

# **KLASIFIKASI INDEKS PENCEMARAN KUALITAS UDARA DI DKI JAKARTA TAHUN 2021 MENGGUNAKAN ALGORITMA RANDOM FOREST**

## ***AIR QUALITY POLLUTION INDEX CLASSIFICATION IN DKI JAKARTA IN 2021 USING RANDOM FOREST ALGORITHM***

**Rainal Zulian Oktavianto**

Program Studi S1 Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa, Cikarang.

### **Abstrak**

Pencemaran udara merupakan suatu masalah yang berdampak buruk bagi kehidupan adalah aktivitas transportasi, industri, jasa, kegiatan lainnya yang meningkat, dan buangan sisa-sisa kegiatan ke udara. Proses metabolisme yang terjadi dalam tubuh makhluk hidup nitrogen oksida, ozon dan zat lain sebagainya. Menurut Air Quality Live Indeks (AQLI), pada bulan April 2021, DKI Jakarta merupakan kota ke-6 (enam) dengan kualitas udara terburuk. Hal ini ditandai dengan nilai indeks AQI Jakarta memiliki nilai 156 dengan kategori tidak sehat. Indeks Standar Pencemaran Udara (ISPU) digunakan sebagai parameter untuk mengukur kualitas udara. 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 Tahun 2021 untuk memantau kualitas udara yang ada di DKI Jakarta. Berdasarkan penjelasan di atas, pada penelitian ini akan melakukan klasifikasi kualitas udara di DKI Jakarta dengan algoritma Random Forest. Penelitian ini diharapkan dapat menghasilkan informasi bagaimana kualitas udara di DKI Jakarta.

**Kata Kunci :** Pencemaran, ISPU, Udara, Klasifikasi

### *Abstract*

*Air pollution is a problem that has adverse effects on life due to increased transportation, industrial, service, and other activities, as well as emissions from various activities into the air. Metabolic processes occurring in living organisms produce nitrogen oxides, ozone, and other substances. According to the Air Quality Live Index (AQLI), in April 2021, DKI Jakarta ranked 6th with the worst air quality. This is indicated by Jakarta's AQI index value of 156, categorized as unhealthy. The Air Pollution Standard Index (ISPU) is used as a parameter to measure air quality. Data classification analysis of air quality in DKI Jakarta is conducted using data obtained from the Jakarta Open Data website, provided by the Jakarta Environmental Agency in 2021 to monitor air quality in DKI Jakarta. Based on the above explanation, this research will classify air quality in DKI Jakarta using the Random Forest algorithm. This study is expected to provide information on the air quality in DKI Jakarta.*

**Keywords :** Pollution, ISPU, Air, Classification

## 1. Pendahuluan

Pencemaran udara merupakan suatu masalah yang berdampak buruk bagi kehidupan adalah aktivitas transportasi, industri, jasa, kegiatan lainnya yang meningkat, dan buangan sisa-sisa kegiatan ke udara[1]. Udara yang tercemar dapat memberikan dampak negatif bagi kesehatan bahkan dapat menyebabkan kematian. Udara mengambil peran penting dalam menjaga kehidupan makhluk hidup di bumi.

Proses metabolisme yang terjadi dalam tubuh makhluk hidup nitrogen oksida, ozon dan zat lain sebagainya. Beberapa zat - zat tersebut masih dapat ditoleransi oleh tubuh dalam konsentrasi yang masih di bawah batas wajar, namun jika melampaui batas wajar dapat menyebabkan masalah serius pada kesehatan manusia seperti penyakit jantung, infeksi pernapasan, stroke, asma hingga penyakit paru. Salah satu faktor penyebab peningkatan konsentrasi zat - zat tersebut yang ada di dalam udara adalah aktivitas manusia[2][3][4].

Menurut Air Quality Live Indeks (AQLI), pada bulan April 2021, DKI Jakarta merupakan kota ke-6 (enam) dengan kualitas udara terburuk. Hal ini ditandai dengan nilai indeks AQI Jakarta memiliki nilai 156 dengan kategori tidak sehat. Polutan utama yang menyebabkan penurunan kualitas udara yaitu PM2.5, di mana polutan ini jumlahnya tidak boleh melebihi 10 mikron saat berada di udara. Di DKI Jakarta, polutan ini tercatat mencapai 57 mikron per kubik yang menandakan bahwa kualitas udara di DKI Jakarta sangat buruk[5].

Indeks Standar Pencemaran Udara (ISPU) digunakan sebagai parameter untuk mengukur kualitas udara. Pada pengukuran ISPU menetapkan lima parameter pencemaran udara yang digunakan untuk pengamatan, yaitu Karbon Monoksida (CO), Sulfur Dioksida (SO<sub>2</sub>), Nitrogen Dioksida (NO<sub>2</sub>), Ozon Permukaan (O<sub>3</sub>), Partikel Debu (PM<sub>10</sub>)[6]. Berdasarkan pemantauan kualitas udara yang dilakukan oleh Dinas Lingkungan Hidup Provinsi DKI Jakarta laporan dari hasil pemantauan kualitas udara yang memberikan keterangan mengenai kualitas udara yang bersih ataupun tercemar[7].

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 Tahun 2021 untuk memantau kualitas udara yang ada di DKI Jakarta[8]. Dataset ini berisi mengenai Indeks Standar Pencemaran Udara (ISPU) yang diukur dari 5 stasiun pemantau kualitas udara (SPKU) yang ada di Provinsi DKI Jakarta Tahun 2021 yang terletak, di Bundaran Hotel Indonesia (HI), Kelapa Gading, Jagakarsa, Lubang Buaya dan Kebon Jeruk[9].

Berdasarkan penjelasan di atas, pada penelitian ini akan melakukan klasifikasi kualitas udara di DKI Jakarta dengan algoritma Random Forest. Penelitian ini diharapkan dapat menghasilkan informasi bagaimana kualitas udara di DKI Jakarta[10]. 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.

## 2. Metodologi

### 2.1 Skema Alur Penelitian

Penelitian dalam mengklasifikasikan level ISPU yang dilakukan untuk menentukan kualitas udara di DKI Jakarta Tahun 2021 menggunakan skema alur penelitian sebagai berikut :



**Gambar 1.** Skema Alur Penelitian

## 2.2 Objek Penelitian

Obyek yang diteliti pada penelitian ini adalah data pencemaran udara dari Indeks Pencemar Standar Udara di DKI Jakarta Tahun 2021. Data ini diambil dari Website Jakarta Open Data yang dilakukan oleh Dinas Lingkungan Hidup DKI Jakarta untuk memantau kualitas udara yang ada di DKI Jakarta.

Dari data tersebut diteliti untuk membuat perbandingan antar algoritma klasifikasi untuk mencari algoritma terbaik dalam memprediksi kualitas udara yang baik, sedang, tidak sehat, sangat tidak sehat dan berbahaya.

## 2.3 Studi Literatur

Tahap awal dalam merumuskan penelitian ini dilakukan dengan cara mengumpulkan informasi yang menunjang penelitian yakni berkaitan dengan kualitas udara, ISPU dan berbagai algoritme klasifikasi yang diperoleh dari artikel, jurnal, dan dokumen lainnya.

## 2.4 Data Selection

Dataset Indeks Pencemar Standar Udara DKI Jakarta dikumpulkan dari laman situs terbuka pemerintah DKI Jakarta <http://www.data.jakarta.go.id/> berupa CSV yang diambil per SPKU dan per daerah yakni Jakarta Pusat (DKI1), Jakarta Utara (DKI2), Jakarta Selatan (DKI3), Jakarta Timur (DKI4), dan Jakarta Barat (DKI5) dari bulan Januari 2021 hingga bulan Desember 2021.

## 2.5 Data Preparation

Data yang sudah dikumpulkan dan diseleksi akan memasuki tahap data preparation. Pada tahap ini data akan di reduksi dengan menghapus data yang memiliki nilai kosong, data duplikat, data yang terisi namun tidak memiliki data secara lengkap, dan penghapusan atribut yang tidak digunakan. Kemudian karena klasifikasi masuk dalam supervised learning maka dibutuhkan tahap labeling pada atribut yang akan diklasifikasikan.

## 2.6 Data Transformation

Pada tahap ini dari dataset yang masih terpisah karena laporan ISPU atau frekuensi penerbitannya selama satu bulan sekali, jadi antar berkas CSV dari bulan januari 2021 hingga bulan desember 2021 digabungkan menjadi satu berkas CSV, sehingga menjadi kesatuan dataset untuk mempermudah dalam proses berikutnya.

## 3. Hasil dan Pembahasan

### 3.1 Persiapan Data

Data ISPU pada penelitian ini disusun oleh Dinas Lingkungan Hidup Provinsi DKI Jakarta. 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 seperti pada gambar berikut :

	tanggal	stasiun	pm10	pm25	so2	co	o3	no2	max	critical	kategori
0	2021-01-01	DKI1 (Bunderan HI)	38	53	29	6	31	13	53	PM25	SEDANG
1	2021-01-02	DKI1 (Bunderan HI)	27	46	27	7	47	7	47	O3	BAIK
2	2021-01-03	DKI1 (Bunderan HI)	44	58	25	7	40	13	58	PM25	SEDANG
3	2021-01-04	DKI1 (Bunderan HI)	30	48	24	4	32	7	48	PM25	BAIK
4	2021-01-05	DKI1 (Bunderan HI)	38	53	24	6	31	9	53	PM25	SEDANG
...	...	...	...	...	...	...	...	...	...	...	...
1820	2021-12-27	DKI5 (Kebon Jeruk) Jakarta Barat	54	76	36	14	21	47	76	PM25	SEDANG
1821	2021-12-28	DKI5 (Kebon Jeruk) Jakarta Barat	44	68	20	11	21	33	68	PM25	SEDANG
1822	2021-12-29	DKI5 (Kebon Jeruk) Jakarta Barat	34	54	28	8	25	29	54	PM25	SEDANG
1823	2021-12-30	DKI5 (Kebon Jeruk) Jakarta Barat	53	75	25	15	23	44	75	PM25	SEDANG
1824	2021-12-31	DKI5 (Kebon Jeruk) Jakarta Barat	60	87	28	19	30	53	87	PM25	SEDANG

Gambar 2. Atribut Dataset

```
df.kategori.value_counts() # Cek value kolom target

SEDANG          1349
TIDAK SEHAT      272
BAIK             187
TIDAK ADA DATA   16
Name: kategori, dtype: int64
```

Gambar 3. Nilai kategori

	pm10	pm25	so2	co	o3	no2	max
count	1517.000000	1517.000000	1517.000000	1517.000000	1517.000000	1517.000000	1517.000000
mean	52.874753	78.238629	35.720501	11.914305	31.338167	20.423204	79.238629
std	14.705020	23.178350	12.627751	4.779081	14.843322	9.458014	22.310143
min	15.000000	13.000000	2.000000	2.000000	8.000000	3.000000	20.000000
25%	45.000000	63.000000	26.000000	9.000000	21.000000	14.000000	64.000000
50%	54.000000	78.000000	36.000000	11.000000	28.000000	19.000000	79.000000
75%	62.000000	92.000000	45.000000	14.000000	38.000000	26.000000	93.000000
max	179.000000	174.000000	82.000000	43.000000	151.000000	65.000000	179.000000

**Gambar 4.**

### 3.2 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 : Tanggal pengukuran kualitas udara
- stasiun : Lokasi pengukuran di stasiun
- pm10 : Partikulat salah satu parameter yang diukur
- pm25 : Partikulat salah satu parameter yang diukur
- so2 : Sulfida (dalam bentuk SO<sub>2</sub>) salah satu parameter yang diukur
- co : Carbon Monoksida salah satu parameter yang diukur
- o3 : Ozon salah satu parameter yang diukur
- no2 : Nitrogen dioksida salah satu parameter yang diukur
- max : Nilai ukur paling tinggi dari seluruh parameter yang diukur dalam waktu yang sama
- critical : Parameter yang hasil pengukurannya paling tinggi
- kategori : Kategori hasil perhitungan indeks standar pencemaran udara

tanggal	stasiun	pm10	pm25	so2	co	o3	no2	max	critical	kategori	kategori_encoded
2021-01-01	DKI1 (Bunderan HI)	38	53	29	6	31	13	53	PM25	SEDANG	0.0
2021-01-02	DKI1 (Bunderan HI)	27	46	27	7	47	7	47	O3	BAIK	2.0
2021-01-03	DKI1 (Bunderan HI)	44	58	25	7	40	13	58	PM25	SEDANG	0.0
2021-01-04	DKI1 (Bunderan HI)	30	48	24	4	32	7	48	PM25	BAIK	2.0
2021-01-05	DKI1 (Bunderan HI)	38	53	24	6	31	9	53	PM25	SEDANG	0.0
2021-01-06	DKI1 (Bunderan HI)	41	58	23	13	46	13	58	PM25	SEDANG	0.0
2021-01-07	DKI1 (Bunderan HI)	35	47	22	6	39	10	47	PM25	BAIK	2.0
2021-01-08	DKI1 (Bunderan HI)	37	54	26	16	17	10	54	PM25	SEDANG	0.0
2021-01-09	DKI1 (Bunderan HI)	47	61	16	27	22	12	61	PM25	SEDANG	0.0
2021-01-10	DKI1 (Bunderan HI)	23	25	16	11	33	8	33	O3	BAIK	2.0
2021-01-11	DKI1 (Bunderan HI)	38	54	17	14	27	10	54	PM25	SEDANG	0.0
2021-01-12	DKI1 (Bunderan HI)	29	50	20	12	26	15	50	PM25	BAIK	2.0
2021-01-13	DKI1 (Bunderan HI)	34	44	17	13	20	9	44	PM25	BAIK	2.0
2021-01-14	DKI1 (Bunderan HI)	36	57	15	15	26	13	57	PM25	SEDANG	0.0
2021-01-15	DKI1 (Bunderan HI)	42	62	14	15	32	12	62	PM25	SEDANG	0.0
2021-01-16	DKI1 (Bunderan HI)	52	73	15	17	25	14	73	PM25	SEDANG	0.0
2021-01-17	DKI1 (Bunderan HI)	52	71	17	15	26	14	71	PM25	SEDANG	0.0
2021-01-18	DKI1 (Bunderan HI)	40	52	16	11	26	12	52	PM25	SEDANG	0.0
2021-01-19	DKI1 (Bunderan HI)	31	41	13	12	25	12	41	PM25	BAIK	2.0
2021-01-20	DKI1 (Bunderan HI)	35	34	12	12	22	13	35	PM10	BAIK	2.0

only showing top 20 rows

**Gambar 5.** Nilai null/nan yang sudah diubah

### 3.3 Proses Klasifikasi

proses klasifikasi dengan menggunakan algoritma Random Forest dan Data yang diolah sebanyak 1824 baris data, didapatkan setelah melakukan pengumpulan dan pembersihan data menggunakan Microsoft Excel. Selanjutnya untuk proses klasifikasi menggunakan tools RapidMiner. Proses pertama yaitu input

dataset yang digunakan kemudian dihubungkan dengan set role untuk penandaan kolom sebagai label. Kemudian split data yang bertujuan untuk membagi menjadi 2 data yaitu data uji dan data latih.

Peneliti menggunakan beberapa rasio pembagian data yang dilakukan pengujian terhadap model. Rasio pembagian data yang digunakan adalah rasio 60:40, 70:30, 80:20, dan 90:10. Rasio tersebut akan diuji satu persatu terhadap model kemudian dibandingkan dengan satu sama lain dengan hasil evaluasi terbaik dan dihubungkan ke algoritma yang digunakan yaitu algoritma Random Forest.

#### 4. 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 Random Forest 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. Daftar Pustaka

- [1] "303-Article Text-1448-1527-10-20201118".
- [2] A. Efendi, I. Iskandar, R. Kurniawan, dan M. Affandes, "KLIK: Kajian Ilmiah Informatika dan Komputer Klasifikasi Kebakaran Hutan Riau Menggunakan Random Forest dan Visualisasi Citra Sentinel-2," *Media Online*, vol. 4, no. 3, hlm. 1602–1612, 2023, doi: 10.30865/klik.v4i3.1521.
- [3] A. A. Nababan, M. Jannah, M. Aulina, dan D. Andrian, "PREDIKSI KUALITAS UDARA MENGGUNAKAN XGBOOST DENGAN SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE (SMOTE) BERDASARKAN INDEKS STANDAR PENCEMARAN UDARA (ISPU)," *Jurnal Teknik Informatika Kaputama (JTik)*, vol. 7, no. 1, 2023.
- [4] S. S. A. Umri, M. S. Firdaus, dan A. Primajaya, "ANALISIS DAN KOMPARASI ALGORITMA KLASIFIKASI DALAM INDEKS PENCEMARAN UDARA DI DKI JAKARTA," *Jurnal Informatika dan Komputer) Akreditasi KEMENRISTEKDIKTI*, vol. 4, no. 2, 2021, doi: 10.33387/jiko.
- [5] A. I. Sang, E. Sutoyo, dan I. Darmawan, "ANALISIS DATA MINING UNTUK KLASIFIKASI DATA KUALITAS UDARA DKI JAKARTA MENGGUNAKAN ALGORITMA DECISION TREE DAN SUPPORT VECTOR MACHINE DATA MINNING ANALYSIS FOR CLASSIFICATION OF AIR QUALITY DATA DKI JAKARTA USING DECISION TREE ALGORTHM AND SUPPORT VECTOR MACHINER ALGORITHM."
- [6] A. Toha, P. Purwono, dan W. Gata, "Model Prediksi Kualitas Udara dengan Support Vector Machines dengan Optimasi Hyperparameter GridSearch CV," *Buletin Ilmiah Sarjana Teknik Elektro*, vol. 4, no. 1, hlm. 12–21, Mei 2022, doi: 10.12928/biste.v4i1.6079.

- [7] H. Rachmi, “Klasifikasi Pencemaran Udara Di DKI Jakarta Menggunakan Metode Naïve Bayes,” *Jurnal Publikasi Ilmu Komputer dan Multimedia (JUPIKOM)*, vol. 2, no. 2, 2023, doi: 10.55606/jupikom.v2i2.18578.
- [8] A. Wicahyo, A. Pudoli, dan D. Kusumaningsih, “Penggunaan Algoritma Naive Bayes dalam klasifikasi Pengaruh Pencemaran Udara,” *Jurnal ICT : Information Communication & Technology*, vol. 20, no. 1, hlm. 103–108, 2021.
- [9] “VOLUME : 6 NO : 1 JURNAL SISTEM & TEKNOLOGI INFORMASI KOMUNIKASI 36.”
- [10] A. Amalia *dkk.*, “PREDIKSI KUALITAS UDARA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR.” [Daring]. Tersedia pada: <https://data.jakarta.go.id/>.