

Klasifikasi Kualitas Udara di Jakarta Pada Bulan Agustus 2024 Menggunakan Algoritma C4.5

Ika Juni Nur Ayu¹, Nayla Rahmania Putri², Ridho Putra Nugraha³, Ryan Febriansyah⁴

^{1,2,3,4}Sistem Informasi, Universitas Bina Sarana Informatika, Indonesia

Email: ¹ikajuninurayu@gmail.com, ²naylarahmaniaputri14@gmail.com, ³ridhoputra2519@gmail.com, ⁴ryanfebriansyah556@gmail.com

ABSTRAK

Udara merupakan salah satu faktor penting dalam keberlangsungan hidup manusia selain tanah dan air. Kualitas udara yang buruk akan mengakibatkan dampak negatif terhadap kesehatan manusia, keseimbangan ekosistem dan perubahan iklim. Penelitian ini bertujuan untuk mengklasifikasi kualitas udara yang ada di Jakarta pada bulan Agustus 2024 menggunakan teknik data mining dengan algoritma C4.5. Data yang dianalisis diperoleh dari website Satu Data Jakarta yang diterbitkan pada 19 Februari 2024, mencakup beberapa parameter pengukuran, yaitu `pm_sepuluh`, `pm_duakomalima`, `sulfur_dioksida`, `karbon_monoksida`, `ozon` dan `nitrogen_dioksida`. Dalam implementasinya, penelitian ini menggunakan `tools` RapidMiner untuk memproses dan menganalisis data. Hasil klasifikasi menunjukkan bahwa kualitas udara di Jakarta pada periode tersebut dapat dikategorikan menjadi dua kelompok, yaitu tidak sehat dan sedang, dengan mayoritas pengukuran masuk dalam kategori sedang. Model klasifikasi yang dihasilkan mencapai tingkat akurasi 99,35%, menunjukkan bahwa algoritma C4.5 sangat efektif dalam mengidentifikasi dan memprediksi kualitas udara di Jakarta. Hasil ini menunjukkan bahwa sebagian besar data pengukuran kualitas udara di Jakarta masih berada dalam kategori yang memenuhi standar kualitas udara yang baik.

Kata Kunci: Kualitas Udara, Algoritma C4.5, Data Mining, Jakarta

ABSTRACT

Air is one of the important factors in human survival besides land and water. Poor air quality will have a negative impact on human health, ecosystem balance and climate change. This study aims to classify air quality in Jakarta in August 2024 using data mining techniques with the C4.5 algorithm. The data analyzed was obtained from the Satu Data Jakarta website published on February 19, 2024, including several measurement parameters, namely `pm_sepuluh`, `pm_duakomalima`, `sulfur_dioxide`, `carbon_monoxide`, `ozone` and `nitrogen_dioxide`. In its implementation, this research uses RapidMiner tools to process and analyze data. The classification results show that air quality in Jakarta during the period can be categorized into two groups, namely unhealthy and moderate, with the majority of measurements falling into the moderate category. The resulting classification model achieved an accuracy rate of 99.35%, indicating that the C4.5 algorithm is very effective in identifying and predicting air quality in Jakarta. This result shows that most of the air quality measurement data in Jakarta is still in a category that meets good air quality standards.

Keywords: Air Quality, C4.5 Algorithm, Data Mining, Jakarta

Penulis Korespondensi:

Ryan Febriansyah

Email: ryanfebriansyah556@gmail.com

Article Info

Diterima: 18 Desember 2024

Direvisi: 8 Januari 2025

Disetujui: 14 Januari 2025

This is an open access article under the [CC BY](#) license.



1. PENDAHULUAN

Udara merupakan salah satu faktor penting dalam keberlangsungan hidup manusia selain tanah dan air. Kualitas udara yang buruk akan mengakibatkan dampak negatif terhadap kesehatan manusia, keseimbangan ekosistem dan perubahan iklim. Jakarta merupakan kota besar di Indonesia yang menghadapi masalah terkait kualitas udara yang harus ditangani. *Data mining*

adalah salah satu teknik yang digunakan untuk menggali informasi yang berharga dari data yang besar dan kompleks. Teknik yang akan digunakan dalam penelitian ini adalah *data mining*.

Data mining adalah teknik analisis data otomatis untuk menemukan atau memisahkan hubungan dari sejumlah besar data yang sebelumnya tidak diketahui [1]. *Data mining* merupakan salah satu teknik untuk menggali atau “menambang” pengetahuan dari sekumpulan besar data.

Data Mining mempunyai 5 fungsi: *classification*, yaitu menyimpulkan definisi-definisi karakteristik sebuah grup. Contoh: pelanggan-pelanggan perusahaan yang telah berpindah kesainan perusahaan yang lain. *Clustering*, yaitu mengidentifikasi kelompok-kelompok dari barang-barang atau produk-produk yang mempunyai karakteristik khusus (*clustering* berbeda dengan *classification*, dimana pada *clustering* tidak terdapat definisi-definisi karakteristik awak yang di berikan pada waktu *classification*.) *Association*, yaitu mengidentifikasi hubungan antara kejadian-kejadian yang terjadi pada suatu waktu, seperti isi-isi dari keranjang belanja. *Sequencing*, hampir sama dengan *association*, *sequencing* mengidentifikasi hubungan-hubungan yang berbeda pada suatu periode waktu tertentu, seperti pelanggan-pelanggan yang mengunjungi supermarket secara berulang-ulang. *Forecasting* memperkirakan nilai pada masa yang akan datang berdasarkan pola-pola dengan sekumpulan data yang besar, seperti peramalan permintaan pasar [2].

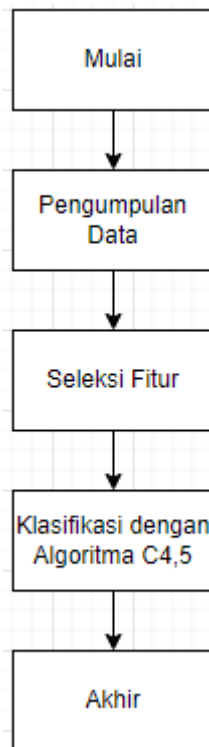
Salah satu metode *data mining* yang paling banyak digunakan adalah algoritma C4.5. Algoritma C4.5 adalah program yang menghasilkan pohon keputusan sebagai keluaran dan menyediakan sekumpulan data yang dilabelkan [3]. Pohon keputusan adalah sebuah struktur data yang terdiri dari simpul (node) dan rusuk (edge) simpul pada sebuah pohon dibedakan menjadi tiga, yaitu simpul akar (root node), simpul percabangan (branch node) dan simpul daun atau leaf node [4]. Manfaat utama menggunakan pohon keputusan adalah kemampuan untuk menyederhanakan proses pengambilan keputusan yang kompleks sehingga pengambil keputusan dapat menafsirkan solusi untuk masalah [5]. Dimana Pohon keputusan termasuk metode klasifikasi dan prediksi yang terkenal karena sangat mudah dipahami dengan bahasa alami sehingga dapat diinterpretasikan dengan cepat [6].

Dalam hal ini, penerapan algoritma C4.5 menjadi jawaban yang efektif dalam mengklasifikasikan kualitas udara di Jakarta. Dengan menganalisa data udara pada dataset yang diperoleh dari website Satu Data Jakarta algoritma C4.5 dapat menemukan pola dan hubungan pada data kualitas udara.

Oleh karena itu, tujuan penelitian ini adalah mengimplementasikan algoritma C4.5 pada data kualitas udara Jakarta untuk mengidentifikasi faktor-faktor yang mempengaruhi kualitas udara dan memberikan rekomendasi untuk meningkatkan kualitas udara di wilayah tersebut.

2. METODE PENELITIAN

Gambar di bawah ini adalah alur tahapan yang akan dilakukan.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Penelitian ini memanfaatkan *dataset* yang berasal dari website Satu data Jakarta yang diterbitkan pada 19 Februari 2024. *Dataset* berisi informasi seperti pengukuran PM10, PM2.5, SO₂, CO, O₃ dan NO₂. PM10 merupakan campuran asap yang kompleks dan heterogen, jelaga, debu, garam, asam dan logam yang mempunyai konsentrasi bervariasi, ukuran, komposisi kimia, luas permukaan dan sumber asalnya [7]. PM2,5 merupakan salah satu zat pencemar yang bebas di udara dan bertahan di udara cukup lama, ukurannya kecil hanya 2,5 mikron dapat masuk kedalam tubuh manusia lewat sistem pernafasan [8]. SO₂ merupakan gas tak berwarna yang memiliki toksisitas tinggi dan berbahaya bagi kesehatan terutama terhadap orang tua, anak-anak dan penderita yang mengalami penyakit kronis pada sistem pernafasan dan kardiovaskular [9].

Karbon monoksida (CO) merupakan gas yang tidak memiliki warna dan bau, yang dapat membahayakan apabila terhirup dengan jumlah yang besar. Sumber gas CO di udara adalah mobil, truk dan kendaraan lainnya, barang yang ada di rumah seperti pemanas ruangan dengan minyak tanah, cerobong asap, dan tungku yang bocor, kompor gas, asap rokok merupakan benda yang dapat melepaskan gas CO di dalam ruangan [10]. Ozon (O₃) merupakan suatu zat gas yang terkandung pada udara yang kita hirup. Ozon yang buruk terbentuk di dekat permukaan tanah ketika polutan (yang disebabkan oleh mobil, pembangkit listrik, industri, kilang minyak, dan pabrik kimia) bereaksi secara kimia dengan cahaya matahari [11]. NO₂ merupakan salah satu jenis polutan yang dapat menurunkan kualitas udara dan berdampak pada kesehatan manusia dan lingkungan. Gas NO₂ termasuk prekursor utama dalam pembentukan Ozon (O₃) di atmosfer [12].

2.2. Seleksi Fitur

Seleksi fitur bertujuan guna mengurangi atribut, sebab banyaknya fitur pada *dataset* yang digunakan dapat mengakibatkan *overfitting* dan perlu menentukan fitur yang dibutuhkan dalam pengolahan data. Selain itu, dilakukan normalisasi dan pemilihan atribut yang paling berpengaruh dalam klasifikasi kualitas udara [13].

2.3. Klasifikasi Dengan Algoritma C4.5

Klasifikasi merupakan salah satu teknik dalam data mining. Klasifikasi (taksonomi) merupakan proses penempatan objek atau konsep tertentu ke dalam satu set kategori berdasarkan objek yang digunakan [14]. Algoritma C4.5 dapat menangani atribut kontiniu dan diskrit, dapat menyelesaikan data tanpa nilai dari atribut., serta dapat meminimalkan pohon keputusan sesudah dibuat [15]. Dalam Penelitian ini untuk memodelkan pohon keputusan peneliti menggunakan *tools* rapidminer. RapidMiner adalah platform perangkat lunak yang kuat untuk ilmu data dan pembelajaran mesin. Ini menyediakan beragam alat untuk persiapan data, pemodelan, evaluasi, dan implementasi. RapidMiner dirancang untuk mudah digunakan dan memungkinkan pengguna untuk dengan mudah membangun dan menguji berbagai model, bahkan tanpa pengalaman pemrograman [16]

3. HASIL DAN PEMBAHASAN

3.1. Pre-Processing

Dataset yang digunakan didapatkan dari website Satu Data Jakarta yang diterbitkan pada 19 Februari 2024 dalam bentuk format XLSX (*Microsoft Excel*). *Dataset* ini memuat pengukuran kualitas udara di Provinsi DKI Jakarta yang diterbitkan pada tahun 2024. Berikut penjelasan parameter yang terdapat di dalam *dataset*:

- periode_data: penjelasan periode data 1 bulan sekali.
- bulan: pengambilan data pengukuran udara.
- tanggal: tanggal pengambilan data pengukuran udara.
- stasiun: lokasi penempatan alat pengukuran udara.
- pm_sepuluh: partikulat dengan ukuran di bawah 10 mikron.
- pm_duakomalima: partikulat dengan ukuran di bawah 2.5 mikron.
- karbon_monoksida: pengukuran zat karbon monoksida (CO).
- ozon: pengukuran zat ozon (O₃).
- nitrogen_dioksida: pengukuran zat nitrogen dioksida (NO₂).
- max: hasil pengukuran tertinggi pada setiap parameter yang ditangkap oleh stasiun pemantau.
- parameter_pencemaran_kritis: nama parameter yang memiliki nilai tertinggi.
- kategori: kategori hasil pengukuran yang didapat dari perhitungan rata-rata setiap parameter yang kemudian diklasifikasikan berdasarkan Indeks Standar Pencemaran Udara.

Tabel 1. Tabel ISPU

No	Nilai ISPU	Kategori
1	0 – 15	Baik
2	15 – 65	Sedang
3	66 – 150	Tidak Sehat
4	151 – 250	Sangat Tidak Sehat
5	>250	Berbahaya

Secara keseluruhan, *dataset* ini memuat 155 *record* data dengan 12 atribut dan 1 kelas. Atribut yang terdapat pada *dataset* ini, yaitu *periode_data*, *bulan*, *tanggal*, *stasiun*, *pm_sepuluh*, *pm_duakomalima*, *sulfur_dioksida*, *karbon_monoksida*, *ozon*, *nitrogen_dioksida*, *max*, *parameter_pencemar_kritis* dan *kategori*.

Tabel 2. Tabel Penggabungan Data

periode_data	bulan	tanggal	stasiun	pm_sepuluh	pm_duakomalima	sulfur_dioksida	karbon_monoksida	ozon	nitrogen_dioksida	max	parameter_pencemar_kritis	kategori
202408	8	1	DKI1 Bundaran Hotel Indonesia (HI)	68	103	13	28	26	44	103	PM25	TIDAK SEHAT
202408	8	2	DKI1 Bundaran Hotel Indonesia (HI)	60	85	11	27	37	44	85	PM25	SEDANG
202408	8	3	DKI1 Bundaran Hotel Indonesia (HI)	47	64	10	28	26	45	64	PM25	SEDANG
202408	8	4	DKI1 Bundaran Hotel Indonesia (HI)	50	74	10	29	29	39	74	PM25	SEDANG
202408	8	5	DKI1 Bundaran Hotel Indonesia (HI)	53	78	10	26	23	37	78	PM25	SEDANG
202408	8	6	DKI1 Bundaran Hotel Indonesia (HI)	53	75	16	24	21	42	75	PM25	SEDANG
202408	8	7	DKI1 Bundaran Hotel Indonesia (HI)	62	91	13	26	27	42	91	PM25	SEDANG
202408	8	8	DKI1 Bundaran Hotel Indonesia (HI)	64	96	16	28	23	42	96	PM25	SEDANG
202408	8	11	DKI1 Bundaran Hotel Indonesia (HI)	56	77	12	22	28	28	77	PM25	SEDANG
202408	8	12	DKI1 Bundaran Hotel Indonesia (HI)	54	78	12	23	24	30	78	PM25	SEDANG
202408	8	13	DKI1 Bundaran Hotel Indonesia (HI)	62	93	19	26	25	44	93	PM25	SEDANG
202408	8	14	DKI1 Bundaran Hotel Indonesia (HI)	62	93	17	26	27	43	93	PM25	SEDANG
202408	8	15	DKI1 Bundaran Hotel Indonesia (HI)	63	90	14	26	38	38	90	PM25	SEDANG
202408	8	16	DKI1 Bundaran Hotel Indonesia (HI)	64	91	15	28	33	48	91	PM25	SEDANG
202408	8	17	DKI1 Bundaran Hotel Indonesia (HI)	65	92	15	25	31	44	92	PM25	SEDANG
202408	8	18	DKI1 Bundaran Hotel Indonesia (HI)	62	93	13	23	30	32	93	PM25	SEDANG
202408	8	19	DKI1 Bundaran Hotel Indonesia (HI)	60	86	13	24	32	38	86	PM25	SEDANG
202408	8	20	DKI1 Bundaran Hotel Indonesia (HI)	64	93	14	26	27	46	93	PM25	SEDANG
202408	8	21	DKI1 Bundaran Hotel Indonesia (HI)	67	96	17	27	27	47	96	PM25	SEDANG
202408	8	22	DKI1 Bundaran Hotel Indonesia (HI)	63	84	19	25	26	40	84	PM25	SEDANG
202408	8	23	DKI1 Bundaran Hotel Indonesia (HI)	56	77	12	23	28	32	77	PM25	SEDANG
202408	8	24	DKI1 Bundaran Hotel Indonesia (HI)	62	98	13	27	24	43	98	PM25	SEDANG
202408	8	25	DKI1 Bundaran Hotel Indonesia (HI)	66	90	13	25	28	37	90	PM25	SEDANG
202408	8	26	DKI1 Bundaran Hotel Indonesia (HI)	55	78	13	24	28	39	78	PM25	SEDANG
202408	8	27	DKI1 Bundaran Hotel Indonesia (HI)	69	106	14	30	36	50	106	PM25	TIDAK SEHAT
202408	8	28	DKI1 Bundaran Hotel Indonesia (HI)	62	86	14	24	33	41	86	PM25	SEDANG
202408	8	29	DKI1 Bundaran Hotel Indonesia (HI)	65	98	14	26	41	42	98	PM25	SEDANG
202408	8	30	DKI1 Bundaran Hotel Indonesia (HI)	68	98	15	28	30	46	98	PM25	SEDANG
202408	8	31	DKI1 Bundaran Hotel Indonesia (HI)	56	75	14	25	24	40	75	PM25	SEDANG

Proses selanjutnya menyeleksi data. Pada penelitian penulis mengambil 7 kolom dengan 6 atribut, yaitu *pm_sepuluh*, *pm_duakomalima*, *sulfur_dioksida*, *karbon_monoksida*, *ozon*, *nitrogen_dioksida* dan *kategori*. Kemudian masuk ke tahap pembersihan data untuk menghapus, memperbaiki dan mengidentifikasi data yang tidak akurat atau tidak valid dengan melakukan penataan ulang data yang terdapat dalam catatan, tabel atau *database*. Sampel yang perlu dilakukan pembersihan data adalah sebagai berikut:

Tabel 3. Tabel Pembersihan Data

pm_sepuluh	pm_duakomalima	sulfur_dioksida	karbon_monoksida	ozon	nitrogen_dioksida	kategori
68	103	13	28	26	44	TIDAK SEHAT
60	85	11	27	37	44	SEDANG
47	64	10	28	26	45	SEDANG
50	74	10	29	29	39	SEDANG
53	78	10	26	23	37	SEDANG
53	75	16	24	21	42	SEDANG
62	91	13	26	27	42	SEDANG
64	96	16	28	23	42	SEDANG
61	91	15	25	19	33	SEDANG
55	81	15	23	21	35	SEDANG
56	77	12	22	28	28	SEDANG
54	78	12	23	24	30	SEDANG
62	93	19	26	25	44	SEDANG
62	93	17	26	27	43	SEDANG
63	90	14	26	38	38	SEDANG
64	91	15	28	33	48	SEDANG
65	92	15	25	31	44	SEDANG
62	93	13	23	30	32	SEDANG
60	86	13	24	32	38	SEDANG
64	93	14	26	27	46	SEDANG
67	96	17	27	27	47	SEDANG
63	84	19	25	26	40	SEDANG
56	77	12	23	28	32	SEDANG
62	98	13	27	24	43	SEDANG
66	90	13	25	28	37	SEDANG
55	78	13	24	28	39	SEDANG
69	106	14	30	36	50	TIDAK SEHAT
62	86	14	24	33	41	SEDANG
65	98	14	26	41	42	SEDANG

Setelah dilakukan pembersihan data maka didapatkan data yang akan siap diolah. Data inilah yang akan dilakukan klasifikasi menggunakan *Tools Rapidminer*. Berikut tampilan data siap olah:

Tabel 4. Tabel Siap Olah

pm_sepuluh	pm_duakomalima	sulfur_dioksida	karbon_monoksida	ozon	nitrogen_dioksida	kategori
68	103	13	28	26	44	TIDAK SEHAT
60	85	11	27	37	44	SEDANG
47	64	10	28	26	45	SEDANG
50	74	10	29	29	39	SEDANG
53	78	10	26	23	37	SEDANG
53	75	16	24	21	42	SEDANG
62	91	13	26	27	42	SEDANG
64	96	16	28	23	42	SEDANG
61	91	15	25	19	33	SEDANG
55	81	15	23	21	35	SEDANG
56	77	12	22	28	28	SEDANG
54	78	12	23	24	30	SEDANG
62	93	19	26	25	44	SEDANG
62	93	17	26	27	43	SEDANG
63	90	14	26	38	38	SEDANG
64	91	15	28	33	48	SEDANG
65	92	15	25	31	44	SEDANG
62	93	13	23	30	32	SEDANG
60	86	13	24	32	38	SEDANG
64	93	14	26	27	46	SEDANG
67	96	17	27	27	47	SEDANG
63	84	19	25	26	40	SEDANG
56	77	12	23	28	32	SEDANG
62	98	13	27	24	43	SEDANG
66	90	13	25	28	37	SEDANG
55	78	13	24	28	39	SEDANG
69	106	14	30	36	50	TIDAK SEHAT
62	86	14	24	33	41	SEDANG
65	98	14	26	41	42	SEDANG

3.2. Proses Klasifikasi dan Pohon Keputusan

Algoritma C4.5 menerapkan pohon keputusan dengan langkah-langkah berikut. Proses pertama adalah memilih atribut sebagai akar pohon keputusan. Selanjutnya, membuat cabang untuk setiap nilai berdasarkan akar tersebut. Kemudian, bagi kasus-kasus tersebut ke dalam cabang. Proses ini diulang pada cabang hingga tercapai kelas yang sama. Algoritma C4.5 menggunakan metode gain untuk menentukan atribut akar. Gain dihitung untuk semua atribut yang ada, apabila atribut dengan nilai gain tertinggi maka akan dipilih sebagai atribut akar pohon keputusan. Maka untuk menghitung klasifikasi data dengan algoritma C4.5 dapat dihitung gain dan entropi untuk membuat pohon keputusan dari semua nilai yang ada pada tabel data. Berikut tampilan hasil perhitungan gain dan entropi.

Tabel 5. Tabel Gain dan Entropi

Attribute		Jumlah (s)	Kepuasan			Entropy	Gain
			Baik	Tidak sehat	Sedang		
Total		155	1	32	122	0,78870534	
pm_sepuluh	<= 50	19	1	0	18	0,00000000	0,788705
	> 50	136	0	104	32	0,00000000	
pm_duakomalima	<=100	122	1	0	121	0,00000000	0,788705
	>100	33	0	33	0	0,00000000	
Sulfur dioksida	<=50	63	1	16	46	0,92832992	0,411384
	>50	92	0	15	77	0,00000000	
Karbon_monoksida	<= 25	135	1	30	104	0,82457524	0,070527
	>25	20	0	2	18	0,00000000	
Ozon	<=50	143	1	28	114	0,77137204	0,077052
	>50	12	0	6	6	0,00000000	
nitrogen_dioksida	<=50	155	1	122	32	0,78870534	0,000000
	>50	0	0	0	0	0,00000000	

Hasil Perhitungan gambar 5 dapat dijabarkan sebagai berikut:

Rumus Entropi:

$$Entropy (S) = \sum_{i=1}^N p_i \log_2 (p_i) \quad (1)$$

Keterangan:

S = Himpunan Kasus

A = Atribut

N = Jumlah partisi s

Pi = Proporsi dari Si terhadap S

Berikut perhitungan entropi dari masing-masing atribut:

entropi total:

$$Entropy (Total) = ((-1/155) * \log_2 (1/155)) + ((-32/155) * \log_2 (32/155)) + ((-122/155) * \log_2 (122/155)) = 0,78870534$$

Entropi sulfur_dioksida:

$$Entropy (sulfur_dioksida) = ((-1/63) * \log_2 (1/63)) + ((-16/63) * \log_2 (16/63)) + ((-46/63) * \log_2 (46/63)) = 0,92832992$$

Entropi karbon_monoksida:

$$Entropy (karbon_monoksida) = ((-1/135) * \log_2 (1/135)) + ((-30/135) * \log_2 (30/135)) + ((-104/135) * \log_2 (104/135)) = 0,82457524$$

Entropi ozon:

$$Entropy (ozon) = ((-1/143) * \log_2 (1/143)) + ((-28/143) * \log_2 (28/143)) + ((-114/143) * \log_2 (114/143)) = 0,77137204$$

Entropi nitrogen_dioksida:

$$Entropy (nitrogen_dioksida) = ((-1/155) * \log_2 (1/155)) + ((-122/155) * \log_2 (122/155)) + ((-3/155) * \log_2 (32/155)) = 0,78870534$$

Lalu perhitungan untuk gain:

Rumus Gain:

$$Gain (S, A) = Entropy (S) - \sum_{i=1}^N p_i * Entropy (S_i) \quad (2)$$

Keterangan:

S = Himpunan Kasus

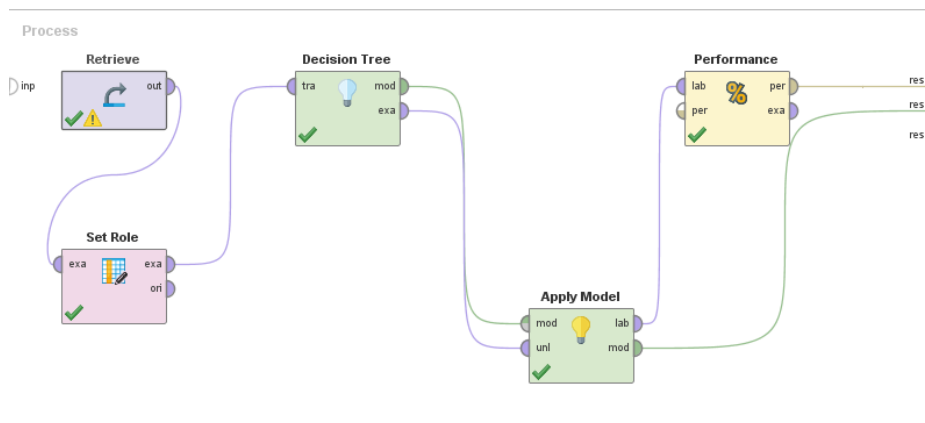
A = Atribut

N = Jumlah partisi pada atribut A

|Si| = Jumlah kasus pada partisi ke-i

$$Gain (pm_duakomalima) = 0,78870534 - ((122/155) * 0,00000000) - ((33/155) * 0,00000000) = 0,788705$$

Selanjutnya adalah klasifikasi menggunakan pohon keputusan pada *tools* rapidminer. Berikut adalah langkah-langkahnya:



Gambar 2. Proses Klasifikasi

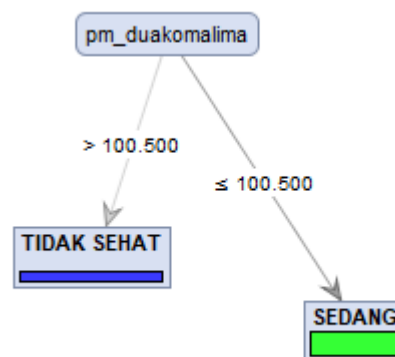
Sementara itu, pada gambar 7 di bawah ini adalah hasil dari proses klasifikasi pada gambar 6.

PerformanceVector

```
PerformanceVector:  
accuracy: 99.35%  
ConfusionMatrix:  
True:   TIDAK SEHAT   SEDANG   BAIK  
TIDAK SEHAT:   32       0       0  
SEDANG:  0       122      1  
BAIK:    0        0       0
```

Gambar 3. Output Klasifikasi

Akhirnya didapatkan pohon keputusan sebagai berikut.



Gambar 4. Pohon Keputusan

Berdasarkan pohon keputusan dari algoritma C4.5 untuk melakukan klasifikasi kualitas udara di Jakarta pada bulan Agustus, diperoleh nilai sebagai berikut:

1. Jika nilai `pm_duakomalima` lebih besar dari 100.500, maka kualitas udara masuk dalam kategori tidak sehat.
2. Jika nilai `pm_duakomalima` kurang dari atau sama dengan 100.500, maka kualitas udara masuk dalam kategori sedang.

4. KESIMPULAN

Berdasarkan pembahasan dan analisis data yang telah diuji menggunakan algoritma C4.5 untuk mengklasifikasi Kualitas udara di Jakarta pada bulan Agustus berdasarkan data yang diperoleh dari website Satu Data Jakarta, data tersebut berisi 7 parameter untuk mengukur kualitas udara, yaitu `pm_sepuluh`, `pm_duakomalima`, sulfur dioksida, karbon monoksida, ozon, nitrogen dioksida dan kategori. Dengan melakukan penerapan *data mining* menggunakan algoritma C4.5 untuk melakukan klasifikasi udara di Jakarta pada bulan Agustus, diketahui bahwa kualitas udara diklasifikasikan menjadi 2 kategori, yaitu tidak sehat dan sedang. Sehingga dapat disimpulkan bahwa kualitas udara di Jakarta pada bulan Agustus masuk ke dalam kategori sedang dengan tingkat akurasi mencapai 99,35%. Hasil ini menunjukkan bahwa model klasifikasi mampu mengidentifikasi dan memprediksi dengan tingkat akurasi yang tinggi dengan menunjukkan bahwa mayoritas data pengukuran kualitas udara berada dalam kategori yang mencukupi standar untuk kualitas udara yang baik.

REFERENSI

- [1] W. W. Kristianto and C. Rudianto, "Penerapan Data Mining Pada Penjualan Produk Menggunakan Metode K- Means Clustering (Studi Kasus Toko Sepatu Kakikaki)," *J. Pendidik. Teknol. Inf.*, no. 5, pp. 90–98, 2020.
- [2] C. Zai, "IMPLEMENTASI DATA MINING SEBAGAI PENGOLAHAN DATA," *J. Portal Data*, vol. 2, no. 3, pp. 1–12, 2022, [Online]. Available: <http://portaldata.org/index.php/portaldata/article/view/107>
- [3] M. Andreas Rayando, I. Fajri, F. Kurniawan, M. Stefanus Rafael Sihotang, and S. Suharjanti, "Analisis Tingkat Kepuasan Masyarakat Terhadap Kinerja Ketua Rw 007 Menggunakan Algoritma C4.5," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 4, pp.

- 7463–7469, 2024, doi: 10.36040/jati.v8i4.10334.
- [4] J. A. Sidette, E. Eko, and O. D. Nurhayati, “Pendekatan Metode Pohon Keputusan Menggunakan Algoritma ID3 Untuk Sistem Informasi Pengukuran Kinerja PNS,” *J. Sist. Inf. Bisnis*, vol. 4, no. 2, pp. 75–86, 2014, doi: 10.21456/vol4iss2pp75-86.
- [5] P. R. Adinda, “Pohon Keputusan C4. 5 Algoritma Untuk Klasifikasi Program Bantuan Belajar,” *J. Portal Data*, vol. 2, no. 9, pp. 1–13, 2022, [Online]. Available: <http://portaldata.org/index.php/portaldata/article/view/215%0Ahttp://portaldata.org/index.php/portaldata/article/download/215/204>
- [6] S. Ucha Putri, E. Irawan, F. Rizky, S. Tunas Bangsa, P. A. -Indonesia Jln Sudirman Blok No, and S. Utara, “Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4.5,” *Januari*, vol. 2, no. 1, pp. 39–46, 2021.
- [7] E. Nur, B. A. Seno, and R. Hidayanti, “Risk of Public Health Problems Due to PM10 Exposure in Padang City,” *J. Kesehatan Lingkungan. Indones.*, vol. 20, no. 2, pp. 97–103, 2021, [Online]. Available: <https://shorturl.asia/6u94Z>
- [8] A. E. Saputra, “Identifikasi Paparan PM2,5 Di Wilayah Kota Cimahi,” *J. Kesehat. Kartika*, vol. 16, no. 3, pp. 104–109, 2021, doi: 10.26874/jkkes.v16i3.205.
- [9] A. R. Nurfadillah and S. Petasule, “ENVIRONMENTAL HEALTH RISK ANALYSIS (SO2, NO2, CO and TSP) IN THE BONE BOLANGO AREA ROAD SEGMENT,” *J. Heal. Sci. Gorontalo J. Heal. Sci. Community*, vol. 6, no. 2, pp. 76–89, 2022, doi: 10.35971/gojhes.v5i3.13544.
- [10] M. A. Rizaldi, R. Azizah, M. T. Latif, L. Sulistyorini, and B. P. Salindra, “Literature Review: Dampak Paparan Gas Karbon Monoksida Terhadap Kesehatan Masyarakat yang Rentan dan Berisiko Tinggi,” *J. Kesehat. Lingkungan. Indones.*, vol. 21, no. 3, pp. 253–265, 2022, doi: 10.14710/jkli.21.3.253-265.
- [11] R. Wahyudi, S. Annas, and Z. Rais, “Analisis Support Vector Regression (Svr) Untuk Meramalkan Indeks Kualitas Udara Di Kota Makassar,” *Variansi J. Stat. Its Appl. Teach. Res.*, vol. 5, no. 3, pp. 104–117, 2023, doi: 10.35580/variasiunm107.
- [12] Y. Serlina, “Pengaruh Faktor Meteorologi Terhadap Konsentrasi NO2 di Udara Ambien (Studi Kasus Bundaran Hotel Indonesia DKI Jakarta),” *J. Serambi Eng.*, vol. 5, no. 3, 2020, doi: 10.32672/jse.v5i3.2146.
- [13] A. Z. D. Nur Adiya, A. F. Desvita, A. Fidela, D. Amelia, and T. Astuti, “Penerapan Data Mining Untuk Klasifikasi Kualitas Udara di Daerah Istimewa Yogyakarta Menggunakan Algoritma C4.5,” *JDMIS J. Data Min. Inf. Syst.*, vol. 2, no. 2, pp. 59–65, 2024, doi: 10.54259/jdmis.v2i2.2800.
- [14] D. Ardiansyah and W. Walim, “Algoritma c4.5 untuk klasifikasi calon peserta lomba cerdas cermat siswa smp dengan menggunakan aplikasi rapid miner,” *J. Inkofar*, vol. 1, no. 2, pp. 5–12, 2018, [Online]. Available: <https://politeknikmeta.ac.id/meta/ojs/index.php/inkofar/article/view/29/45>
- [15] Y. Hendra Kusuma, S. Suprapto, and Y. Setiawan, “Analisis Kepuasan Penumpang pada Maskapai Penerbangan Menggunakan Algoritma C4.5 dan Naïve Bayes,” *SENTIMAS Semin. Nas. Penelit. dan Pengabd. Masy.*, pp. 162–171, 2022, [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas/article/view/320/125>
- [16] M. Rafi Nahjan, Nono Heryana, and Apriade Voutama, “Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 101–104, 2023, doi: 10.36040/jati.v7i1.6094.