

**KLASIFIKASI KUALITAS UDARA BERBASIS IOT DAN
ALGORITMA *DECISION TREE***

Rima Febriyanti¹, Rofilde Hasudungan², Taghfirul Azhima Yoga Siswa³

Universitas Muhammadiyah Kalimantan Timur

E-mail: febriyantirima7@gmail.com¹, rh219@umkt.ac.id², tay758@umkt.ac.id³

Abstrak

Kualitas udara merupakan salah satu faktor penting yang berdampak langsung pada kesehatan masyarakat. Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi kualitas udara berbasis Internet of Things (IoT) yang terintegrasi dengan algoritma Decision Tree. Sistem ini menggunakan sensor MQ-135, MQ-7, dan GP2Y1010AU0F untuk mengukur parameter pencemar udara seperti PM_{2.5}, PM₁₀, SO₂, dan CO. Data sensor dikirimkan secara real-time melalui mikrokontroler ESP32 dan disimpan untuk keperluan analisis klasifikasi. Dataset pelatihan diambil dari data ISPU DKI Jakarta dan digunakan untuk melatih model klasifikasi kualitas udara ke dalam empat kategori: Baik, Sedang, Tidak Sehat, dan Sangat Tidak Sehat. Model Decision Tree yang dibangun diuji menggunakan dua set data uji dari lingkungan kampus, yaitu data dari kantin dan parkir UMKT. Hasil klasifikasi menunjukkan performa yang sangat tinggi dengan akurasi mencapai 99%. Visualisasi confusion matrix dan pohon keputusan mengindikasikan bahwa model mampu mengidentifikasi kelas secara akurat, khususnya pada kelas dominan seperti “Sedang”. Namun, kelas dengan jumlah data kecil seperti “Sangat Tidak Sehat” tetap memiliki risiko salah klasifikasi akibat ketidakseimbangan data. Temuan ini menunjukkan bahwa integrasi teknologi IoT dengan algoritma Decision Tree sangat efektif dalam mengklasifikasikan kualitas udara, dan berpotensi digunakan dalam sistem monitoring lingkungan yang praktis dan akurat.

Kata Kunci — IoT, Kualitas Udara, Decision Tree, Klasifikasi, ISPU.

Abstract

Air quality is one of the key factors directly impacting public health. This study aims to develop an Internet of Things (IoT)-based air quality classification system integrated with a Decision Tree algorithm. The system utilizes MQ-135, MQ-7, and GP2Y1010AU0F sensors to measure air pollutant parameters such as PM_{2.5}, PM₁₀, SO₂, and CO. Sensor data is transmitted in real-time via an ESP32 microcontroller and stored for classification analysis. The training dataset is sourced from Jakarta's ISPU data and used to train the air quality classification model into four categories: Good, Moderate, Unhealthy, and Very Unhealthy. The Decision Tree model was tested using two sets of test data from the campus environment, namely data from the UMKT canteen and parking lot. The classification results showed very high performance with an accuracy of 99%. Visualization of the confusion matrix and decision tree indicated that the model was able to accurately identify classes, especially dominant classes such as “Moderate.” However, classes with small amounts of data, such as “Very Unhealthy,” still have a risk of misclassification due to data imbalance. These findings indicate that the integration of IoT technology with the Decision Tree algorithm is highly effective in classifying air quality and has the potential to be used in practical and accurate environmental monitoring systems.

Keywords: IoT, Air Quality, Decision Tree, Classification, ISPU.

1. PENDAHULUAN

Kualitas udara merupakan salah satu faktor yang sangat berpengaruh terhadap kesehatan fisik makhluk hidup. Lingkungan yang sehat juga mempengaruhi kualitas udara untuk memenuhi standar kesehatan. Udara mengandung oksigen yang di butuhkan untuk

mahluk hidup. Selain oksigen, terdapat juga virus, bakteri, debu, jamur dan juga yang lainnya. Oksigen di luar maupun di dalam ruangan juga dapat terkontaminasi dengan zat-zat berbahaya bagi kesehatan. Terdapat batasan batasaan tertentu pada kadar zat-zat yang dapat dinetralisir namun, jika melampaui batas normal maka dapat merusak oksigen sehingga mengganggu kesehatan. World Health Organization (WHO) menyatakan bahwa terdapat zat-zat yang membahayakan Kesehatan berasal dari bangunan, material konstruksi, peralatan, proses pembakaran atau pemanasan yang dapat memicu kesehatan (Pradiah et al. 2024).

Selain itu, aktivitas manusia juga mempengaruhi kualitas udara yang menimbulkan pencemaran udara seperti asap kendaraan, industri pabrik, pembakaran sampah dan sebagainya (Candrasari et al. 2023). Sehingga memerlukan solusi untuk meminimalisir efek yang dapat mengganggu kesehatan. Untuk melakukan pemantauan secara real-time dan mendapatkan data-data yang mendeteksi kualitas udara, dapat dilakukan dengan membangun perangkat keras yang terhubung dengan sistem pemantauan kualitas udara (Octaviano et al. 2022).

Oleh karena itu untuk mendeteksi kualitas udara dengan menggunakan Internet of Things (IoT) memungkinkan perangkat untuk mengumpulkan data secara real-time, menganalisisnya, dan mengambil tindakan otomatis berdasarkan hasil analisis. Teknologi ini membantu mengoptimalkan proses, meningkatkan produktivitas, dan mengurangi biaya. Perangkat pintar dapat dikontrol dari jarak jauh dan disesuaikan dengan preferensi pengguna, memberikan kenyamanan dan efisiensi dalam kehidupan sehari-hari. Selain itu, IoT menciptakan lingkungan yang lebih terhubung dan aman sesuai dengan kebutuhan pengguna (Madani et al. 2024).

Sang et al. (2021) telah melakukan penelitian terkait analisis data kualitas udara menggunakan metode data mining dengan klasifikasi algoritma Decision Tree dan Support Vector Machine di DKI Jakarta. Dalam penelitian tersebut, Decision Tree menghasilkan Precision sebesar 99,02%, Recall 99,73%, F1-Measure 99,37%, dan Akurasi 99,40%. Sementara itu, Support Vector Machine (SVM) menghasilkan Precision sebesar 95,82%, Recall 88,89%, F1-Measure 92,22%, dan Akurasi 94,93%.

Candrasari et al. (2023) melakukan pemilihan dampak pencemaran udara dengan menggunakan metode Des kriptif Kualitatif dengan tinjauan Literatur dengan hasil pengamatan Untuk mengurangi polusi udara, disarankan berolahraga, menanam pohon, mengurangi kendaraan bermotor, menggunakan transportasi dan produk ramah lingkungan, serta mengelola sampah dengan baik.

Octaviano et al. (2022) melakukan penelitian pemantauan kualitas udara berbasis IoT dengan Sensor yang digunakan adalah sensor MQ-7 untuk gas CO, sensor Sharp GP2Y1010AU0F untuk debu, dan sensor DHT11 untuk suhu dan kelembaban di daerah Tangerang dengan menghasilkan Selama tiga hari pengujian, sistem mencatat suhu rata-rata 29,18°C, kelembaban 80,01%, CO 3983 $\mu\text{g}/\text{m}^3$, dan debu 83,90 $\mu\text{g}/\text{m}^3$. Data kualitas udara ditampilkan real-time melalui Blynk dan tersimpan di ThingSpeak, yang dapat diakses via ThingView dan web.

Hendi Budianto1 et al. (2024) melakukan penelitian perancangan sistem monitoring kualitas udara menggunakan metode penelitian dan pengembangan yang meliputi tahap desain sistem, pembuatan dan integrasi alat, perancangan aplikasi berbasis web, serta pengujian sistem dengan menggunakan Perangkat ini terdiri dari rangkaian sensor, NodeMCU ESP8266, Arduino Uno, dan layar LCD dilakukan didalam ruangan dengan hasil sistem pemantauan berbasis website berhasil mendeteksi CO₂, CO, debu, suhu, dan kelembaban secara real-time serta menyimpan data di database.

Rumampuk et al. (2021) melakukan Perancangan Sistem Monitoring Kualitas Udara Dalam Ruangan Berbasis IoT dengan menggunakan sensor MQ135 untuk mendeteksi Kualitas udara udara, MG811 untuk mendeteksi CO₂, Sensor dust untuk mendeteksi partikel PM2.5, Sensor MQ2 mendeteksi asap, Sensor MQ9 mendeteksi CO, Sensor MQ8 meneteksi H₂, dan menggunakan ESP32 sebagai mikrokontroler didalam ruangan dengan mehasilkan Alat yang dikembangkan dapat mendeteksi polusi dan mengirim data sensor

secara real-time melalui WiFi ke platform IoT OVoRD (Online Value of Real-Time Data).

Irwansyah et al. (2023) melakukan penelitian untuk menentukan kualitas udara menggunakan metode Klasifikasi dengan algoritma Decision Tree, Naive Bayes, dan K-Nearest Neighbor di daerah DKI Jakarta dengan menghasilkan proses evaluasi menggunakan K-5 fold yang telah dilakukan menggunakan tools RapidMiner diperoleh hasil akurasi algoritma decision tree sebesar 95,89%, Akurasi algoritma naïve bayes sebesar 93,15%, dan akurasi algoritma K-NN sebesar 91,78%.

Pradih et al. (2024) melakukan penelitian kualitas udara menggunakan Klasifikasi Data Mining dengan menggunakan algoritma Naive Bayes, K-Nearest Neighbors (KNN), dan Support Vector Machine (SVM) di kota Bandung dengan menghasilkan penelitian menunjukkan bahwa algoritma SVM mencapai akurasi prediksi terbaik sebesar 92,5%.

Subagiyo et al. (2020) melakukan penelitian Rancang Bangun Sensor Node untuk Pemantauan Parameter Kualitas Udara dengan menggunakan sensor node, sensor MQ7, MQ135, MQ136, MQ131, PM10, RTC dan LCD di Pekanbaru dengan menghasilkan network capture menunjukkan rata-rata packet loss sebesar 13,6% dan latency 58 ms.

Muttaqin et al. (2024) melakukan penelitian Pemantauan Kualitas Udara Berbasis IoT menggunakan sensor MQ-135 dan sensor DHT11 yang dilakukan di Laboratorium dengan menghasilkan pengukuran menunjukkan rata-rata koefisien determinasi sebesar 0,91, menandakan sistem mampu merespons perubahan konsentrasi gas dalam kotak uji. Monitoring mencakup suhu, kelembaban, kadar alkohol, aseton, dan CO₂ di laboratorium. Pemantauan dapat dilakukan melalui aplikasi Blynk di smartphone atau Blynk Cloud di desktop PC.

Naufal H(2024) melakukan penelitian Implementasi Sistem Klasifikasi Udara Menggunakan Metode Decision Tree di Jakarta dengan menghasilkan algoritma Decision Tree dengan Criterion entropi, Max Depth 7, dan Test Size 10/90 memberikan hasil terbaik. Akurasi mencapai 91%, dengan precision 0.92, recall 0.91, dan F1-Score 0.95.

Berdasarkan latar belakang diatas, permasalahan dan perkembangan teknologi yang telah diuraikan, penelitian ini dilakukan untuk mengembangkan sistem pemantauan kualitas udara berbasis IoT dengan penerapan algoritma Decision Tree. Sistem ini diharapkan dapat memberikan solusi yang efektif dalam mendeteksi polusi udara, meningkatkan kesadaran masyarakat akan pentingnya udara bersih, serta mendukung upaya pencegahan pencemaran udara secara lebih sistematis dan berbasis data real-time.

2. METODE PENELITIAN

Pada penelitian ini dilakukan untuk memprediksi sebuah kualitas udara di lingkungan Universitas Muhammadiyah Kalimantan Timur yang diambil di beberapa titik kumpul mahasiswa melakukan aktivitas di lingkungan kampus sering kali terpapar berbagai sumber polusi, seperti polusi kendaraan, aktivitas industri di sekitar, dan kegiatan sehari-hari mahasiswa dan staf. Kualitas udara yang buruk dapat berdampak negatif pada kesehatan, terutama bagi mahasiswa yang menghabiskan banyak waktu di luar ruangan. seperti di area parkir gedung A dan Kantin Gedung G. Sistem yang dirancang akan mengumpulkan dan menganalisis data kualitas udara dari pembagi parameter seperti gas CO₂, CO, NO₂, SO₂, PM2.5, PM10, suhu, dan kelembaban udara untuk memberikan informasi dan prediksi yang akurat mengenai kondisi kualitas udara (Mariano et al., 2024).

Dengan menggunakan teknologi Internet of Things (IoT), data dari berbagai sensor akan dikumpulkan secara real-time dan dianalisis menggunakan algoritma Decision Trees untuk memprediksi kualitas udara. Hasil dari penelitian ini diharapkan dapat memberikan informasi yang berguna bagi pihak universitas dan masyarakat sekitar dalam mengambil langkah-langkah mitigasi untuk meningkatkan kualitas udara.

3. HASIL DAN PEMBAHASAN

1. Sensor Udara Berbasis IoT

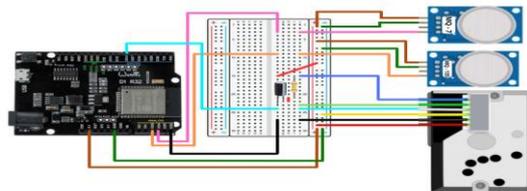
Tahap implementasi perangkat keras merupakan langkah krusial dalam pengembangan sistem pemantauan kualitas udara berbasis *Internet of Things* (IoT). Pada tahap ini, dilakukan penyusunan dan penyambungan komponen sesuai dengan desain sistem agar dapat beroperasi secara maksimal dalam mendeteksi berbagai parameter pencemar udara. Sistem ini dibangun menggunakan mikrokontroler Wemos D1 R32 (berbasis ESP-12/ESP32), yang terhubung dengan sejumlah sensor untuk memantau kondisi udara di sekitarnya.

Sensor-sensor yang digunakan meliputi MQ-7 untuk mendeteksi konsentrasi gas Karbon Monoksida (CO), MQ-135 untuk mengukur kadar gas Sulfur Dioksida (SO₂), serta GP2Y1010AU0F untuk mendeteksi partikel debu seperti PM2.5 dan PM10. Data dari masing-masing sensor langsung diproses oleh mikrokontroler dan dikirim secara real-time ke platform Blynk sebagai antarmuka monitoring, serta dicatat secara otomatis di Google Sheets untuk keperluan dokumentasi dan analisis data.

Dalam penerapannya, tiap sensor dihubungkan ke pin analog ADC pada Wemos D1 R32 berdasarkan konfigurasi yang telah ditentukan sebelumnya. Khusus untuk sensor debu GP2Y1010AU0F, dibutuhkan pin tambahan untuk mengendalikan LED internalnya. Seluruh rangkaian disusun di atas *breadboard*, dan skema hubungan antar komponen divisualisasikan dalam diagram pada Gambar 2.7, sementara rincian konfigurasi koneksi pin antar perangkat ditampilkan dalam Tabel 1.

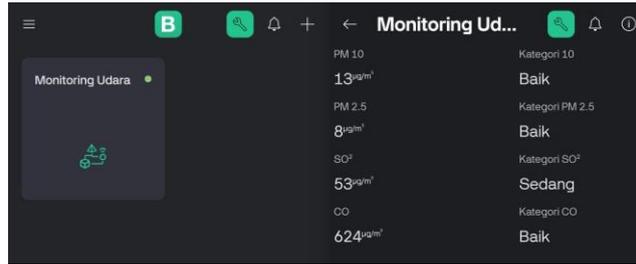
Tabel 1. Konfigurasi Pin Antar Perangkat

No.	Komponen	Pin Sensor	Pin Wemos D1 R23
1.	MQ-7	AOUT	GPIO34
		VCC	V5
		GND	GND
2.	MQ-135	AOUT	GPIO35
		VCC	V5
		GND	GND
3.	GP2Y1010AU0F	AOUT	GPIO39
		VCC	V5
		GND	GND
		V-LED	GPIO14
		LED-GND	GND



Gambar 1. Diagram Rangkaian Alat

Tampilan antarmuka pengguna pada platform Blynk dirancang dengan konsep yang sederhana namun tetap menyajikan informasi yang lengkap. Antarmuka ini menampilkan data parameter kualitas udara secara real-time. Gambar 1. menunjukkan tampilan monitoring kualitas udara melalui aplikasi Blynk, di mana pengguna dapat memantau nilai konsentrasi masing-masing parameter (PM10, PM2.5, SO₂, dan CO) lengkap dengan klasifikasi tingkat kualitas udaranya.



Gambar 2. Tampilan Antarmuka Pada Platform Blynk

1) Kandungan Udara Menggunakan *Internet of Things* (IoT)

Setelah sistem pemantauan kualitas udara berhasil diterapkan, proses pengambilan data dilakukan secara langsung di dua titik lokasi dalam kawasan Universitas Muhammadiyah Kalimantan Timur, yaitu di area parkir Gedung A dan kantin Gedung G. Pengumpulan data berlangsung selama empat hari dengan pembagian lokasi dan waktu: Senin dan Selasa di parkir Gedung A, serta Rabu dan Kamis di kantin Gedung G. Waktu pengamatan dimulai pukul 10.00 WITA hingga pukul 03.00 WITA keesokan harinya. Rentang waktu ini dipilih untuk menyesuaikan dengan jam operasional kegiatan kampus, sehingga data yang diperoleh mencerminkan kondisi kualitas udara saat aktivitas perkuliahan berlangsung secara normal.

Di lokasi parkir Gedung A, alat pemantau diletakkan di area terbuka tanpa hambatan bangunan, memungkinkan sensor mendeteksi udara sekitar secara optimal. Dilihat pada Tabel 2. lokasi ini memiliki sirkulasi udara yang cukup baik dan intensitas lalu lintas kendaraan yang tergolong sedang. Berdasarkan hasil pemantauan dan data awal yang dikumpulkan, tingkat konsentrasi polutan di area ini cenderung rendah dan stabil selama periode pengambilan data, meskipun terjadi sedikit peningkatan pada jam-jam sibuk seperti saat siang hari.

Tabel 2. Data Kandungan Gas Pada Area Parkiran Gedung A

NO.	PM10	PM25	SO2	CO
1.	2	1	30	980
2.	2	1	33	986
3.	2	1	36	990
4.	2	1	39	993
5.	2	1	39	995
...
1006	8	5	47	507
1007	8	5	55	449
1008	8	5	56	449
1009	7	5	56	452
1010	8	5	49	470

Di lokasi kantin Gedung G, alat monitoring ditempatkan di area semi-terbuka yang memiliki kondisi lingkungan berbeda dari parkir. Aktivitas di sekitar kantin lebih ramai, terutama saat jam makan siang dan sore. Kegiatan seperti memasak dan merokok di sekitar area ini diduga menjadi penyebab utama tingginya kadar beberapa polutan, terutama gas sulfur dioksida (SO₂) dan partikel debu seperti PM2.5 dan PM10 dapat dilihat pada Tabel 2. Secara keseluruhan, hasil pengukuran sensor di kantin menunjukkan angka konsentrasi polutan yang lebih tinggi dibandingkan dengan yang ada di parkir.

Tabel 3. Data Kandungan Gas Pada Area Kantin Gedung G

NO.	PM10	PM25	SO2	CO
1	19	13	85	664
2	19	13	81	661
3	19	13	78	652
4	19	13	75	621
5	20	13	73	628

...
1006	7	5	30	760
1007	5	4	30	832
1008	5	3	30	882
1009	5	4	31	918
1010	5	3	32	942

Sistem ini menggunakan tiga jenis sensor, yaitu MQ-7 untuk mendeteksi gas karbon monoksida (CO), MQ-135 untuk mengukur gas sulfur dioksida (SO₂), dan GP2Y1010AU0F untuk mengukur partikel debu PM2.5 dan PM10. Semua sensor tersebut terhubung ke mikrokontroler ESP32, yang bertugas mengubah sinyal analog dari sensor menjadi angka konsentrasi polutan dalam satuan mikrogram per meter kubik ($\mu\text{g}/\text{m}^3$).

Proses pengolahan data meliputi pengukuran tegangan dari sensor, perhitungan rasio R_s/R_o sesuai karakteristik sensor, konversi dari satuan ppm ke $\mu\text{g}/\text{m}^3$, dan penyaringan data menggunakan metode low-pass filter agar hasilnya lebih stabil dan tidak mudah berubah-ubah.

Setiap data yang dibaca oleh sensor akan dikirim secara langsung (*real-time*) ke aplikasi Blynk sehingga bisa dipantau lewat ponsel. Pada saat yang sama, data tersebut juga dikirim ke Google Sheets melalui metode HTTP POST yang diarahkan ke *Google Apps Script*. Data yang dikirim mencakup nilai konsentrasi masing-masing polutan (CO, SO₂, PM10, dan PM2.5), informasi tentang polutan dengan kadar tertinggi, parameter pencemar paling dominan (*critical parameter*), serta kategori kualitas udara berdasarkan standar ISPU. Selama proses pengambilan data, setiap lokasi mencatat lebih dari 1.000 baris data.

Mikrokontroler

Mikrokontroler adalah komponen utama dalam sistem pemantauan kualitas udara berbasis IoT yang digunakan dalam penelitian ini. Mikrokontroler berfungsi sebagai pusat kendali untuk mengolah data yang diterima dari sensor-sensor, lalu mengirimkan hasilnya secara *real-time* ke platform Blynk. Pada sistem ini, digunakan Wemos D1 R32 yang berbasis ESP32. Pemilihan ESP32 karena mikrokontroler ini memiliki keunggulan, seperti konektivitas Wi-Fi dan Bluetooth bawaan, prosesor dual-core yang cepat, serta jumlah pin input-output yang cukup banyak sehingga cocok untuk menghubungkan berbagai sensor sekaligus.

Dalam sistem ini, berbagai sensor dihubungkan ke mikrokontroler. Sensor-sensor tersebut di antaranya adalah MQ-7 untuk mendeteksi gas karbon monoksida (CO), MQ-135 untuk mengukur gas sulfur dioksida (SO₂), serta GP2Y1010AU0F yang digunakan untuk mendeteksi partikel debu PM2.5 dan PM10. Selain itu, terdapat juga sensor DHT22 yang digunakan untuk mengukur suhu dan kelembaban udara. Semua data dari sensor akan masuk ke mikrokontroler melalui pin analog, kemudian diolah untuk menghasilkan nilai konsentrasi gas atau partikel dalam satuan yang sesuai.

Setelah data dari sensor diproses, mikrokontroler akan mengirimkan hasilnya ke aplikasi Blynk melalui jaringan Wi-Fi. Aplikasi ini berfungsi sebagai tampilan monitoring yang bisa diakses pengguna lewat smartphone. Dengan begitu, pengguna bisa melihat kondisi kualitas udara secara langsung dan real-time. Selain itu, data juga dicatat secara otomatis di Google Sheets sebagai dokumentasi untuk analisis lebih lanjut. Mikrokontroler ini juga mengendalikan perangkat output seperti lampu LED dan buzzer. Lampu LED digunakan sebagai indikator kualitas udara, misalnya jika kualitas udara baik maka lampu akan berwarna hijau, jika sedang maka berwarna kuning, dan jika kualitas udara tidak sehat maka lampu akan berwarna merah. Sedangkan buzzer berfungsi sebagai alarm peringatan jika kualitas udara mencapai tingkat berbahaya.

Secara keseluruhan, mikrokontroler ESP32 ini memegang peran penting dalam sistem karena menjadi pusat dari seluruh proses mulai dari pengambilan data, pengolahan, hingga pengiriman informasi ke pengguna. Dengan kemampuan ini, ESP32 sangat cocok digunakan dalam proyek pemantauan kualitas udara berbasis IoT seperti pada penelitian ini.

Implementasi Algoritma *Decision Tree*

Untuk mengklasifikasikan kualitas udara, diterapkan algoritma *Decision Tree*. Data dari kedua lokasi diproses, dilatih, dan kemudian di uji untuk mengevaluasi kinerja model.

1) Preprocessing Data

Sebelum data digunakan untuk melatih dan menguji model klasifikasi, data harus dipersiapkan terlebih dahulu melalui proses pra-pemrosesan. Tahapan ini penting dilakukan agar data yang digunakan benar-benar bersih, rapi, dan siap untuk dianalisis. Dalam penelitian ini, data dibagi menjadi dua bagian utama, yaitu data latih (*training data*) dan data uji (*testing data*).

Data ISPU (Indeks Standar Pencemar Udara) adalah indikator yang digunakan untuk menggambarkan tingkat pencemaran udara berdasarkan konsentrasi polutan tertentu. Di Indonesia, ISPU dihitung berdasarkan tujuh parameter pencemar utama, yaitu: *PM10*, *PM2.5*, *SO₂* (Sulfur Dioksida), *CO* (Karbon Monoksida), *O₃* (Ozon), *NO₂* (Nitrogen Dioksida), dan *HC* (Hidrokarbon). Masing-masing parameter memiliki ambang batas dan rumus konversi tersendiri yang kemudian disesuaikan dalam satuan nilai ISPU, sehingga dapat diklasifikasikan ke dalam kategori kualitas udara seperti: Baik, Sedang, Tidak Sehat, Sangat Tidak Sehat, dan Berbahaya.

Data ISPU DKI Jakarta yang digunakan dalam penelitian ini sebenarnya memiliki atribut lengkap dari ketujuh parameter pencemar tersebut. Namun, untuk menyesuaikan dengan keterbatasan perangkat IoT yang digunakan dalam pengukuran lapangan, hanya empat parameter utama yang dipilih, yaitu *PM10*, *PM2.5*, *SO₂*, dan *CO* karena data untuk keempat parameter tersebut tersedia secara konsisten dan memiliki relevansi tinggi terhadap evaluasi kualitas udara.

Model klasifikasi dalam penelitian ini menggunakan pendekatan *supervised learning* dengan memanfaatkan data latih yang telah memiliki label kategori ISPU dari sumber resmi. Dengan demikian, meskipun hanya empat parameter yang tersedia dalam data uji dari sensor, model tetap dapat melakukan klasifikasi berdasarkan pola dan relasi yang telah dipelajari dari data latih yang lengkap. Model tidak menghitung ulang nilai indeks ISPU secara numerik, melainkan memetakan kombinasi nilai *PM10*, *PM2.5*, *SO₂*, dan *CO* ke dalam kategori kualitas udara. Pada kode tersebut, data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Bagian *test_size=0.2* artinya 20% data digunakan untuk pengujian. Sedangkan *random_state=42* digunakan agar pembagian data selalu sama saat kode dijalankan berulang kali, sehingga hasil pengujian bisa konsisten dan mudah diulang. Dengan proses ini, model bisa dievaluasi secara adil dan hasilnya dapat dipertanggungjawabkan karena menggunakan data yang telah diproses dengan baik.

2) Preprocessing Data ISPU

Data Indeks Standar Pencemar Udara (ISPU) digunakan sebagai acuan utama untuk menilai tingkat kualitas udara berdasarkan kadar beberapa zat pencemar. Pada penelitian ini, data ISPU yang digunakan berasal dari wilayah DKI Jakarta dengan rentang waktu dan lokasi pengamatan yang bervariasi. Dataset yang digunakan memiliki beberapa kolom, seperti *periode_data*, *bulan*, *tanggal*, *stasiun*, *PM10*, *PM2.5*, *SO₂*, *CO*, *O₃*, *NO₂*, *max_value*, *parameter_kritis*, dan *kategori*.

Namun, dalam penelitian ini hanya fokus pada empat parameter utama, yaitu *PM10*, *PM2.5*, *SO₂*, dan *CO*. Pemilihan parameter tersebut dilakukan karena data untuk empat

parameter ini paling lengkap dan paling relevan untuk menilai kualitas udara di lokasi penelitian. Data pelatihan (*training data*) diambil dari dataset ISPU Jakarta tersebut. Awalnya, dataset terdiri dari 2.281 baris data, namun setelah dilakukan proses pembersihan data untuk menghapus data yang kosong atau tidak valid, hanya tersisa 2.107 baris data yang dinyatakan bersih dan layak untuk digunakan dalam pelatihan model klasifikasi.

Tabel 4. Data ISPU DKI Jakarta Setelah di Cleaning Data

No.	PM10	PM25	SO2	CO	Kategori
1	51	65	45	9	Sedang
2	27	34	45	5	Sedang
3	46	65	46	8	Sedang
4	37	55	47	7	Sedang
5	43	62	50	7	Sedang
7	40	52	48	8	Sedang
...
129	55	103	35	22	Tidak Sehat
7					
129	51	107	37	22	Tidak Sehat
8					
129	55	107	34	21	Tidak Sehat
9					
...
210	48	58	31	14	Sedang
6					
210	51	62	30	16	Sedang
7					
210	26	37	28	14	Baik
8					

Data ISPU digunakan sebagai data referensi atau acuan untuk menentukan label kelas kualitas udara. Label ini sangat penting dalam proses pelatihan model, karena setiap data yang digunakan harus memiliki kategori tertentu, misalnya Baik, Sedang, atau Tidak Sehat, yang ditentukan berdasarkan konsentrasi polutan di udara.

Dengan adanya label ini, model dapat belajar mengenali pola dari data sebelumnya (data historis). Tujuannya agar ketika ada data baru dari sensor di sistem IoT, model bisa memprediksi secara otomatis masuk kategori kualitas udara yang mana. Proses ini memungkinkan sistem untuk memberikan hasil klasifikasi yang akurat sesuai dengan kondisi udara saat itu.

3) *Preprocessing* Data Sensor

Setelah data ISPU digunakan sebagai label dalam proses klasifikasi, langkah selanjutnya adalah mengolah data sensor yang akan menjadi fitur atau input bagi sistem klasifikasi. Data sensor ini diambil secara langsung dari perangkat mikrokontroler Wemos D1 R32 yang telah dipasangkan dengan beberapa sensor kualitas udara. Sensor-sensor tersebut meliputi sensor partikulat PM2.5, sensor gas karbon monoksida (CO), serta sensor untuk mengukur suhu dan kelembaban udara. Data dari sensor dikirimkan secara *real-time* ke aplikasi Blynk sehingga pengguna dapat memantau kondisi udara secara langsung. Selain itu, data juga otomatis dicatat ke dalam *Google Sheets* sebagai data mentah untuk proses analisis lanjutan.

Agar data sensor ini bisa digunakan dalam sistem klasifikasi, terlebih dahulu dilakukan proses pembersihan data dan penyesuaian format agar sesuai dengan kebutuhan analisis. Karena data sensor masih berupa satuan konsentrasi dalam mikrogram per meter kubik ($\mu\text{g}/\text{m}^3$), maka perlu dilakukan konversi ke dalam skala ISPU. Hal ini bertujuan agar

data dari sensor bisa disejajarkan dengan label ISPU yang sudah ada sebelumnya. Proses konversi ini mengacu pada ambang batas ISPU yang telah ditetapkan oleh Kementerian Lingkungan Hidup dan Kehutanan (KLHK). Tabel 5. dalam penelitian ini menyajikan batas-batas ISPU untuk masing-masing parameter pencemar udara yang digunakan sebagai acuan dalam proses konversi tersebut.

Tabel 5. Batas ISPU Berdasarkan KLHK

ISPU	PM 10 ($\mu\text{g}/\text{m}^3$)	PM 2.5 ($\mu\text{g}/\text{m}^3$)	SO ₂ ($\mu\text{g}/\text{m}^3$)	CO ($\mu\text{g}/\text{m}^3$)	O ₃ ($\mu\text{g}/\text{m}^3$)	NO ₂ ($\mu\text{g}/\text{m}^3$)	HC ($\mu\text{g}/\text{m}^3$)
0-50	50	15,5	52	4000	120	80	45
51-100	150	55,4	180	8000	235	200	100
101-200	350	150,4	400	15000	400	1130	215
201-300	420	250,4	800	30000	800	2260	432
≥ 300	500	500	1200	45000	1000	3000	648

Untuk melakukan konversi dari Sensor ke ISPU, digunakan rumus interpolasi linier sebagai berikut :

$$ISPU = \frac{I_{high}}{C_{high}} + \frac{I_{low}}{C_{low}} \times (C - C_{low}) + I_{low} \quad (\text{Rumus 5})$$

Keterangan :

1. C = Nilai konsentrasi aktual ($\mu\text{g}/\text{m}^3$)
2. C_{low}, C_{high} = Batas bawah dan atas dari rentang konsentrasi
3. I_{low}, I_{high} = Batas bawah dan atas dari rentang ISPU

Contoh perhitungan

Jika nilai PM2.5 yang terdeteksi adalah $5 \mu\text{g}/\text{m}^3$, maka nilai ini termasuk dalam kategori **Baik** dengan rentang :

- $C_{low} = 0, C_{high} = 15$
- $I_{low} = 0, I_{high} = 50$

Maka:

$$ISPU = \left(\frac{50}{15} + \frac{0}{0} \right) + (5 - 0) = \frac{50}{15} \times 5 = 3.3 \times 5 = 16.67 \quad (\text{Rumus 6})$$

Jadi, nilai ISPU untuk PM2.5 sebesar $5 \mu\text{g}/\text{m}^3$ adalah sekitar 16.67, yang termasuk dalam kategori **Baik**.

Setelah proses konversi dilakukan, data dari masing-masing sensor yang sebelumnya masih dalam satuan asli, yaitu mikrogram per meter kubik ($\mu\text{g}/\text{m}^3$) untuk PM2.5, PM10, dan SO₂, serta dalam satuan part per million (ppm) untuk gas CO, telah diubah ke dalam skala Indeks Standar Pencemar Udara (ISPU). Dengan konversi ini, data dari sensor sudah sesuai dengan standar kualitas udara nasional yang ditetapkan oleh Kementerian Lingkungan Hidup dan Kehutanan (KLHK). Data hasil konversi inilah yang kemudian digunakan sebagai fitur input dalam proses pelatihan dan pengujian model klasifikasi.

Pengambilan data dilakukan di dua lokasi yang berbeda, yaitu kantin dan parkir. Karena kedua lokasi ini memiliki kondisi lingkungan yang berbeda, data yang dihasilkan pun bervariasi sesuai dengan tingkat pencemarannya masing-masing. Dengan mengubah data sensor ke dalam satuan ISPU, perbandingan antara kedua lokasi menjadi lebih adil dan terstandarisasi. Hal ini memungkinkan sistem klasifikasi dapat menganalisis dan membandingkan data dari kedua lokasi dengan cara yang konsisten dan objektif.

Selanjutnya data yang di dapatkan oleh sensor kemudian di konversi dengan standar ISPU dapat dilihat pada Tabel 6 dan Tabel 7 pada tabel tersebut sudah memenuhi format

yang sesuai, yaitu berupa Indeks Standar Pencemar Udara (ISPU) yang terukur dan terstruktur dengan baik. Dengan format seperti ini, data menjadi siap untuk digunakan dalam penerapan sistem klasifikasi berbasis machine learning. Proses konversi ini tidak hanya bertujuan untuk menyamakan satuan dari berbagai parameter pencemar, tetapi juga memastikan bahwa data tersebut benar-benar valid, konsisten, dan siap dipakai sebagai input pada proses pelatihan dan pengujian model klasifikasi.

Tabel 6. Data Parkiran Sesuai Standar ISPU

No.	PM10	PM25	SO2	CO
1	2	1	30	980
2	2	1	33	986
3	2	1	36	990
4	2	1	39	993
5	2	1	39	995
...
1006	8	5	47	507
1007	8	5	55	449
1008	8	5	56	449
1009	7	5	56	452
1010	7	5	56	452

Tabel 7. Data Kantin Sesuai Standar ISPU

No.	PM10	PM25	SO2	CO
1	19	13	85	664
2	19	13	81	661
3	19	13	78	652
4	19	13	75	621
5	20	13	73	628
...
1006	7	5	30	760
1007	5	4	30	832
1008	5	3	30	882
1009	5	4	31	918
1010	5	3	32	942

Berdasarkan data yang disajikan pada Tabel 6. dan Tabel 7, penelitian ini menggunakan dataset Indeks Standar Pencemar Udara (ISPU) wilayah parkir dan kantin yang pada awalnya memiliki tujuh atribut polutan utama, yakni *PM10*, *PM2.5*, *SO₂*, *CO*, *O₃*, *NO₂*, dan *HC*. Namun, dalam implementasinya, hanya empat atribut yang diambil, yaitu *PM10*, *PM2.5*, *SO₂*, dan *CO*. Pemilihan empat atribut tersebut menyesuaikan dengan jenis sensor yang terpasang pada sistem monitoring yang dirancang, antara lain sensor *MQ-7* untuk mendeteksi karbon monoksida (*CO*), sensor *MQ-135* untuk mendeteksi sulfur dioksida (*SO₂*), serta sensor *GP2Y1010AU0F* yang dapat mengukur konsentrasi partikel debu *PM10* dan *PM2.5*. Sedangkan sensor untuk mendeteksi polutan lain seperti ozon (*O₃*), nitrogen dioksida (*NO₂*), dan hidrokarbon (*HC*) belum tersedia pada perangkat yang dikembangkan.

Keputusan untuk menyederhanakan jumlah atribut dari tujuh menjadi empat tersebut belum melalui validasi oleh pakar (*expert*). Namun, langkah ini diambil agar data yang digunakan tetap relevan dengan kemampuan perangkat keras (sensor) yang tersedia, sehingga hasil pemantauan yang diperoleh tetap dapat mewakili kondisi kualitas udara di lingkungan pemantauan. Selain itu, dalam metode perhitungan ISPU, penentuan status kualitas udara mengikuti prinsip bahwa nilai ISPU tertinggi dari parameter yang diukur menjadi penentu kategori kualitas udara secara keseluruhan. Artinya, meskipun sistem

hanya memantau empat parameter, status kualitas udara akhir tetap diambil berdasarkan nilai ISPU tertinggi di antara parameter tersebut. Dengan demikian, hasil klasifikasi yang diperoleh tetap dapat mencerminkan kondisi pencemaran udara paling dominan pada periode pengukuran

4) Penerapan Algoritma *Decision Tree*

Pada tahap ini, algoritma *Decision Tree* C4.5 digunakan sebagai metode klasifikasi untuk memprediksi kategori kualitas udara berdasarkan data sensor yang telah dikumpulkan dan diproses sebelumnya. Salah satu keunggulan dari *Decision Tree* C4.5 adalah kemampuannya dalam menangani data kontinu dan data dengan banyak kategori. Berbeda dengan beberapa algoritma lain, *Decision Tree* tidak memerlukan proses normalisasi atau penyamaan skala data, karena metode ini bekerja dengan membagi data berdasarkan nilai ambang yang ditentukan secara otomatis saat proses pelatihan.

Dalam algoritma C4.5, proses pemilihan fitur terbaik dilakukan dengan menggunakan konsep *Entropy*, yaitu ukuran ketidakpastian dalam suatu dataset. Semakin tinggi *Entropy*, semakin tinggi ketidakpastian dalam data tersebut. Sistem akan memilih fitur yang dapat mengurangi *Entropy* paling besar, melalui proses yang disebut *Information Gain*. Namun, karena *Information Gain* terkadang bias terhadap fitur dengan banyak nilai, C4.5 menyempurnakan metode ini dengan menggunakan *Gain Ratio*. *Gain Ratio* adalah perbandingan antara *Information Gain* dan *Split Information*, yang membuat pemilihan fitur menjadi lebih adil dan seimbang.

Langkah-langkah pembentukan *Decision Tree* C4.5 dimulai dengan membagi data menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*). Data latih digunakan untuk membangun pohon keputusan, sedangkan data uji digunakan untuk mengevaluasi kinerja model. Pohon keputusan dibentuk dengan memilih fitur yang memiliki nilai *Gain Ratio* tertinggi sebagai node utama. Proses ini dilakukan secara berulang hingga semua data dapat diklasifikasikan dengan baik atau hingga memenuhi kriteria penghentian. Pada tahap akhir, model yang telah dibangun digunakan untuk memprediksi data uji. Hasil prediksi dibandingkan dengan label sebenarnya untuk mengukur tingkat akurasi dan performa klasifikasi model.

Sebagai contoh, Tabel 8 menampilkan lima data sampel yang masing-masing berisi empat parameter pencemar udara, yaitu PM10, PM2.5, SO₂, dan CO, serta satu kolom label yang menunjukkan kategori kualitas udara sebagai target klasifikasi. Dataset kecil ini digunakan untuk mendemonstrasikan proses pembentukan *Decision Tree* C4.5 secara manual, termasuk perhitungan *Entropy*, *Information Gain*, dan *Gain Ratio* sebagai dasar pemilihan fitur utama dalam pohon keputusan. Distribusi kelas pada dataset contoh terdiri dari dua data dengan kategori "BAIK", satu data "SEDANG", dan dua data "TIDAK SEHAT", sehingga mencerminkan variasi kategori yang cukup beragam. Contoh ini bertujuan untuk mempermudah pemahaman tentang cara kerja algoritma *Decision Tree* C4.5 sebelum diterapkan pada dataset yang lebih besar.

Tabel 8. Data Sampel

PM10	PM2.5	SO2	CO
27.0	34.0	45	5
33.0	44.0	47	14
37.0	47.0	7	26
67.0	102.0	13	34
78.0	111.0	58	15

Karena data ini belum berlabel, kita harus memberikan label dulu. Misalnya pakai metode ISPU maksimum (yang terbesar di antara PM10, PM2.5, SO₂, CO) untuk menentukan kualitas udara:

Tabel 9. Klasifikasi Menggunakan ISPU

ISPU	Kategori
≤ 50	BAIK
51 - 100	SEDANG
> 100	TIDAK SEHAT

Sehingga dapat kita klasifikasikan secara manual menjadi :

PM10	PM2.5	SO2	CO	Kelas
27.0	34.0	45	5	BAIK
33.0	44.0	47	14	BAIK
37.0	47.0	7	26	SEDANG
67.0	102.0	13	34	TIDAK SEHAT
78.0	111.0	58	15	TIDAK SEHAT

Pada proses klasifikasi menggunakan algoritma Decision Tree C4.5, ada beberapa langkah utama yang dilakukan untuk membentuk pohon keputusan. Langkah-langkah ini melibatkan perhitungan Entropy, Information Gain, dan Gain Ratio. Berikut penjelasannya secara bertahap:

1) Hitung *Entropy* Dataset

Rumus

$$\text{Entropy (S)} = - \sum_{i=1}^k p_i \log_2 p_i \quad (\text{Rumus 7})$$

Kelas :

BAIK: 2 Data $\rightarrow p = \frac{2}{5}$

SEDANG: 1 Data $\rightarrow p = \frac{1}{5}$

TIDAK SEHAT : 2 Data $\rightarrow p = \frac{2}{5}$

Perhitungan:

$$\begin{aligned} \text{Entropy}(S) &= - \binom{2}{5} \log_2 \binom{2}{5} - \binom{1}{5} \log_2 \binom{1}{5} - \binom{2}{5} \log_2 \binom{2}{5} \\ &= -0.4 \times (-1.3219) - 0.2 \times (-2.3219) - 0.4 \times (1.3219) \\ &= 0.5288 + 0.4644 + 0.5288 = 1.5219 \end{aligned}$$

Entropy dataset = 1.5219

2) Tentukan Fitur Terbaik

C4.5 memilih fitur terbaik berdasarkan *Gain Ratio*. Kita akan mencoba fitur PM2.5 karena ini yang paling signifikan.

Split PM2.5 (Threshold 50)

PM2.5 ≤ 50 → Data 1, 2, 3 → Kelas: BAIK, BAIK, SEDANG

PM2.5 > 50 → Data 4, 5 → Kelas: TIDAK SEHAT, TIDAK SEHAT

3) Hitung Entropy setelah split

Untuk $PM2.5 \leq 50$:

Kategori	Jumlah
BAIK	2
SEDANG	1
TIDAK SEHAT	0

$$\begin{aligned} \text{Entropy}(S1) &= - \binom{2}{3} \log_2 \binom{2}{3} - \binom{1}{3} \log_2 \binom{1}{3} \\ &= -0.6667 \times (-0.5849) - 0.3333 \times (-1.5849) = 0.3899 + 0.5283 = 0.9182 \end{aligned}$$

Untuk $PM2.5 > 50$:

Semua data adalah TIDAK SEHAT Maka Entropy = 0

4) Hitung *Information Gain*

$$\begin{aligned} Gain(S, PM2.5) &= Entropy(S) - \left(\frac{3}{5} \times 0.9182 + \frac{2}{5} \times 0 \right) \\ &= 1.5219 - (0.6 \times 0.9182) = 1.5219 - 0.5509 = 0.971 \end{aligned}$$

5) Hitung *Split Information*

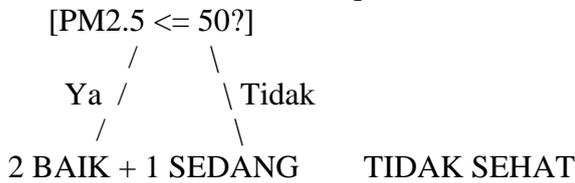
$$\begin{aligned} SplitInfo(PM2.5) &= -\binom{3}{5} \log_2 \binom{3}{5} - \binom{2}{5} \log_2 \binom{2}{5} \\ &= -0.6 \times (-0.7369) - 0.4 \times (-1.3219) = 0.4421 + 0.5288 = 0.9709 \end{aligned}$$

6) Hitung *Gain Ratio*

$$Gain Ratio = \frac{Gain}{Split Info} = \frac{0.971}{0.9709} = 1.0001$$

Karena *Gain Ratio* mendekati 1, maka fitur *PM2.5* dipilih sebagai Node Utama

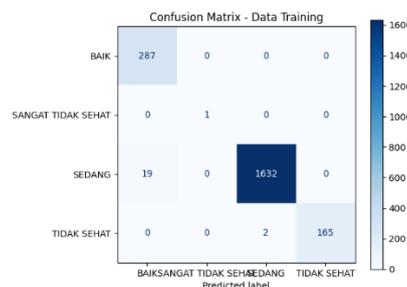
7) Bentuk Pohon Keputusan



Pengujian

Pengujian sistem dalam penelitian ini dilakukan dengan menggunakan metode *Confusion Matrix* sebagai alat evaluasi utama. Penggunaan *Confusion Matrix* dipilih karena metode ini mampu memberikan gambaran yang lebih detail mengenai performa model dalam proses klasifikasi. *Confusion Matrix* tidak hanya menunjukkan berapa banyak data yang berhasil diklasifikasikan dengan benar, tetapi juga memperlihatkan jumlah data yang salah klasifikasi pada masing-masing kategori. Dengan menggunakan matrix ini, evaluasi dapat dilakukan secara lebih menyeluruh karena setiap hasil prediksi akan dibandingkan langsung dengan kelas sebenarnya, baik itu kategori “Baik”, “Sedang”, maupun “Tidak Sehat”.

Alasan penggunaan *Confusion Matrix* dalam pengujian adalah karena klasifikasi kualitas udara merupakan masalah multi-kelas, sehingga diperlukan metode evaluasi yang dapat memberikan informasi secara spesifik untuk masing-masing kelas. Selain itu, *Confusion Matrix* juga membantu dalam menghitung metrik lain seperti *akurasi*, *precision*, *recall*, dan *F1-score*, yang dapat memberikan gambaran lebih lengkap tentang performa model. Melalui pengujian ini, peneliti dapat mengetahui seberapa baik sistem yang dibangun mampu mengklasifikasikan data dengan benar dan apakah terdapat kecenderungan model untuk salah dalam mengklasifikasikan data pada kelas tertentu. Dengan informasi ini, proses evaluasi menjadi lebih transparan dan dapat dijadikan dasar untuk perbaikan model di masa mendatang.



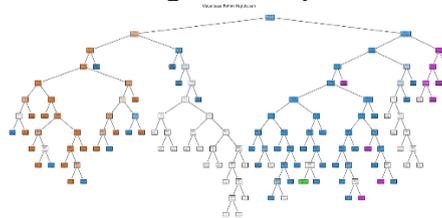
Gambar 3. *Confusion Matrix*

Hasil Gambar 3. Berdasarkan visualisasi confusion matrix data training di atas, kita dapat menganalisis kinerja model klasifikasi Decision Tree dalam mengenali kualitas udara berdasarkan kategori ISPU: BAIK, SEDANG, TIDAK SEHAT, dan SANGAT TIDAK SEHAT.

Model menunjukkan kinerja klasifikasi yang sangat baik, terutama pada kategori SEDANG, di mana terdapat 1.632 data yang berhasil diprediksi dengan benar dari kelas tersebut. Ini merupakan jumlah terbanyak dari keseluruhan prediksi yang dilakukan. Kategori BAIK juga memiliki akurasi tinggi dengan 287 data diklasifikasikan dengan benar, tanpa adanya kesalahan prediksi ke kelas lain. Untuk kategori TIDAK SEHAT, model mampu mengklasifikasikan dengan tepat sebanyak 165 data, walaupun terdapat 2 data dari kelas ini yang salah diklasifikasikan sebagai SEDANG.

Namun, terdapat beberapa kesalahan klasifikasi yang tercatat. Sebanyak 19 data dari kelas SEDANG justru diprediksi sebagai BAIK, serta 1 data dari kelas SANGAT TIDAK SEHAT diprediksi sebagai SEDANG, yang menandakan bahwa model masih memiliki kelemahan dalam membedakan kelas-kelas dengan distribusi data yang rendah atau fitur yang tumpang tindih.

Secara umum, model Decision Tree mampu mengklasifikasikan mayoritas data dengan benar, terutama pada kelas dengan jumlah data dominan. Kesalahan klasifikasi paling sering terjadi antara kelas SEDANG dan BAIK, serta TIDAK SEHAT dan SEDANG, yang menunjukkan bahwa fitur yang digunakan mungkin memiliki kemiripan nilai antar kelas-kelas tersebut. Evaluasi lebih lanjut seperti akurasi, precision, recall, dan F1-score dibutuhkan untuk memberikan gambaran performa model secara menyeluruh.



Gambar 4. Pohon Keputusan

Gambar 4. adalah visualisasi pohon keputusan (*Decision Tree*) menunjukkan proses klasifikasi kualitas udara berdasarkan data latih *IspuTrain* menggunakan algoritma *Decision Tree*. Pohon ini terdiri dari beberapa cabang yang berasal dari satu node utama (*root node*) dan mengarah ke berbagai simpul (*nodes*) hingga mencapai simpul akhir (*leaf nodes*) yang merepresentasikan hasil klasifikasi. Pohon ini dibangun dengan cara membagi data berdasarkan nilai ambang (*threshold*) dari fitur-fitur tertentu, seperti *pm10*, *co*, *so2*, dan kemungkinan *o3* atau *pm2_5*. Fitur yang ditempatkan di bagian paling atas, yaitu *root node*, merupakan fitur yang dianggap paling berpengaruh dalam proses pemisahan kelas berdasarkan tingkat impurity atau heterogenitas data, yang diukur menggunakan nilai gini impurity.

Setiap simpul pada pohon ini memuat informasi penting, yaitu: fitur yang dijadikan dasar pemisahan, nilai *threshold*-nya, nilai gini impurity (yang menunjukkan kemurnian data dalam node tersebut), jumlah sampel yang masuk ke dalam simpul, distribusi kelas dalam bentuk *array (value)*, dan label kelas mayoritas yang dijadikan hasil klasifikasi pada simpul tersebut (*class*). Sebagai contoh, jika suatu simpul memiliki informasi $pm10 \leq 30.5$, $gini = 0.2$, $samples = 120$, $value = [110, 5, 5, 0]$, dan $class = BAIK$, maka simpul tersebut memuat 120 data, 110 di antaranya tergolong sebagai kelas BAIK, dan simpul tersebut diputuskan sebagai klasifikasi BAIK.

Pohon keputusan ini memiliki struktur bercabang yang kompleks, mencerminkan bahwa model dapat mengidentifikasi pola yang cukup spesifik dalam data ISPU. Masing-

masing simpul bercabang lagi berdasarkan fitur dan nilai threshold yang berbeda, sampai akhirnya mencapai simpul akhir (daun), yang merepresentasikan hasil akhir klasifikasi ke dalam salah satu dari empat kelas: BAIK, SEDANG, TIDAK SEHAT, dan SANGAT TIDAK SEHAT. Warna pada setiap node juga mencerminkan kelas hasil klasifikasi semakin kuat warna yang tampak, maka semakin dominan kelas tersebut pada simpul tersebut. Misalnya, simpul dengan warna biru tua biasanya mengindikasikan dominasi kelas SEDANG, sementara oranye atau kuning dapat mewakili kelas BAIK, tergantung skema warna yang digunakan.

Model *Decision Tree* ini mampu menjelaskan secara transparan bagaimana pengambilan keputusan dilakukan berdasarkan nilai sensor kualitas udara. Dengan membaca jalur dari akar ke daun, dapat disusun aturan logis berbentuk IF-THEN yang dapat dipahami dengan mudah oleh manusia, seperti: “Jika nilai pm10 \leq 30.5 dan co \leq 4.2, maka kualitas udara diklasifikasikan sebagai BAIK.” Hal ini menjadikan model sangat interpretatif, yang merupakan keunggulan utama dari algoritma *Decision Tree* dibanding algoritma lain yang bersifat black box.

Secara keseluruhan, visualisasi pohon keputusan ini menunjukkan bahwa model mampu mengidentifikasi batas-batas yang jelas antar kelas kualitas udara berdasarkan parameter ISPU. Dengan pemisahan data yang cukup dalam dan kompleks, model ini memiliki ketelitian tinggi dalam data pelatihan, meskipun perlu dilakukan validasi lebih lanjut untuk menghindari *overfitting*, terutama jika pohon terlalu dalam atau terlalu spesifik.

```

=== Classification Report ===

```

	precision	recall	f1-score	support
BAIK	0.94	1.00	0.97	287
SANGAT TIDAK SEHAT	1.00	1.00	1.00	1
SEDANG	1.00	0.99	0.99	1651
TIDAK SEHAT	1.00	0.99	0.99	167
accuracy			0.99	2106
macro avg	0.98	0.99	0.99	2106
weighted avg	0.99	0.99	0.99	2106

Gambar 5. *Classification Report*

Berdasarkan hasil evaluasi model klasifikasi menggunakan algoritma Decision Tree terhadap data kualitas udara, diperoleh hasil classification report dengan performa yang sangat baik. Untuk kelas "BAIK", nilai precision mencapai 0,94, yang menunjukkan bahwa dari seluruh data yang diprediksi sebagai "BAIK", sebesar 94% benar-benar sesuai dengan kelas aslinya. Sementara itu, nilai *recall* mencapai 1,00, yang berarti seluruh data yang benar-benar tergolong "BAIK" berhasil dikenali dengan tepat oleh model. *F1-score* untuk kelas ini adalah 0,97, yang mencerminkan keseimbangan yang baik antara precision dan recall. Adapun jumlah data aktual (*support*) untuk kelas ini sebanyak 287 data.

Pada kelas "SANGAT TIDAK SEHAT", seluruh metrik evaluasi yaitu *precision*, *recall*, dan *f1-score* menunjukkan nilai sempurna, yakni 1,00. Ini berarti model secara akurat mampu mengidentifikasi satu data yang tergolong dalam kelas ini tanpa kesalahan. Meskipun demikian, karena jumlah data untuk kelas ini hanya satu, nilai evaluasinya cenderung bias dan tidak cukup mewakili kemampuan model secara umum dalam mengklasifikasikan kelas tersebut jika terdapat lebih banyak data.

Selanjutnya, untuk kelas "SEDANG" yang merupakan kelas dengan jumlah data terbanyak yaitu sebanyak 1.651 data, model menunjukkan performa yang sangat tinggi dengan nilai precision sebesar 1,00, yang berarti semua prediksi model terhadap kelas "SEDANG" benar. Namun, *recall*-nya sedikit lebih rendah yaitu 0,99, menunjukkan bahwa ada sebagian kecil data yang sebenarnya "SEDANG" namun diklasifikasikan ke kelas lain. F1-score untuk kelas ini berada pada angka 0,99, yang tetap menggambarkan akurasi prediksi yang sangat baik.

Untuk kelas "TIDAK SEHAT", model juga memperlihatkan performa yang sangat memuaskan dengan precision dan recall masing-masing sebesar 1,00 dan 0,99, serta f1-score 0,99. Dari 167 data yang tergolong "TIDAK SEHAT", hampir seluruhnya dapat diklasifikasikan dengan tepat oleh model.

Secara keseluruhan, model menghasilkan akurasi sebesar 0,99, yang berarti 99% data dapat diklasifikasikan dengan benar. Nilai *macro average* untuk *precision* dan *recall* masing-masing sebesar 0,98 dan 0,99, serta nilai *f1-score* juga 0,99. *Macro average* ini menunjukkan rata-rata kinerja model tanpa memperhitungkan jumlah data per kelas. Sementara itu, *weighted average* untuk *precision*, *recall*, dan *f1-score* semuanya sebesar 0,99, yang merupakan rata-rata yang mempertimbangkan jumlah data pada masing-masing kelas. Hal ini menegaskan bahwa model bekerja sangat baik dalam konteks distribusi data yang tidak seimbang sekalipun. Dengan performa evaluasi yang sangat tinggi ini, dapat disimpulkan bahwa model *Decision Tree* yang dibangun sangat efektif dalam melakukan klasifikasi kualitas udara berdasarkan data yang digunakan.

KESIMPULAN

Penelitian ini berhasil merancang dan mengimplementasikan sistem klasifikasi kualitas udara berbasis Internet of Things (IoT) yang terintegrasi dengan algoritma Decision Tree. Sistem ini mampu mengumpulkan data parameter pencemar seperti PM10, PM2.5, SO₂, dan CO menggunakan sensor berbasis ESP32 dan menyimpannya untuk keperluan klasifikasi. Model klasifikasi dilatih menggunakan dataset ISPU Jakarta, kemudian diuji dengan data aktual dari lingkungan kampus Universitas Muhammadiyah Kalimantan Timur, yaitu dari area kantin dan parkir. Hasil evaluasi menunjukkan bahwa model Decision Tree mampu mengklasifikasikan data dengan akurasi sangat tinggi, mencapai 99%. Nilai precision dan recall yang tinggi tercatat pada seluruh kelas utama, terutama "Sedang" dan "Tidak Sehat". Visualisasi pohon keputusan memperlihatkan bahwa fitur PM10 dan CO merupakan atribut paling dominan dalam proses klasifikasi. Selain itu, confusion matrix menunjukkan tingkat kesalahan klasifikasi yang sangat rendah. Dengan demikian, dapat disimpulkan bahwa sistem klasifikasi berbasis IoT dan Decision Tree ini sangat efektif dan berpotensi digunakan untuk sistem monitoring kualitas udara secara praktis, khususnya di lingkungan terbatas seperti kampus.

Saran

1. Perluasan lokasi dan waktu pengambilan data: disarankan untuk menambah lokasi pengambilan data di luar lingkungan kampus serta melakukan pengambilan data di waktu yang bervariasi untuk memperkaya variasi dan representasi kondisi udara.
2. Peningkatan jumlah data kelas minoritas: perlu dilakukan penambahan data pada kelas yang jarang muncul seperti sangat tidak sehat, melalui simulasi data, penambahan titik pantau, atau teknik oversampling agar model dapat mengenali semua kategori secara seimbang.
3. Penggunaan model ensemble: untuk meningkatkan akurasi dan mengurangi bias klasifikasi, penelitian lanjutan disarankan menggunakan model ensemble seperti random forest atau gradient boosting yang lebih stabil terhadap data yang tidak seimbang.
4. Integrasi sistem notifikasi dini: sistem dapat dikembangkan lebih lanjut dengan fitur peringatan dini berbasis notifikasi otomatis apabila kualitas udara memburuk, agar pengguna dapat segera mengambil tindakan preventif.
5. Validasi dengan data resmi: disarankan untuk membandingkan hasil klasifikasi model dengan data dari instansi resmi seperti dinas lingkungan hidup atau bmkg untuk meningkatkan validitas ilmiah dan kepercayaan terhadap sistem yang dikembangkan.

DAFTAR PUSTAKA

- Budianto, H., & Sumanto, B. (2024). Perancangan Sistem Monitoring Kualitas Udara dalam Ruang Berbasis Internet of Things. *Jurnal Listrik, Instrumentasi, dan Elektronika Terapan*, 5(1), 9–17.
- Hidayah, R. R., Nurcahyo, S., & Dewatama, D. (2024). Implementasi Pengaturan Suhu Menggunakan Mikrokontroler ESP32. *METROTECH*, 3(3), 106–115.
- Irwansyah, I., Wiranata, A. D., & M, T. T. (2023). Komparasi Algoritma Decision Tree, Naive Bayes Dan K-Nearest Neighbor Untuk Menentukan Kualitas Udara Di Provinsi Dki Jakarta. *Infotech: Journal of Technology Information*, 9(2), 193–198. <https://doi.org/10.37365/jti.v9i2.203>
- Junaedi, A., Puspitasari, M. D. M., & Maulidina, M. (2021). Pengaruh (Intensor) Induktor Heater Menggunakan Thermal Sensor Berbasis Mikrokontroler Arduino Nano Dalam Mengolah Logam. *Nusantara of Engineering (NOE)*, 4(2), 169–175. <https://doi.org/10.29407/noe.v4i2.16754>
- Mariano, K. D. P., Almada, F. L. N., & Dutra, M. A. (2024). Smart Air Quality Monitoring for Automotive Workshop Environments. *arXiv*. <https://doi.org/10.48550/arXiv.2410.03986>
- Muttaqin, R., Prayitno, W. S. W., Setyaningsih, N. E., & Nurbaiti, U. (2024). Rancang Bangun Sistem Pemantauan Kualitas Udara Berbasis Iot (Internet Of Things) dengan Sensor DHT11 dan Sensor MQ135. *Jurnal Pengelolaan Laboratorium Pendidikan (JPLP)*, 6(2), 102–115.
- Octaviano, A., Sofiana, S., Agustino, D. O., & Rosyani, P. (2022). Pemantauan Kualitas Udara Berbasis Internet Of Things. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 3(2), 147–156.
- Pradih, A. R., Az-Zahra, A. S., Lintang, A. M., Suci, M. A., & Putri, F. S. (2024). Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Kualitas Udara Di Kota Bandung. *SCIENTICA*, 2(8), 312–317.
- Ramadhon, R. N., Ogi, A., Agung, A. P., Putra, R., Febrihartina, S. S., & Firdaus, U. (2024). Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak Aktif pada Data Bank. *Karya Ilmiah Mahasiswa Bertauhid (Karimah Tauhid)*, 3(2), 1860–1874. <https://doi.org/10.30997/karimahtauhid.v3i2.11952>
- Rumampuk, G. C., Poekoel, V. C., & Rumagit, A. M. (2021). Internet of Things-Based Indoor Air Quality Monitoring System Design Perancangan Sistem Monitoring Kualitas Udara Dalam Ruang Berbasis Internet of Things. *Jurnal Teknik Informatika*, 17, 11–18.
- Subagiyo, H., Wahyuni, R. T., Akbar, M., & Ulfa, F. (2020). Rancang Bangun Sensor Node untuk Pemantauan Parameter Kualitas Udara. *Jurnal Sains dan Teknologi Industri*, 18(1), 72–79.