

# ANALISIS KLASIFIKASI DATA KUALITAS UDARA DKI JAKARTA MENGGUNAKAN ALGORITMA C.45

Maya Astriyani, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Ida Nur Laela, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Dwi Puji Lestari, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Laudiana Anggraeni, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Tri Astuti, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto

*Abstract*— Pertumbuhan dan perkembangan suatu kota merupakan salah satu faktor penyebab meningkatnya pencemaran udara karena kualitas udara yang sudah tercampur dengan berbagai komponen. Polusi udara merupakan suatu keadaan dimana substansi fisik, biologi atau kimia di lapisan udara bumi yang jumlahnya cukup banyak dapat menyebabkan bahaya bagi kesehatan tubuh manusia dan makhluk hidup lainnya. Salah satu kota dengan tingkat pencemaran udara tertinggi adalah DKI Jakarta. Dinas Lingkungan Hidup Pemerintah Provinsi DKI Jakarta mengoperasikan Stasiun Pemantau Kualitas Udara (SPKU) untuk memantau kualitas udara setiap hari. Penggunaan metode data *mining* digunakan untuk menganalisis faktor-faktor penyebab polusi udara di DKI Jakarta. Metode ini dapat mengolah data parameter ISPU menjadi informasi yang memberitahukan tingkat kualitas udara per harinya dengan menggunakan algoritma C.45 atau *decision tree*.

Kata Kunci : DKI Jakarta, polusi udara, ISPU, klasifikasi, *decision tree*

*Abstract* — *The growth and development of a city is one of the factors causing the increase in air pollution because the air quality is mixed with various components. Air pollution is a condition in which a large number of physical, biological or chemical substances in the earth's air can cause harm to the health of the human body and other living things. One of the cities with the highest level of air pollution is DKI Jakarta. The Environmental Service of the DKI Jakarta Provincial Government operates an Air Quality Monitoring Station (SPKU) to monitor air quality every day. The use of data mining methods is used to analyze the factors that cause air pollution in DKI Jakarta. This method can process ISPU parameter data into information that tells the level of air quality per day using the decision C45 or tree algorithm.*

*Keyword* : *DKI Jakarta, air pollution, ISPU, classification, decision tree*

## I. PENDAHULUAN

Perkembangan teknologi yang semakin maju berpengaruh dalam segala aspek kehidupan manusia baik di bidang ekonomi, politik, sosial dan budaya, bahkan di dunia pendidikan. Lingkungan yang sehat merupakan lingkungan yang mempunyai kualitas udara yang bersih dan nyaman ketika

dihirup. Bersihnya udara pada lingkungan yang sehat disebabkan oleh minimnya polusi atau pencemaran udara pada lingkungan tersebut. Oleh karena itu lingkungan yang sehat merupakan suatu kebutuhan yang penting bagi seluruh manusia untuk mempertahankan kehidupan.

Kota-kota besar di Indonesia yang memiliki pertumbuhan dan perkembangan penduduk yang padat dan pesat menjadi salah satu faktor terjadinya pencemaran udara. Lingkungan yang padat penduduk menyebabkan penurunan udara karena beberapa aktivitas yang dilakukan manusia seperti asap rokok, kegiatan industri, transportasi, pembakaran lahan, dan lain-lain. DKI Jakarta menjadi salah satu kota dengan kualitas udara yang tidak sehat, kualitas udara dapat diketahui melalui pengukuran kualitas udara yang dilakukan oleh Dinas Lingkungan Hidup Provinsi DKI Jakarta.

Analisis klasifikasi data kualitas udara di DKI Jakarta menggunakan data yang diambil dari *Website* Jakarta Open Data yang dilakukan oleh Dinas Lingkungan Hidup DKI Jakarta untuk memantau kualitas udara yang ada di DKI Jakarta. Pengukuran ini didokumentasikan dalam suatu indeks yang disebut Indeks Standar Pencemar Udara (ISPU). ISPU merupakan laporan dari hasil pemantauan kualitas udara yang memberikan keterangan mengenai kualitas udara yang bersih ataupun tercemar serta pengaruhnya terhadap kesehatan.

Dinas Lingkungan Hidup Provinsi DKI Jakarta memerlukan pengolahan data yang dapat memberikan informasi atau yang biasa disebut data *mining*. Data *mining* berfungsi untuk menggali informasi lebih dalam dengan jumlah data yang banyak. Implementasi data *mining* menjadi solusi untuk mengetahui kualitas udara DKI Jakarta menggunakan teknik klasifikasi. Klasifikasi merupakan sesuatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia.

Berdasarkan penjelasan di atas, pada penelitian ini akan melakukan klasifikasi kualitas udara di DKI Jakarta dengan algoritma C.45 atau *decision tree*. Penelitian ini diharapkan dapat menghasilkan informasi bagaimana kualitas udara di DKI Jakarta. Selain itu, tujuan penelitian ini adalah diharapkan hasil dari penelitian dapat menjadi acuan untuk Dinas Lingkungan Hidup dapat melakukan klasifikasi data kualitas udara berdasarkan atribut yang ada, agar pihak Dinas Lingkungan Hidup dapat melakukan penanganan lebih cepat terhadap pencemaran udara yang terjadi di DKI Jakarta.

II. LANDASAN TEORI

2.1. Polusi Udara

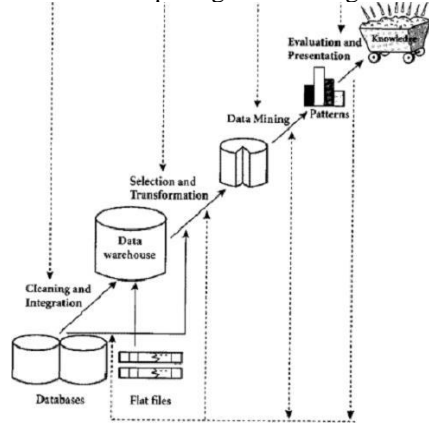
Polusi udara adalah suatu keadaan dimana substansi fisik, biologi atau kimia di lapisan udara bumi yang jumlahnya cukup banyak dapat menyebabkan bahaya bagi kesehatan tubuh manusia dan makhluk hidup lainnya. Polusi udara merupakan permasalahan dunia yang cukup menjadi perhatian di sebagian negara, mengingat dampak dan efeknya cukup besar untuk kesehatan masyarakat. Adanya polusi udara yang semakin serius dapat dikembangkan sistem peringatan dini untuk prakiraan kualitas udara guna memantau dan mengendalikan kualitas udara. Pencemaran udara dapat berakibat secara langsung terhadap kesehatan manusia seperti saluran pernafasan, iritasi mata, dan alergi kulit.

2.2. Data Mining

Data Mining merupakan proses menganalisis data dari berbagai sudut pandang dan meringkasnya menjadi sebuah informasi yang berguna. Data mining juga digunakan untuk menganalisis data dari berbagai dimensi, mengelompokkan, dan meringkas hubungan yang diidentifikasi. Beberapa tahun terakhir, mengolah data dengan teknis data mining dapat berubah dengan perkembangan teknologi yang canggih dan kemudahan pengguna dalam penggunaan alat untuk peningkatan jumlah para peneliti untuk menerapkan data mining. Hal penting menurut Novianti et al (2016) yang terkait dengan data mining adalah:

1. Data mining adalah pemrosesan otomatis data yang ada.
2. Data yang akan diolah memiliki bentuk data yang sangat besar.
3. Tujuan dari data mining adalah untuk menemukan hubungan atau pola yang dapat memberikan indikasi yang berguna.

Data mining merupakan serangkaian proses yang dapat dikelompokkan menjadi beberapa tahap. Tahap-tahap tersebut diilustrasikan pada gambar sebagai berikut :



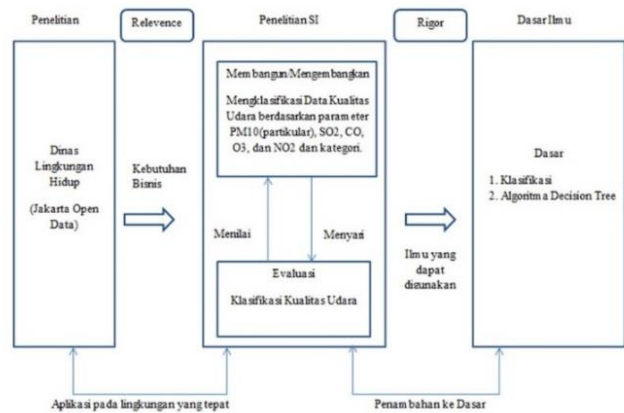
Gambar 1. Tahap-tahap data mining

2.3. Klasifikasi

Klasifikasi adalah proses pencarian sekumpulan model, pola, atau fungsi yang menggambarkan pada objek data untuk dikelompokkan kedalam kelas tertentu dari sejumlah kelas yang tersedia dihitung dengan menentukan persentase instance yang diberi label kelas yang benar dan nantinya dapat berguna untuk

memperkirakan kelas dari suatu objek yang kelasnya tidak diketahui. Klasifikasi penelitian dapat dilakukan dengan beberapa tinjauan yaitu dengan bidang ilmu, pendekatan, tempat pelaksanaan, tujuan umum, taraf, dan tidak adanya intervensi pada variabel yang ada.

III. METODE



Gambar 2. Model konseptual

Pada gambar di atas merupakan model konseptual yang berisi tentang data-data yang dibutuhkan dalam proses penelitian yang digunakan untuk menggambarkan konsep permasalahan yang akan diteliti agar mudah dipahami. Lingkungan yang terdapat pada penelitian ini adalah DKI Jakarta dimana dataset diperoleh dari website Jakarta Open Data. Pada bagian dasar ilmu ditampilkan metode data mining untuk mengklasifikasi data kualitas udara berdasarkan parameter PM10, PM25, SO<sub>2</sub>, CO, O<sub>3</sub>, NO<sub>2</sub>. Untuk melakukan klasifikasi yang akurat maka algoritma yang digunakan yaitu Algoritma C.45

IV. HASIL DAN PEMBAHASAN

4.1 Pre-Processing

Dataset yang digunakan didapatkan dari Website Jakarta Open Data dalam bentuk format CSV (Comma Separated Values). Dataset tersebut berbentuk file terpisah berdasarkan bulan dari bulan Januari hingga bulan Desember 2021. Dataset ini berisi mengenai Indeks Standar Pencemar Udara (ISPU) yang diukur dari 5 stasiun pemantau kualitas udara (SPKU) yang ada di Provinsi DKI Jakarta Tahun 2021. Penjelasan variabel atau parameter yang terdapat pada dataset adalah sebagai berikut :

- Tanggal: waktu pengukuran kualitas udara
- PM10: partikel udara yang berukuran lebih kecil dari 10 mikron (mikrometer)
- PM25: partikel udara yang berukuran lebih kecil dari 2.5 mikron (mikrometer)
- SO<sub>2</sub>: Sulfur dioksida adalah salah satu spesies dari gas-gas oksida sulfur
- CO: Karbon Monoksida

- O<sub>3</sub>: molekul anorganik radikal yang terdiri dari tiga atom oksigen yang bersifat oksidator kuat. Secara alamiah ozon dihasilkan dari molekul oksigen (O<sub>2</sub>) pada atmosfer bumi yang berinteraksi dengan sinar ultraviolet atau aktivitas elektrik pada atmosfer
- NO<sub>2</sub>: Nitrogen Dioksida
- Max: nilai ukur paling tinggi dari seluruh parameter yang diukur dalam waktu yang sama
- Critical: parameter yang hasil pengukurannya paling tinggi
- Kategori: diperoleh dari perhitungan rata-rata parameter PM10, PM25, SO<sub>2</sub>, CO, O<sub>3</sub>, dan NO<sub>2</sub>, kemudian diklasifikasi berdasarkan ISPU (Indeks Standar Pencemar Udara).

Tabel 1. Indeks 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 1825 record data dengan 10 atribut dan 1 kelas. Atribut-atribut yang terdapat dalam dataset ini antara lain tanggal, stasiun, pm10, pm25, so2, co, o3, no2, max, critical, dan kategori. Kemudian data digabungkan menjadi satu server basis data yang sama. Data sampel yang digabungkan seperti pada table berikut:

Tabel 2. Penggabungan Data

tanggal	stasiun	pm10	pm25	so2	co	o3	no2	max	critical	kategori
01/01/2021	DKI1 (Bunderan HI)	38	53	29	6	31	13	53	PM25	Sedang
02/01/2021	DKI1 (Bunderan HI)	27	46	27	7	47	7	47	O3	Baik
03/01/2021	DKI1 (Bunderan HI)	44	58	25	7	40	13	58	PM25	Sedang
04/01/2021	DKI1 (Bunderan HI)	30	48	24	4	32	7	48	PM25	Baik
05/01/2021	DKI1 (Bunderan HI)	38	53	24	6	31	9	53	PM25	Sedang
06/01/2021	DKI1 (Bunderan HI)	41	58	23	13	46	13	58	PM25	Sedang
07/01/2021	DKI1 (Bunderan HI)	35	47	22	6	39	10	47	PM25	Baik
08/01/2021	DKI1 (Bunderan HI)	37	54	26	16	17	10	54	PM25	Sedang
09/01/2021	DKI1 (Bunderan HI)	47	61	16	27	22	12	61	PM25	Sedang
10/01/2021	DKI1 (Bunderan HI)	23	25	16	11	33	8	33	O3	Baik
11/01/2021	DKI1 (Bunderan HI)	38	54	17	14	27	10	54	PM25	Sedang
12/01/2021	DKI1 (Bunderan HI)	29	50	20	12	26	15	50	PM25	Baik
13/01/2021	DKI1 (Bunderan HI)	34	44	17	13	20	9	44	PM25	Baik
14/01/2021	DKI1 (Bunderan HI)	36	57	15	15	26	13	57	PM25	Sedang
15/01/2021	DKI1 (Bunderan HI)	42	62	14	15	32	12	62	PM25	Sedang

Setelah proses penggabungan data, selanjutnya adalah data selection atau seleksi data. Pada penelitian ini mengambil sebanyak 7 kolom dengan 6 atribut parameter kualitas udara seperti PM10, PM25, SO<sub>2</sub>, CO, O<sub>3</sub>, NO<sub>2</sub>, dan kategori. Kemudian masuk ke tahap *cleaning* data untuk 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*. Hasil dari *cleaning* data dari 1825 baris data menghasilkan 1809 data. Sampel data yang perlu dilakukan *cleaning* data seperti pada tabel berikut:

Tabel 3. Sampel *Cleaning* Data

pm10	pm25	so2	co	o3	no2	kategori
---	72	17	15	56	4	Sedang
---	58	18	9	37	4	Sedang
26	47	16	6	39	3	Baik
29	78	20	14	40	8	Sedang
24	82	19	16	64	6	Sedang
70	92	19	16	60	7	Sedang
58	86	22	11	43	9	Sedang
51	64	21	7	38	5	Sedang
42	56	19	7	37	6	Sedang
46	---	17	17	41	10	Baik
52	---	18	19	44	9	Sedang
84	112	20	33	57	10	Tidak-Sehat
89	126	19	42	64	13	Tidak-Sehat
64	90	19	31	34	10	Sedang
60	95	18	23	67	8	Sedang

Setelah dilakukan proses *cleaning* data maka didapatkan data siap olah. Berikut tabel output yang dihasilkan setelah melakukan preprocessing data. Data inilah yang selanjutnya akan dilakukan klasifikasi menggunakan Tools Weka. Berikut tampilan data siap olah:

Tabel 4. Data Siap Olah

pm10	pm25	so2	co	o3	no2	kategori
---	72	17	15	56	4	Sedang
---	58	18	9	37	4	Sedang
26	47	16	6	39	3	Baik
29	78	20	14	40	8	Sedang
24	82	19	16	64	6	Sedang
70	92	19	16	60	7	Sedang
58	86	22	11	43	9	Sedang
51	64	21	7	38	5	Sedang
42	56	19	7	37	6	Sedang
46	---	17	17	41	10	Baik
52	---	18	19	44	9	Sedang
84	112	20	33	57	10	Tidak-Sehat
89	126	19	42	64	13	Tidak-Sehat
64	90	19	31	34	10	Sedang
60	95	18	23	67	8	Sedang

4.2 Proses Klasifikasi dan Pohon Keputusan

Algoritma C4.5 menggunakan pohon keputusan dengan cara yang pertama memilih atribut sebagai akar. Kemudian dibuat cabang untuk tiap-tiap nilai di dalam akar. Langkah selanjutnya yaitu dengan membagi kasus dalam cabang. Lakukan perulangan proses pada setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama. Dalam pemilihan atribut akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus sebagai berikut:

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S)$$

Keterangan :

S : Himpunan kasus

A : Atribut

N : Jumlah partisi atribut A

|S<sub>i</sub>| : Jumlah kasus pada partisi ke-i

|S| : Jumlah kasus dalam S

Sementara itu, penghitungan nilai entropi dapat dilihat pada persamaan :

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i$$

Keterangan :

S = Himpunan kasus

A = Atribut

N = Jumlah partisi S

Pi = Proporsi dari Si terhadap S

Berikut hasil klasifikasi dan perhitungan pohon keputusan di Tools Weka

ARFF-Viewer - C:\Users\mayat\Music\Klasifikasi-Udara.arff

File Edit View

Klasifikasi-Udara.arff

Relation: kualitas\_predicted

No.	1: pm10	2: pm25	3: so2	4: co	5: o3	6: no2	7: prediction margin	8: predicted class	9: class
	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Nominal	Nominal
1	38.0	53.0	29.0	6.0	31.0	13.0	0.971735	Sedang	Sedang
2	27.0	46.0	27.0	7.0	47.0	7.0	1.0	Baik	Baik
3	44.0	58.0	25.0	7.0	40.0	13.0	0.971735	Sedang	Sedang
4	30.0	48.0	24.0	4.0	32.0	7.0	1.0	Baik	Baik
5	38.0	53.0	24.0	6.0	31.0	9.0	0.811254	Sedang	Sedang
6	41.0	58.0	23.0	13.0	46.0	13.0	0.971735	Sedang	Sedang
7	35.0	47.0	22.0	6.0	39.0	10.0	1.0	Baik	Baik
8	37.0	54.0	26.0	16.0	17.0	10.0	0.971735	Sedang	Sedang
9	47.0	61.0	16.0	27.0	22.0	12.0	0.971735	Sedang	Sedang
10	23.0	25.0	16.0	11.0	33.0	8.0	1.0	Baik	Baik
11	38.0	54.0	17.0	14.0	27.0	10.0	0.971735	Sedang	Sedang
12	29.0	50.0	20.0	12.0	26.0	15.0	1.0	Baik	Baik
13	34.0	44.0	17.0	13.0	20.0	9.0	1.0	Baik	Baik
14	36.0	57.0	15.0	15.0	26.0	13.0	0.971735	Sedang	Sedang
15	42.0	62.0	14.0	15.0	32.0	12.0	0.971735	Sedang	Sedang
16	52.0	73.0	15.0	17.0	25.0	14.0	0.971735	Sedang	Sedang
17	52.0	71.0	17.0	15.0	26.0	14.0	0.971735	Sedang	Sedang
18	40.0	52.0	16.0	11.0	26.0	12.0	0.971735	Sedang	Sedang
19	31.0	41.0	13.0	12.0	25.0	12.0	1.0	Baik	Baik
20	35.0	34.0	12.0	12.0	22.0	13.0	1.0	Baik	Baik
21	43.0	51.0	13.0	14.0	28.0	15.0	0.971735	Sedang	Sedang
22	49.0	63.0	14.0	14.0	28.0	15.0	0.971735	Sedang	Sedang
23	72.0	108.0	14.0	43.0	44.0	20.0	0.980059	Tidak-Sehat	Tidak...
24	33.0	48.0	12.0	14.0	23.0	10.0	1.0	Baik	Baik
25	56.0	82.0	13.0	29.0	62.0	15.0	0.971735	Sedang	Sedang
26	52.0	70.0	14.0	19.0	28.0	13.0	0.971735	Sedang	Sedang
27	20.0	20.0	13.0	10.0	19.0	8.0	1.0	Baik	Baik
28	27.0	41.0	14.0	18.0	17.0	11.0	1.0	Baik	Baik
29	30.0	53.0	16.0	17.0	24.0	11.0	0.971735	Sedang	Sedang
30	27.0	37.0	13.0	16.0	24.0	12.0	1.0	Baik	Baik
31	36.0	51.0	13.0	16.0	26.0	10.0	0.971735	Sedang	Sedang
32	38.0	58.0	2.0	11.0	65.0	6.0	0.811254	Sedang	Sedang
33	29.0	46.0		12.0	80.0	12.0	0.966752	Sedang	Sedang
34	37.0	59.0		10.0	86.0	13.0	0.971735	Sedang	Sedang
35	30.0	49.0		9.0	77.0	7.0	0.966752	Sedang	Sedang
36	35.0	52.0		8.0	77.0	7.0	0.811254	Sedang	Sedang
37	48.0	60.0		12.0	85.0	14.0	0.971735	Sedang	Sedang

Gambar 3. Klasifikasi Data

Classifier output

```

J48 pruned tree
-----

pm25 <= 100
| pm25 <= 50
| | o3 <= 50
| | | pm10 <= 49
| | | | so2 <= 44: Baik (140.84/3.45)
| | | | so2 > 44: Sedang (5.13/2.02)
| | | pm10 > 49: Sedang (18.34/0.57)
| | | o3 > 50: Sedang (22.37/0.36)
| | pm25 > 50
| | | no2 <= 5
| | | | no2 <= 4
| | | | pm10 <= 41: Baik (9.73/0.09)
| | | | pm10 > 41: Sedang (5.81/1.77)
| | | no2 > 4: Sedang (13.33/0.74)
| | | no2 > 5: Sedang (1311.36/1.48)
pm25 > 100
| pm10 <= 54: Sedang (10.42/2.28)
| pm10 > 54: Tidak-Sehat (271.68/3.08)

Number of Leaves :    10

Size of the tree :    19

Time taken to build model: 0.25 seconds
    
```

Gambar 4. Hasil klasifikasi

Classifier output

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      1797      99.3367 %
Incorrectly Classified Instances     12         0.6633 %
Kappa statistic                     0.9827
Mean absolute error                  0.0119
Root mean squared error              0.0643
Relative absolute error              4.6133 %
Root relative squared error         17.9563 %
Total Number of Instances          1809

=== Detailed Accuracy By Class ===

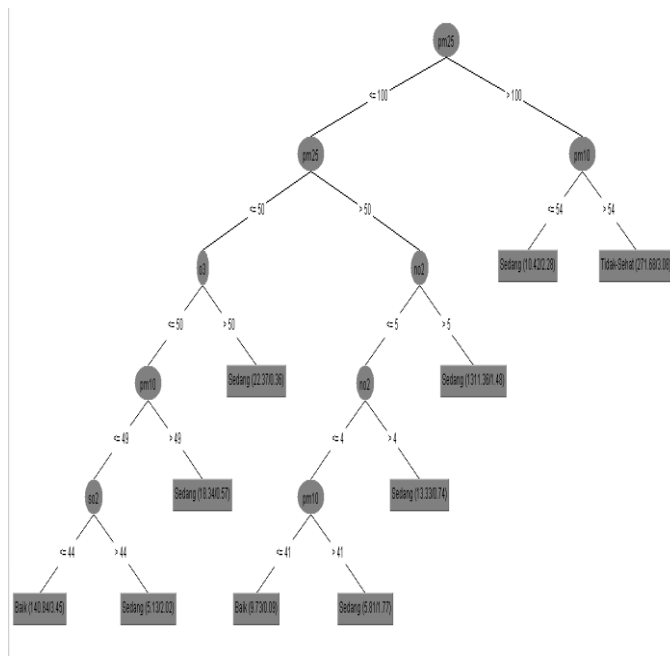
                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0.997    0.019    0.994      0.997    0.996      0.982    0.997    0.998    Sedang
                1.000    0.000    1.000      1.000    1.000      1.000    1.000    1.000    Tidak-Sehat
                0.949    0.002    0.974      0.949    0.961      0.958    0.993    0.976    Baik
Weighted Avg.   0.993    0.015    0.993      0.993    0.993      0.982    0.997    0.996

=== Confusion Matrix ===

  a  b  c  <-- classified as
1380 0  4 | a = Sedang
  0 269 0 | b = Tidak-Sehat
  8  0 148 | c = Baik
    
```

Gambar 5. Output klasifikasi

Hasil klasifikasi didapatkan 10 keputusan atau leaves dengan tingkat akurasi keberhasilan 99,3367% dan tingkat kegagalan 0,6633%.



Gambar 6. Pohon keputusan

Berdasarkan pohon keputusan dari algoritma C.45 untuk melakukan klasifikasi kualitas udara DKI Jakarta pada Gambar 6, diperoleh nilai sebagai berikut:

1. Jika nilai PM25 lebih besar dari 100 dan PM10 lebih besar dari 54 maka masuk kedalam kategori tidak sehat.
2. Jika nilai PM25 lebih besar dari 100 tetapi PM10 kurang dari sama dengan 54 maka masuk kedalam kategori sedang.
3. Jika nilai PM25 kurang dari sama dengan 100 dan PM25 lebih dari 50 tetapi NO<sub>2</sub> lebih dari 5 maka dikategorikan sedang.
4. Jika nilai PM25 kurang dari sama dengan 100 dan PM25 lebih dari 50 tetapi NO<sub>2</sub> kurang dari sama dengan 5 dan NO<sub>2</sub> lebih dari 4 maka dikategorikan sedang.
5. Jika nilai PM25 kurang dari sama dengan 100 dan PM25 lebih dari 50 tetapi NO<sub>2</sub> kurang dari sama dengan 5 dan NO<sub>2</sub> kurang dari 4 dan PM10 lebih dari 41 maka dikategorikan sedang.
6. Jika nilai PM25 kurang dari sama dengan 100 dan PM25 lebih dari 50 tetapi NO<sub>2</sub> kurang dari sama dengan 5 dan NO<sub>2</sub> kurang dari sama dengan dari 4 dan PM10 kurang dari sama dengan 41 maka dikategorikan baik.
7. Jika nilai PM25 kurang dari sama dengan 100 dan PM25 kurang dari sama dengan 50 tetapi O<sub>3</sub> lebih dari 50 maka dikategorikan sedang.
8. Jika nilai PM25 kurang dari sama dengan 100 dan PM25 kurang dari sama dengan 50 tetapi O<sub>3</sub> kurang dari sama dengan 50 dan PM10 lebih dari 49 maka dikategorikan sedang.
9. Jika nilai PM25 kurang dari sama dengan 100 dan PM25 kurang dari sama dengan 50 tetapi O<sub>3</sub> kurang dari sama dengan 50 dan PM10 kurang dari sama

dengan 49 tetapi so<sub>2</sub> lebih dari 44 maka dikategorikan sedang.

Jika nilai PM25 kurang dari sama dengan 100 dan PM25 kurang dari sama dengan 50 tetapi O<sub>3</sub> kurang dari sama dengan 50 dan PM10 kurang dari sama dengan 49 tetapi SO<sub>2</sub> kurang dari sama dengan 44 maka dikategorikan baik.

## V. KESIMPULAN

### 5.1 Kesimpulan

Berdasarkan analisis data dan pembahasan yang telah kita uji dengan menggunakan metode Algoritma C.45 untuk mengelompokkan atau mengklasifikasi data kualitas udara DKI Jakarta berdasarkan data yang diambil dari Website Jakarta Open Data, data tersebut menggunakan 6 parameter pengukuran kualitas udara yaitu PM10, PM25, SO<sub>2</sub>, CO, O<sub>3</sub>, NO<sub>2</sub>, dan kategori. Hasil dari penerapan algoritma C.45 data mining untuk klasifikasi kualitas udara di DKI Jakarta didapatkan bahwa kualitas udara dikategorikan menjadi tiga macam yaitu sedang, baik, dan tidak sehat dengan rasio yang berbeda-beda setiap harinya. Setelah dilakukan klasifikasi maka dapat disimpulkan kualitas udara yang dianalisis pada bulan Januari-Desember 2021 masuk kedalam kategori sedang dimana tingkat kualitas udara masih dapat diterima untuk kesehatan manusia, hewan, dan tumbuhan. Berdasarkan kesimpulan tersebut maka diharapkan masyarakat DKI Jakarta untuk meminimalisir pencemaran udara dengan cara menggunakan transportasi publik, melakukan reboisasi, menggunakan bahan bakar ramah lingkungan, dan lain sebagainya. Dengan begitu maka angka harapan hidup dari masyarakat DKI Jakarta akan semakin meningkat.

### 5.2 Saran

Saran untuk penelitian selanjutnya adalah peneliti dapat menggunakan metode yang lain dan juga dapat menggunakan data di Provinsi atau Kota lainnya di Indonesia.

## DAFTAR PUSTKA

Amalia, A., Zaidiah, A., & Isnainiyah, I. N. (2022). Prediksi kualitas udara menggunakan algoritma k- nearest neighbor. 07(April 2021), 496–507.

Binus (2016) *Tahap-Tahap Data Mining, 15 December 2016*. Available at: <https://sis.binus.ac.id/2016/12/15/tahap-tahap-data-mining/>.

B. Novianti, T. Rismawan, and S. Bahri, “Implementasi Data Mining Dengan Algoritma C4.5 Untuk Penjurusan Siswa (Studi Kasus: Sma Negeri 1 Pontianak),” J. Coding, Sist. Komput. Untan, vol. 04, no. 3, pp. 75–84, 2016.

P. B. N. Setio., et al. 2020. “Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5,” PRISMA, Prosiding Seminar Nasional Matematika, vol.3, pp. 64–71.

Sang, A.I., Sutoyo, E. and Darmawan, I. (2021) ‘Analisis Data Mining Untuk Klasifikasi Data Kualitas Udara DKI Jakarta Menggunakan Algoritma Decision Tree Dan Support Vector Machine’, *e-Proceeding of Engineering*, 8(5), pp. 8954–8963.

- Syihabuddin Azmil Umri, S. (2021) “Analisis Dan Komparasi Algoritma Klasifikasi Dalam Indeks Pencemaran Udara Di Dki Jakarta,” *JIKO (Jurnal Informatika dan Komputer)*, 4(2), hal. 98–104. doi:10.33387/jiko.v4i2.2871.
- Wiryono, Singgih. (2021) *Data Pemprov DKI, Kendaraan Bermotor Jadi Masalah Utama Pencemaran Udara Jakarta*, *Kompas.com*. Available at: <https://megapolitan.kompas.com/read/2021/11/11/14321471/data-pemprov-dki-kendaraan-bermotor-jadi-masalah-utama-pencemaran-udara?page=all>.



Tri Astuti, S.Kom., M.Eng. Lahir di Banyumas Provinsi Jawa Tengah. Pendidikan yang ditempuh dari SD Negeri Gumelar Kidul, SMP Negeri 1 Sumpiuh, dan SMA Negeri 1 Purwokerto. Selanjutnya penulis melanjutkan Studi S-1 Teknik Informatika di Universitas Amikom Purwokerto dan S-2 Teknik Elektro Universitas Gadjah Mada. Minat Penelitian Data Mining, Machine Learning.