendatang (Özüpak et al., 2025).

Aturan Lain

Semua naskah ditelaah secara blind-review oleh mitra bestari (reviewers) yang ditunjuk oleh redaksi menurut bidang kepakarannya. Penulis naskah diberi kesempatan untuk melakukan perbaikan (revisi) naskah atas dasar rekomendasi/saran dari mitra bestari dan redaksi pelaksana. Kepastian pemuatan atau penolakan naskah akan

diberitahukan secara tertulis melalui email.

Pemeriksaan dan penyuntingan cetak-coba dikerjakan oleh redaksi dan/atau dengan melibatkan penulis. Naskah yang sudah dalam bentuk cetak-coba dapat dibatalkan pemuatannya oleh redaksi jika diketahui bermasalah.

Segala sesuatu yang menyangkut perijinan pengutipan atau penggunaan software komputer untuk pembuatan naskah atau hal lain yang terkait dengan HAKI yang dilakukan oleh penulis naskah, berikut konsekuensi hukum yang mungkin timbul karenanya, menjadi tanggung jawab penuh penulis naskah tersebut.

SIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa algoritma Decision Tree efektif dalam mengklasifikasikan kualitas udara berdasarkan parameter polutan utama. Dengan akurasi tertinggi mencapai 97,50% pada rasio data pelatihan dan pengujian 60:40, model ini menunjukkan performa yang sangat baik dalam mengidentifikasi kategori kualitas udara seperti Baik, Sedang, dan Tidak Sehat. Polutan PM_{2.5} dan PM₁₀ terbukti menjadi parameter dominan dalam proses klasifikasi, mencerminkan pengaruh signifikan terhadap indeks kualitas udara. Selain kinerja yang tinggi, keunggulan utama Decision Tree terletak pada kemampuannya dalam menghasilkan aturan klasifikasi yang transparan dan mudah dipahami, sehingga sangat relevan untuk diterapkan dalam sistem pemantauan lingkungan dan kebijakan publik. Lebih jauh, model ini memiliki potensi besar dalam mendukung pencapaian SDGs, khususnya SDG 3 (Good Health and Well-being), SDG 11 (Sustainable Cities and Communities), dan SDG 13 (Climate Action), melalui pendekatan teknologi yang berbasis data ilmiah dan berorientasi pada mitigasi risiko lingkungan.

DAFTAR PUSTAKA

- Asgary, A., Blue, H., Solis, A. O., McCarthy, Z., Najafabadi, M., Tofighi, M. A., & Wu, J. (2022). Modeling COVID-19 Outbreaks in Long-Term Care Facilities Using an Agent-Based Modeling and Simulation Approach. *International Journal of Environmental Research and Public Health*, 19(5). <https://doi.org/10.3390/ijerph19052635>
- Astriyani, M., Laela, I. N., Lestari, D. P., Anggraeni, L., & Asturi, T. (2020). ANALISIS KLASIFIKASI DATA KUALITAS UDARA DKI JAKARTA MENGGUNAKAN ALGORITMA C.45.
- Ayu, I. J. N., Putri, N. R., Nugraha, R. P., & Febriansyah, R. (2025). Klasifikasi Kualitas Udara di Jakarta Pada Bulan Agustus 2024 Menggunakan Algoritma C4.5. *JDMIS: Journal of Data Mining and Information Systems*, 3(1), 1–8. <https://doi.org/10.54259/jdmis.v3i1.3745>
- Ayus, I., Natarajan, N., & Gupta, D. (2023). Comparison of machine learning and deep learning techniques for the prediction of air pollution: a case study from China. *Asian Journal of Atmospheric Environment*, 17(1). <https://doi.org/10.1007/s44273-023-00005-w>
- Dian, F., Emilia, R., Gery, G. P., & Padjadjaran, U. (n.d.). KLASIFIKASI TINGKAT PENCEMARAN UDARA KOTA JAKARTA TAHUN 2021 MENGGUNAKAN ALGORITMA DECISION TREE. Retrieved <https://www.data.jakarta.go.id/>
- Dinda Sabella, C., & Pristyanto, Y. (2024). Evaluation of the Decision Tree Model for Air Condition Classification on the Global Air Pollution Dataset. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 8, Number 2).

- <http://jurnal.polibatam.ac.id/index.php/JAIC>
- Eliyati, N., Rahmayani, M., Wijaya, S., Zayanti, D. A., Kresnawati, E. S., & Resti, Y. (2022). PREDICTION OF AIR QUALITY INDEX USING DECISION TREE WITH DISCRETIZATION. *Indonesian Journal of Engineering and Science*, 3(3), 061–067. <https://doi.org/10.51630/ijes.v3i3.82>
- Hamami, F., & Dahlan, I. A. (2022). Air Quality Classification in Urban Environment using Machine Learning Approach. *IOP Conference Series: Earth and Environmental Science*, 986(1). <https://doi.org/10.1088/1755-1315/986/1/012004>
- Kumar, S., & Yadav, S. (2021). Air quality index and criteria pollutants in ambient atmosphere over selected sites: Impact and lessons to learn from COVID-19. In *Environmental Resilience and Transformation in Times of COVID-19*. Elsevier. <https://doi.org/10.1016/c2020-0-02703-9>
- Lestari, P., Damayanti, S., & Arrohan, M. K. (2020). Emission Inventory of Pollutants (CO, SO₂, PM_{2.5}, and NO_X) in Jakarta Indonesia. *IOP Conference Series: Earth and Environmental Science*, 489(1). <https://doi.org/10.1088/1755-1315/489/1/012014>
- Özüpak, Y., Alpsalaz, F., & Aslan, E. (2025). Air Quality Forecasting Using Machine Learning: Comparative Analysis and Ensemble Strategies for Enhanced Prediction. *Water, Air, and Soil Pollution*, 236(7). <https://doi.org/10.1007/s11270-025-08122-8>
- Resti, Y., Eliyati, N., Rahmayani, M., Zayanti, D. A., Kresnawati, E. S., Cahyono, E. S., & Yani, I. (2024). Ensemble of naive Bayes, decision tree, and random forest to predict air quality. *IAES International Journal of Artificial Intelligence*, 13(3), 3039–3051. <https://doi.org/10.11591/ijai.v13.i3.p3039-3051>
- Syuhada, G., Akbar, A., Hardiawan, D., Pun, V., Darmawan, A., Heryati, S. H. A., Siregar, A. Y. M., Kusuma, R. R., Driejana, R., Ingole, V., Kass, D., & Mehta, S. (2023). Impacts of Air Pollution on Health and Cost of Illness in Jakarta, Indonesia. *International Journal of Environmental Research and Public Health*, 20(4). <https://doi.org/10.3390/ijerph20042916>
- Wang, Y., & Kong, T. (2019). Air Quality Predictive Modeling Based on an Improved Decision Tree in a Weather-Smart Grid. *IEEE Access*, 7, 172892–172901. <https://doi.org/10.1109/ACCESS.2019.2956599>
- Yue, H., He, C., Huang, Q., Zhang, D., Shi, P., Moallemi, E. A., Xu, F., Yang, Y., Qi, X., Ma, Q., & Bryan, B. A. (2024). Substantially reducing global PM_{2.5}-related deaths under SDG3.9 requires better air pollution control and healthcare. *Nature Communications*, 15(1). <https://doi.org/10.1038/s41467-024-46969-3>