

KLASIFIKASI KUALITAS UDARA PADA PROVINSI DKI JAKARTA MENGGUNAKAN ALGORITMA RANDOM FOREST

1st Farhan Hasan Al amudi
Fakultas Rekayasa Industri
Universitas Telkom
Bandung, Indonesia
farhanhasan@student.telkomuniversit
y.ac.id

2nd Faqih Hamami S.Kom., M.T.
Fakultas Rekayasa Industri
Universitas Telkom
Bandung, Indonesia
faqihhamami@telkomuniversity.ac.id

3rd Ahmad Almaarif, S.T., M.T.
Fakultas Rekayasa Industri
Universitas Telkom
Bandung, Indonesia
ahmadalmaarif@telkomuniversity.ac.id

Abstrak : Udara merupakan salah satu unsur penting bagi lingkungan dan menjadi kebutuhan utama untuk menopang kehidupan makhluk hidup. Proses metabolisme pada makhluk hidup tidak mungkin berlangsung tanpa oksigen yang diambil dari udara. Selain oksigen, udara juga mengandung zat lain seperti karbon monoksida, karbon dioksida, formaldehida, jamur, virus, dan sebagainya. Zat-zat tersebut masih dapat dinetralkan selama berada dalam batas aman, namun ketika melebihi ambang batas, proses netralisasi akan terganggu. Peningkatan kandungan zat-zat tersebut di udara umumnya diakibatkan oleh aktivitas manusia. Di tahun 2019 Indonesia mencapai titik pencemaran udara terburuk yang sudah mencapai titik merah yang menandakan tidak sehatnya udara yang ada pada DKI Jakarta. Salah satu cara untuk memantau informasi mengenai kualitas udara adalah dengan menggunakan metode klasifikasi. Pada penelitian ini klasifikasi dilakukan menggunakan dataset ISPU pencemaran udara Provinsi DKI Jakarta dari tahun 2019 sampai tahun 2022. Klasifikasi yang tepat dapat sangat membantu pemerintah dalam merumuskan kebijakan. Kebijakan ini bertujuan untuk mengendalikan polusi agar sesuai dengan standar kualitas udara yang bermanfaat bagi kelangsungan hidup makhluk hidup. Dalam penelitian ini, model klasifikasi yang digunakan adalah random forest, dengan lima atribut yaitu PM10, SO₂, NO, O₃, dan CO₂, serta kategori sebagai target label.

Kata kunci— Udara, DKI Jakarta, ISPU, Klasifikasi, Random Forest

I. PENDAHULUAN

Udara merupakan komponen lingkungan yang sangat penting dan menjadi kebutuhan utama untuk kehidupan. Metabolisme dalam tubuh makhluk tidak bisa berlangsung tanpa hadirnya oksigen dari udara. Selain oksigen, udara juga mengandung zat lain seperti karbon monoksida, karbon dioksida, formaldehida, jamur, virus, dan lain-lain. Selama zat-zat ini masih dalam batas aman, mereka bisa dinetralkan, tetapi jika melebihi ambang batas, proses penetralan bisa terganggu. Peningkatan kadar zat-zat tersebut sering kali disebabkan akibat aktivitas manusia (Wulandari, R. A., et al., 2008).

Udara normal mengandung gas-gas dengan komposisi 78% nitrogen, 20% oksigen, 0,93% argon, 0,03% karbon dioksida, serta sejumlah kecil gas lainnya seperti neon, helium, metana, dan hidrogen. Komposisi ini mendukung keberlangsungan hidup manusia, tetapi aktivitas manusia yang tidak lingkungan seringkali menyebabkan penurunan kualitas udara, baik dari segi sifat fisik maupun kimiawi. Perubahan tersebut dikenal sebagai pencemaran udara.

Kualitas udara yang buruk di Indonesia, terutama di provinsi DKI Jakarta, disebabkan oleh polusi udara. Beberapa faktor yang mempengaruhi pencemaran udara di wilayah ini meliputi kendaraan bermotor, industri, pembakaran batubara, debu dari jalan, dan debu dari aktivitas konstruksi. Namun, penyumbang terbesar adalah meningkatnya jumlah kendaraan, yang berkontribusi sekitar 31 hingga 40% terhadap pencemaran udara di provinsi tersebut.

Melihat permasalahan yang ada, penelitian ini akan berfokus pada identifikasi dan pemantauan kualitas udara dengan menggunakan data pencemaran udara di Provinsi DKI Jakarta. Temuan dari penelitian ini diharapkan dapat membantu pemerintah untuk merumuskan kebijakan dan praktik pengelolaan yang lebih efektif untuk menangani isu kualitas udara di daerah ini.

Dari permasalahan yang ada peneliti akan melakukan identifikasi untuk memantau kualitas udara yang terdapat pada data pencemaran kualitas udara Provinsi DKI Jakarta yang akan menjadi acuan kepada pemerintah untuk membuat kebijakan dan pengelolaan yang baik terhadap kualitas udara di Provinsi DKI Jakarta.

Algoritma data mining telah diterapkan untuk klasifikasi polusi udara, alasan untuk menerapkan data mining adalah untuk memanfaatkan informasi tersembunyi dalam database yang memiliki volume data besar, menggunakan analisis statistik dan model matematika (Meilani & Asadulloh, 2015). Klasifikasi merupakan salah satu metode dalam pengolahan atau komputasi data untuk mengelompokkan data sesuai dengan ciri-ciri yang sama (Aji dll., 2018).

Dalam penelitian ini, algoritma random forest dipilih karena merupakan algoritma machine learning yang fleksibel dan mudah diterapkan, bahkan tanpa perlu melakukan penyetelan hyper-parameter, tapi sering kali mendapat hasil yang bagus. Random forest merupakan algoritma yang paling populer digunakan karena kemudahannya dan variasinya. Ini juga yang menjadi salah satu keunggulan besar dari algoritma random forest adalah kemampuannya untuk diterapkan pada masalah klasifikasi dan regresi, yang merupakan bagian penting dari banyak sistem machine learning saat ini.

II. KAJIAN TEORI

A. Udara

Udara adalah atmosfer yang melindungi bumi dan berfungsi untuk melindungi planet ini dari gangguan luar. Namun, udara di lingkungan tidak sepenuhnya bersih karena adanya polusi yang dihasilkan oleh aktivitas manusia. Kualitas udara menjadi parameter penting untuk menilai kondisi udara yang layak di suatu daerah. Penurunan kualitas udara disebabkan dari polusi berbagai jenis gas, asap dari kendaraan, emisi industri, serta limbah udara yang dihasilkan oleh rumah tangga (Prayudha et al., 2018).

Udara di bumi mengandung berbagai jenis zat, salah satunya adalah oksigen (O₂) yang diperlukan untuk pernapasan manusia dan hewan. Selain itu, karbondioksida (CO₂) berperan dalam proses fotosintesis pada tumbuhan, sementara O₃ berfungsi menyerap radiasi sinar ultraviolet dari pancaran matahari.

Udara yang terdapat bebas di permukaan bumi pada lapisan troposfer. Udara ini berperan penting dalam bidang kesehatan. Sebagai salah satu komponen lingkungan yang esensial, kualitas udara perlu dipelihara dan ditingkatkan agar dapat mendukung kehidupan makhluk hidup dengan optimal. Pencemaran udara biasanya diartikan sebagai situasi di mana udara mengandung satu atau lebih zat kimia dalam konsentrasi tinggi yang dapat menimbulkan gangguan atau risiko bagi manusia, hewan, tumbuhan, serta aset lainnya (Zakaria & Azizah, 2013).

B. Parameter Pencemaran Udara

Parameter yang digunakan untuk menentukan nilai indeks standar pencemaran udara (ISPU) dapat ditemukan pada tabel 1.

Tabel 1

(Lampiran keputusan kepala bapedal No. 107 tahun 1997)

No.	Parameter	Waktu Pengukuran (rata-rata)
1	Partikulat (PM ₁₀)	24 jam
2	Sulfur dioksida (SO ₂)	24 jam
3	Karbon monoksida	8 jam
4	Ozon (O ₃)	1 jam
5	Nitrogen dioksida (NO ₂)	1 jam

Jenis-jenis parameter yang terkait dengan pencemaran udara meliputi:

- Nitrogen dioksida (NO₂): adalah gas yang beracun bagi manusia. Secara umum, NO₂ dapat menyebabkan gangguan sistem pernapasan. NO₂ bisa masuk ke dalam paru-paru bereaksi dengan akan membentuk senyawa asam nitrit (HN02) dan asam nitrat (HN03), ini dapat merusak mukosa tubuh manusia.
- Sulfur dioksida (SO₂): Sulfur dioksida (SO₂) yang terdapat dalam udara dapat mengiritasi saluran pernapasan, membunuh jaringan daun, dan dapat menyebabkan hujan asam.
- Carbon Monoksida (CO): Keracunan akibat CO ditandai oleh gejala ringan seperti sakit kepala, dan mual. Pada tingkat keracunan lebih parah, dapat menyebabkan masalah serius.
- Partikulat Matter (PM₁₀): Partikulat Matter (PM₁₀) juga sering disebut sebagai partikel debu. Menurut ahli lingkungan dan kesehatan masyarakat, infeksi saluran pernapasan dapat disebabkan di saluran pernapasan, khususnya di bronkus dan alveoli.
- Ozon(O₃): O₃ Saat ini menjadi isu penting terkait global dari pencemaran udara, yaitu menipisnya lapisan ozon pada atmosfer bumi.

C. Indeks Kualias Udara

Nilai indeks kualitas udara dihitung untuk setiap parameter yang terukur, sehingga diperoleh nilai indeks kualiras udara untuk masing-masing parameter kualitas udara.

Tabel 2

Angka dan kategori indek kualitas udara

Kategori	Rentang angka
Baik	1-50
Sedanng	51-100
Tidak sehat	101-200
Sangat tidak sehat	201-300
Berbahaya	>301

D. Data Mining

Data mining merupakan proses penggalian informasi penting dari kumpulan data yang besar dengan menggunakan teknik analisis statistik, algoritma pembelajaran mesin, dan kecerdasan buatan untuk mengidentifikasi pola, tren, dan hubungan yang tersembunyi dalam data (Hand & Adams, 2015).

Data mining memiliki beberapa teknik:

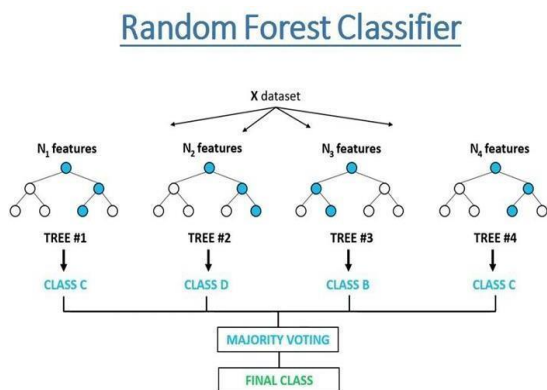
- a) Deskripsi: Teknik Deskripsi digunakan untuk menemukan pola maupun tren tersembunyi dalam data serta mendeskripsikannya.
- b) Estimasi: Teknik estimasi mirip dengan klasifikasi, tetapi variabel targetnya berbentuk numerik daripada pengelompokan kategori.
- c) Klasifikasi: Teknik ini untuk mengelompokkan data ke dalam kategori berdasarkan variabel tujuan.
- d) Pengelompokan: Teknik ini berfokus pada pengelompokan rekaman, observasi, atau kasus berdasarkan kesamaan dalam suatu kelas.
- e) Asosiasi: Teknik ini bertujuan untuk mengidentifikasi hubungan antara berbagai peristiwa yang terjadi bersamaan.

E. Klasifikasi

Klasifikasi adalah suatu proses pengolahan data dengan cara pengelompokan data berdasarkan karakteristik objek yang akan diklasifikasikan, dengan tujuan menemukan fungsinya yang menggambarkan suatu kelas data. (Aji et al., 2018).

F. Random Forest

Random forest adalah sebuah metode pembelajaran mesin yang menggunakan sejumlah pohon keputusan untuk menyelesaikan masalah klasifikasi maupun regresi. Metode ini bekerja dengan membangun banyak pohon dari subset data yang berbeda secara acak, dan setiap pohon memberikan prediksi. Hasil akhir kemudian ditentukan berdasarkan mayoritas prediksi yang dihasilkan oleh setiap pohon. Pendekatan ini dapat meningkatkan akurasi karena mengurangi risiko overfitting dan memanfaatkan kekuatan dari keputusan kolektif berbagai pohon (Nugraha et al., 2018).

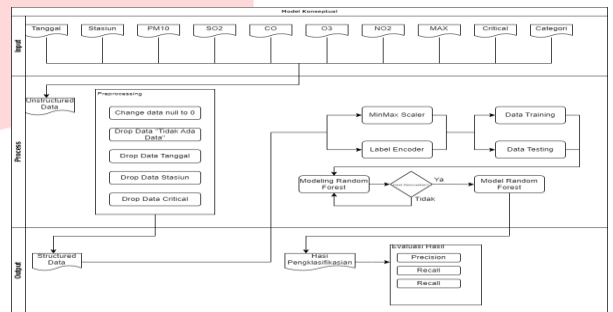


Gambar 1
(Klasifikasi *random Forest*)

III. METODE

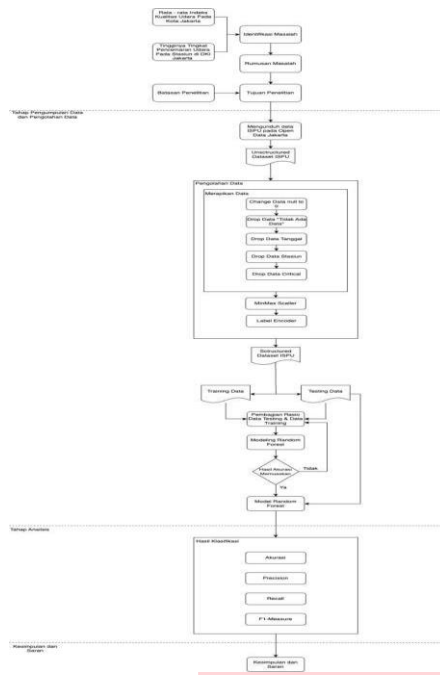
A. Model Konseptual

Model konseptual adalah tahap memastikan peneliti membangun model sesuai dengan keperluan, pengetahuan sebelumnya serta pengalaman. Model ini bukan representasi struktur dasar sistem melakukan model yang membantu dalam membaca penggunaan sistem secara efektif.



Gambar 2
(Model konseptual)

B. Sistematika penyelesaian masalah



Gambar 3
(Sistematika penyelesaian masalah)

a) Identifikasi Masalah

Penulis melakukan pencarian studi kasus dan mengangkat tentang kualitas udara DKI Jakarta. Pada tahap ini dilakukan studi literatur, lalu menentukan pemecahan masalah yang dibahas menentukan tujuan penelitian, menentukan batasan masalah dan diakhiri dengan penulis.

b) Pengolahan Data

Selanjutnya penulis melakukan pencarian dataset serta mengumpulkan melalui website Jakarta open data. penulis mengumpulkan data dan melakukan preprocessing terhadap dataset tersebut, data dibersihkan dari nilai null, variabel yang tidak sesuai dan menyeleksi variabel-variabel yang akan digunakan, penulis melakukan splitting data menjadi data latih data uji. Setelah itu data yang sudah terolah tersebut akan diimplementasikan dengan memakai model algoritma *random forest*.

c) Hasil dan evaluasi

Pada tahap analisis, akan dilakukan pengukuran akurasi terhadap data yang sudah dikumpulkan dan diproses menggunakan algoritma *random forest* yang telah dilakukan pada tahap pengolahan data.

d) Pengumpulan Data

Data dikumpulkan dari situs resmi Jakarta Open Data Pemerintah Provinsi DKI Jakarta serta Dinas Lingkungan Hidup DKI Jakarta, yang menyediakan informasi mengenai Indeks Standar Kualitas Udara. Data yang dipakai sebagai data training memiliki atribut "pm10", "So2", "Co", "o3", dan "No2", dan target label empat yaitu "BAIK", "SEDANG", "TIDAK SEHAT", dan "SANGAT TIDAK SEHAT".

e) Metode evaluasi

Kesimpulan diambil setelah seluruh tahapan sebelumnya telah diselesaikan, hasil akhir ditentukan berdasarkan analisis metode yang telah diterapkan, kemudian evaluasi dan validasi diketahui setelah semua proses penelitian melalui pengujian dan percobaan selesai

IV. HASIL DAN PEMBAHASAN

A. Implementasi Algoritma

a) Proses klasifikasi meliputi proses dari pembagian data training, data testing, dan modeling. Pada tahap ini dibagi proses data training dan testing dengan penentuan pengujian sebagai berikut:

Tabel 3
(Pengujian pada proses data latih dan uji)

Data latih	Data uji	Random state	N_estimator	akurasi
60%	40	20	5	91%
70%	30	20	5	91%
80%	20	20	5	90%
90%	10	20	5	92%

Didapatkan, hasil dari pengujian ratio data pada 90:10, maka data yang digunakan sebanyak 90% data digunakan untuk data latih, sementara 10% sisanya untuk data uji..

Setelah itu dilakukan proses pengujian terhadap random state dengan tujuan mencari akurasi yang lebih baik.

b) *Random state testing*

Pada tahap ini dilakukan pengujian terhadap random state

Tabel 4
(Pengujian random state)

Data Training	Data testing	Random state	N_estimator	Akurasi
90%	10%	5	5	92%
90%	10%	10	5	90%
90%	10%	20	5	91%
90%	10%	40	5	91%
90%	10%	80	5	91%

Dari hasil diatas dapat disimpulkan dengan menggunakan random state sebanyak 5 memberikan hasil akurasi yang lebih baik dibandingkan pengujian random state yang lainnya.

c) *N_estimator testing*

Selanjutnya dilakukan pengujian terhadap N_Estimator atau tree pada tabel di bawah ini:

Tabel 5
(N_estimator testing)

Data Training	Data testing	Random state	N_estimator	Akurasi
90%	10%	5	2	90%
90%	10%	5	3	88%
90%	10%	5	4	91%
90%	10%	5	5	92%
90%	10%	5	6	91%

dari pengujian tree atau N_Estimator dapat disimpulkan bahwa tree atau N_Estimator yang paling baik sebanyak 5 tree atau N_Estimator.

d) Modeling terbaik

Dari semua parameter-parameter yang telah diuji maka dapat disimpulkan bahwa modeling terbaik adalah:

Tabel 6
(Modeling terbaik)

Data Training	Data testing	Random state	N_estimator	Akurasi
90%	10%	5	5	92%

B. Classification Report

Pada tahap ini, dilakukan analisis hasil pada Classification Report yang bertujuan untuk menyimpulkan hasil klasifikasi dapat dikatakan baik atau tidaknya. dari hasil pengujian tersebut didapatkan classification matrix sebagai berikut.

Tabel 7
(Confusion Matrix)

Label	Precision (%)	Recall (%)	F1-score (%)
0 (Baik)	78	80	79
1 (Sangat tidak sehat)	100	100	100
2 (Sedang)	94	95	94
3 (Tidak sehat)	89	80	85
Macro avg	90	89	89
Weighted avg	92	92	91

V. KESIMPULAN

- a) Dalam penelitian ini, data terkait kualitas udara ISPU diklasifikasikan menggunakan algoritma Random Forest dengan parameter PM10, SO2, CO, O3, NO2 sebagai output label yang diklasifikasikan. Pada tahap simulasi dilakukan splitting data menjadi empat rasio 60:40, 70:30, 80:20, 90:10. Kelas yang digunakan adalah "Baik" akan bernilai 0, "Sangat tidak sehat" 1, "Sedang" 2, dan "Tidak Sehat" 3..
- b) Pada evaluasi performasi didapatkan sebesar classification report dengan rincian:
 - a) a. Data 1: precision (78%), recall (80%), dan f1-score (79%)
 - b) b. Data 2: precision (100%), recall

(100%), dan f1-score (100%)

- c) c. Data 3: precision (94%), recall (95%), dan f1-score (94%)
- d) d. Data 4: precision (89%), recall (80%), dan f1-score (85%)

REFERENSI

Laila Fitria, Ririn Arminsih Wulandari, Ema Hermawati, Dewi Susanna. (2008), Kualitas Udara Dalam Ruang Perpustakaan Universitas "X" Ditinjau Dari Kualitas Biologi, Fisik, Dan Kimiawi.

Budanis Dwi Meilani, Muhammad Asadulloh. (2015). Data Mining Untuk Menggali Pola Mahasiswa Baru Menggunakan Metode Frequent Pattern Growth (Studi Kasus : Institut Teknologi Adhi Tama Surabaya).

Aji Primajaya, Betha Nurina Sari. (2018). Random Forest Algorithm for Prediction of Precipitation.

Xueheng Qiua, Le Zhanga, Ponnuthurai Nagarathnam Suganthana, Gehan A. J. Amaratungab. (2017). Oblique Random Forest Ensemble via Least Square Estimation for Time Series Forecasting.

Jaka Prayudha, Ardianto Pranata, Afdal Al Hafiz. (2018). Implementasi Metode Fuzzy Logic Untuk Sistem Pengukuran Kualitas Udara Di Kota Medan Berbasis Internet Of Things (IOT).

Nurdin Zakaria, R. Azizah. (2013). Analisis Pencemaran Udara (SO2), Keluhan Iritasi Tenggorokan dan Keluhan Kesehatan Iritasi Mata Pada Pedagang Makanan di Sekitar Terminal Joyoboyo Surabaya.

Agusta Kurniawan. (2017). Pengukuran Parameter Kualitas Udara (CO,NO2 ,SO2 ,O3 DAN PM10) Di Bukit Kototabang Berbasis ISPU.

David J. Hand, Niall M. Adams. (2015). Data Mining. Karen Hao. (2018). What is Machine Learning? Diakses dari

<https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart/>

Tagliaferri, Lisa. (2017). An Introduction to Machine Learning. Diakses pada 10 April 2021, dari <https://www.digitalocean.com/community/tutorials/anintroduction-to-machine-learning.html>.

M. F. Sanner. (1999). Python: A Programming Language For Software Integration And Development.

Muhammad Yasoob Ullah Khalid. (2019). Intermediate Python.

Vincentius Riandaru Prasetyo, Mirella Mercifia, Anasthasya Averina, Lauren Sunyoto, Budiarjo. (2022). Prediksi Rating Film Pada Website IMDB Menggunakan Metode Neural Network.

Mohamad Dena Nugraha, Judhistira Aria Utama, Santi Sulistiani. (2018). Implementasi Metode Random Forest Dalam Memprediksi Peristiwa Flare Di Siklus Ke-23 Dan 24 Menggunakan WEKA Data Mining.

Maricar, M. Azmar, Widiadnyana, Putu Arta Wijaya, I Wayan. (2017). Analysis of Data Mining for Forecasting Total Goods Delivery with Moving Average Method.

Aji Prasetya Wibawa, Muhammad Guntur Aji Purnama, Muhammad Fathony Akbar, Felix Andika Dwiyanto. (2018). Metode-metode Klasifikasi

