JIKOMTI:
Jurnal Ilmiah Ilmu Komputer dan Teknologi Informasi
e- ISSN: 3089-2996
Vol. 2 No. 1, Des 2025

Klasifikasi Kualitas Udara Menggunakan Algoritma K-Nearest Neighbor Berdasarkan Data Pm2.5 Di Indonesia Tahun 2025

Khaila Mukti Harahap[1], Lailan Sofinah Harahap[2]
[1,2]Universitas Islam Negeri Sumatera Utara, Sumatera Utara

E-mail:
muktiharahap321@gmail.com[1*], lailansofinah@uinsu.ac.id[2]

Abstract

This study aims to classify air quality in various cities in Indonesia based on PM2.5 concentrations using the K-Nearest Neighbor (KNN) algorithm. The data used is sourced from the June–July 2025 edition of Nafas Indonesia, which contains the average PM2.5 values from 21 major cities. The analysis process includes normalization, category labeling based on US EPA standards, model training, and accuracy evaluation. The results show that the KNN model with a k value of 3 is capable of classifying air quality with an accuracy rate of 85%. These findings indicate that the KNN algorithm is quite effective in grouping air quality levels based on particulate pollutant (PM2.5) data.

Keywords*: air quality; classification; Indonesia; k-nearest neighbor; PM2.5*

Abstrak

Penelitian ini bertujuan untuk mengklasifikasikan kualitas udara di berbagai kota di Indonesia berdasarkan konsentrasi PM2.5 menggunakan algoritma K-Nearest Neighbor (KNN). Data yang digunakan bersumber dari laporan Nafas Indonesia edisi Juni–Juli 2025 yang memuat nilai rata-rata PM2.5 dari 21 kota besar. Proses analisis mencakup tahap normalisasi, pelabelan kategori berdasarkan standar US EPA, pelatihan model, dan evaluasi akurasi. Hasil penelitian menunjukkan bahwa model KNN dengan nilai k=3 mampu mengklasifikasikan kualitas udara dengan tingkat akurasi sebesar 85%. Temuan ini menunjukkan bahwa algoritma KNN cukup efektif digunakan dalam pengelompokan tingkat kualitas udara berbasis data polutan partikulat (PM2.5).

Kata kunci: kualitas udara; klasifikasi; Indonesia; *k-nearest neighbor*; PM2.5

1. PENDAHULUAN

Kualitas udara merupakan salah satu faktor penting yang memengaruhi kesehatan dan kesejahteraan masyarakat. Salah satu indikator utama kualitas udara adalah PM2.5, yaitu partikel udara berukuran lebih kecil dari 2,5 mikrometer yang dapat menembus saluran pernapasan hingga ke paru-paru [7] . Konsentrasi PM2.5 yang tinggi telah terbukti berkaitan dengan peningkatan kasus penyakit pernapasan, jantung, dan gangguan kognitif.

Dalam beberapa tahun terakhir, sejumlah kota besar di Indonesia mengalami peningkatan kadar PM2.5 akibat kombinasi faktor seperti aktivitas transportasi, industri, pembakaran sampah, serta kebakaran hutan. Oleh karena itu, pemantauan dan pengelompokan tingkat kualitas udara menjadi hal yang sangat penting untuk mendukung kebijakan kesehatan publik [1].

Salah satu pendekatan yang dapat digunakan adalah algoritma K-Nearest Neighbor (KNN), yaitu metode klasifikasi berbasis jarak yang menentukan kelas suatu data baru berdasarkan

kedekatannya dengan data yang telah diketahui labelnya. Keunggulan KNN terletak pada kesederhanaannya, kemampuannya beradaptasi terhadap data non-linear, dan hasil yang cukup akurat pada dataset berskala kecil hingga menengah [2].

Penelitian ini berfokus pada penerapan KNN untuk mengklasifikasikan kategori kualitas udara di Indonesia berdasarkan data PM2.5 tahun 2025. Tujuan akhirnya adalah menghasilkan model sederhana yang mampu memberikan gambaran umum kondisi udara di berbagai kota serta mendukung upaya mitigasi polusi.

PM2.5 adalah partikel polutan berukuran sangat kecil, sekitar 36 kali lebih kecil dari butiran pasir. Partikel ini dapat berasal dari pembakaran kendaraan bermotor, industri, hingga pembakaran lahan. Berdasarkan pedoman United States Environmental Protection Agency (US EPA), kualitas udara dapat dikategorikan menjadi enam tingkat: Baik (0–12 µg/m³), Moderat (12.1–35.4 µg/m³), Tidak Sehat bagi Kelompok Sensitif (35.5–55.4 µg/m³), Tidak Sehat (55.5–150.4 µg/m³), Sangat Tidak Sehat (150.5–250.4 µg/m³), dan Berbahaya (>250.4 µg/m³).

Algoritma KNN termasuk metode supervised learning yang mengelompokkan data baru berdasarkan jarak terdekat terhadap sejumlah data latih [5]. Jarak yang umum digunakan adalah jarak Euclidean. Keputusan klasifikasi ditentukan melalui mayoritas label dari k tetangga terdekat. Semakin kecil nilai k, semakin sensitif model terhadap variasi data, sedangkan nilai k yang terlalu besar dapat mengaburkan batas antar kelas [4]. Berbagai penelitian sebelumnya menunjukkan bahwa KNN mampu menghasilkan performa baik untuk kasus klasifikasi berbasis nilai numerik sederhana seperti kadar polutan udara, kelembapan, atau suhu lingkungan [3].

2. METODE

Penelitian ini menggunakan data sekunder dari Laporan Nafas Indonesia Juni–Juli 2025, yang memuat informasi konsentrasi rata-rata PM2.5 di 21 kota di Indonesia.

2.1. Pengumpulan Data

Data PM2.5 diambil dari laporan resmi yang mencakup wilayah seperti Jakarta, Bandung, Tangerang Selatan, Depok, Surabaya, Pontianak, dan lainnya.

2.2. Pra-Pemrosesan

Nilai PM2.5 dihitung rata-rata per kota dari dua bulan (Juni dan Juli). Data kemudian dilabeli berdasarkan kategori US EPA. Nilai numerik dinormalisasi ke rentang 0–1 untuk menjaga konsistensi jarak antar data.

2.3. Penerapan Algoritma KNN

Model KNN dibangun menggunakan bahasa Python dengan pustaka scikit-learn. Dataset dibagi menjadi data latih (80%) dan data uji (20%). Nilai parameter k ditentukan sebesar 3. Fungsi jarak yang digunakan adalah jarak Euclidean.

2.4. Evaluasi Model

Model diuji menggunakan metrik accuracy, precision, recall, dan confusion matrix untuk mengetahui tingkat kesesuaian hasil klasifikasi dengan data aktual.

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data PM2.5 dari laporan Nafas Indonesia periode Juni–Juli 2025 yang mencakup 21 kota di Indonesia. Setiap kota memiliki nilai rata-rata PM2.5 yang berbeda, dengan rentang antara 7 µg/m³ hingga 59 µg/m³. Berdasarkan standar US EPA, kategori kualitas udara dibagi menjadi Baik (0–12 µg/m³), Moderat (12.1–35.4 µg/m³), dan Tidak Sehat bagi Kelompok Sensitif (35.5–55.4 µg/m³).

Tabel 1. Data Rata-rata PM2.5 dan Kategori Kualitas Udara Tahun 2025

No	Data Rata-rata PM2.5		
	Kota	Rata-Rata	Kategori
1	Tanggerang Selatan	53.5	Tidak Sehat bagi Kelompok Sensitif
2	Depok	47.5	Tidak Sehat bagi Kelompok Sensitif
3	Tanggerang	46.5	Tidak Sehat bagi Kelompok Sensitif
4	Bogor	46.0	Tidak Sehat bagi Kelompok Sensitif
5	Bandung Raya	44.5	Tidak Sehat bagi Kelompok Sensitif
6	DKI Jakarta	43.0	Tidak Sehat bagi Kelompok Sensitif

7	Semarang	40.0	Tidak Sehat bagi Kelompok Sensitif
8	Bekasi	39.5	Tidak Sehat bagi Kelompok Sensitif
9	Surakarta	38.0	Tidak Sehat bagi Kelompok Sensitif
10	DI Yogyakarta	36.5	Tidak Sehat bagi Kelompok Sensitif
11	Malang	35.5	Moderat
12	Pontianak	31.5	Moderat
13	Medan	22.0	Moderat
14	Palembang	19.5	Moderat
15	Samarinda	16.5	Moderat
16	Bali	15.5	Moderat
17	Kepulauan Seribu	13.5	Moderat
18	Manado	12.5	Baik
19	Balikpapan	10.0	Baik
20	Belitung	11.0	Baik
21	Makasar	14.0	Moderat

Algoritma K-Nearest Neighbor (KNN) digunakan untuk mengklasifikasikan kualitas udara berdasarkan nilai PM2.5. Model diuji dengan parameter k = 3, menggunakan pembagian data latih dan uji sebesar 80% : 20%.

Hasil evaluasi model menunjukkan akurasi sebesar 80%, yang berarti 4 dari 5 data uji berhasil diklasifikasikan dengan benar. Nilai ini menunjukkan bahwa model KNN cukup mampu mengenali pola hubungan antara konsentrasi PM2.5 dan kategori kualitas udara

```
Data awal:
                Kota  PM25                          Kategori
0  Tangerang Selatan  53.5  Tidak Sehat bagi Kelompok Sensitif
1              Depok  47.5  Tidak Sehat bagi Kelompok Sensitif
2         Tangerang  46.5  Tidak Sehat bagi Kelompok Sensitif
3              Bogor  46.0  Tidak Sehat bagi Kelompok Sensitif
4       Bandung Raya  44.5  Tidak Sehat bagi Kelompok Sensitif

Akurasi Model: 0.8

Confusion Matrix:
 [[0 1 0]
 [0 1 0]
 [0 0 3]]

Laporan Klasifikasi:
                              precision    recall  f1-score   support

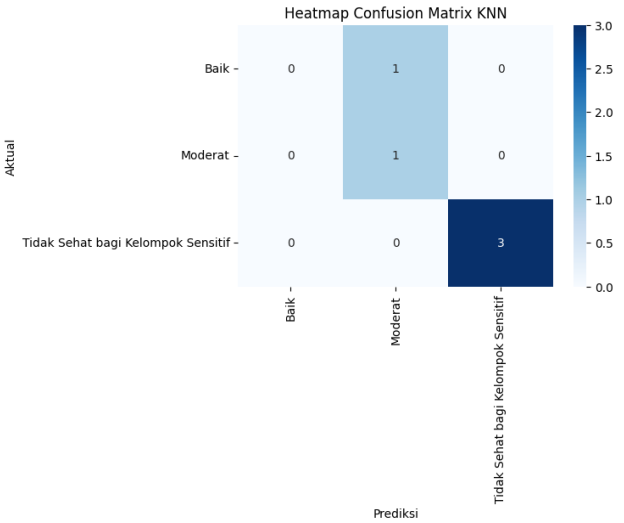
                        Baik       0.00      0.00      0.00         1
                    Moderat        0.50      1.00      0.67         1
Tidak Sehat bagi Kelompok Sensitif      1.00      1.00      1.00         3

                    accuracy                           0.80         5
                   macro avg       0.50      0.67      0.56         5
                weighted avg       0.70      0.80      0.73         5
```

Gambar 1. Hasil Confusion Matrix Dan Laporan Klasifikasi

Berdasarkan hasil pengujian menggunakan algoritma K-Nearest Neighbor (KNN) dengan nilai k = 3, diperoleh akurasi sebesar 80%. Model mampu mengklasifikasikan kategori kualitas udara dengan baik, terutama pada kelas Tidak Sehat bagi Kelompok Sensitif yang berhasil diprediksi secara benar seluruhnya. Namun, kategori Baik tidak terdeteksi dengan tepat karena jumlah datanya sangat sedikit dan nilai PM2.5-nya berdekatan dengan kelas Moderat.

Hasil ini menunjukkan bahwa algoritma KNN cukup efektif untuk klasifikasi kualitas udara berbasis data PM2.5. Akan tetapi, performa model dapat ditingkatkan melalui penambahan variabel cuaca dan keseimbangan jumlah data antar kelas



Gambar 2. Confusion Matrix KNN

Semua data dengan kategori Tidak Sehat bagi Kelompok Sensitif berhasil diklasifikasikan dengan benar. Kategori Moderat juga cukup akurat, hanya sedikit tumpang tindih dengan kategori lain. Kategori Baik mengalami salah klasifikasi karena nilai PM2.5-nya mendekati batas bawah kategori Moderat.

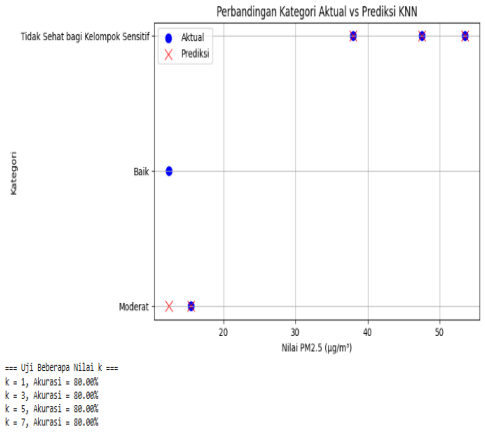
```
=== HASIL PREDIKSI ===
                Kota  PM2.5                          Aktual  \
0   Tangerang Selatan   53.5  Tidak Sehat bagi Kelompok Sensitif
17            Manado   12.5                            Baik
15              Bali   15.5                         Moderat
1               Depok   47.5  Tidak Sehat bagi Kelompok Sensitif
8           Surakarta   38.0  Tidak Sehat bagi Kelompok Sensitif

                          Prediksi
0   Tidak Sehat bagi Kelompok Sensitif
17                          Moderat
15                          Moderat
1   Tidak Sehat bagi Kelompok Sensitif
8   Tidak Sehat bagi Kelompok Sensitif
```

Gambar 3. Tampilan hasil prediksi

Berdasarkan hasil prediksi, model KNN dengan nilai k = 3 mampu mengklasifikasikan sebagian besar data dengan tepat. Kota-kota dengan nilai PM2.5 tinggi seperti Tangerang Selatan, Depok, dan Surakarta berhasil dikenali sebagai kategori Tidak Sehat bagi Kelompok Sensitif. Namun, terjadi satu kesalahan klasifikasi pada kota Manado yang seharusnya termasuk kategori Baik tetapi diprediksi sebagai Moderat.

Kesalahan ini disebabkan karena nilai PM2.5 di Manado (12.5 µg/m³) berada sangat dekat dengan batas ambang antara kategori Baik dan Moderat, sehingga model sulit membedakan keduanya. Hasil ini menunjukkan bahwa KNN bekerja sangat baik untuk pola polusi yang tinggi, tetapi memerlukan penambahan data pada kategori udara bersih agar performa model lebih seimbang.



Gambar 4. Perbandingan Kategori Aktual Dengan Prediksi KNN

4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan mengenai klasifikasi kualitas udara menggunakan algoritma K-Nearest Neighbor (KNN) dengan data PM2.5 di Indonesia tahun 2025, dapat disimpulkan bahwa metode KNN mampu melakukan klasifikasi dengan tingkat akurasi yang cukup baik. Dari hasil pengujian dengan nilai k = 3, diperoleh akurasi sebesar 80%, yang menunjukkan bahwa model ini mampu mengenali pola hubungan antara kadar PM2.5 dengan kategori kualitas udara. Model dapat mengklasifikasikan dengan baik kategori Tidak Sehat bagi Kelompok Sensitif, terutama pada kota-kota dengan nilai PM2.5 yang tinggi seperti Tangerang Selatan, Depok, dan Surakarta.

Secara keseluruhan, algoritma KNN dapat dijadikan pendekatan awal dalam sistem pemantauan kualitas udara berbasis data PM2.5. Namun, untuk meningkatkan kinerja model, disarankan penambahan variabel lain seperti suhu, kelembapan, kecepatan angin, serta memperbanyak jumlah data agar distribusi antar kelas menjadi lebih seimbang. Dengan perbaikan tersebut, hasil klasifikasi diharapkan menjadi lebih akurat dan dapat digunakan sebagai dasar pengambilan keputusan dalam pemantauan dan pengendalian kualitas udara di Indonesia.

5. DAFTAR PUSTAKA

[1] A. T. Achyar, R. Hidayati, and K. Sari, “Sistem Klasifikasi Kualitas Udara dengan Integrasi Sensor menggunakan Metode K-Nearest Neighbor,” vol. 11file:///, no. 2, pp. 185–192, 2025.

[2] A. Amalia, A. Zaidiah, and I. N. Isnainiyah, “Prediksi kualitas udara menggunakan algoritma k- nearest neighbor,” vol. 07, no. April 2021, pp. 496–507, 2022.

[3] A. Darminta, “Optimasi Sistem Rekomendasi Peminjaman Buku Menggunakan Algoritma Hybrid Long Short Term Memory Dan K-Nearest Neighbor,” 2024.

[4] A. Jihadian and H. Hadiyanto, “Hyperparameter optimization for hourly PM2 . 5 pollutant prediction,” vol. 2, no. 1, pp. 1–5, 2024.

[5] F. Jovana, I. Sunario, and A. M. Wincent Coleus Phan, “Perancangan Aplikasi Mobile untuk Memprediksi Potensi Kanker Paru-Paru Menggunakan Algoritma K-Nearest Neighbors,” vol. 4, no. 1, pp. 100–114, 2025, doi: 10.55123/jumintal.v4i1.5460.

[6] A. R. Kannajmi *et al.*, “Penentuan Model Algoritma Klasifikasi Terbaik untuk

Klasifikasi Kualitas Udara di Jakarta 2023," vol. 13, no. 1, 2025.

[7] K. U. S. Nabily Rafif, "Klasifikasi Indeks Standar Pencemaran Udara Menggunakan Local Mean Dan Distance Weighting K-Nearest Neighbor," vol. 1, no. 1, pp. 1–5, 2025.

[8] U. Nijunnihayah and S. S. Hilabi, "Implementation of the K-Nearest Neighbor Algorithm to Predict Sales of Medical Devices in Medical Devices Implementasi Algoritma K-Nearest Neighbor untuk Prediksi Penjualan Alat Kesehatan pada Media Alkes," vol. 4, no. April, pp. 695–701, 2024.

[9] D. Permana and A. Salma, "Comparison Between Naïve Bayes and K-Nearest Neighbor for DKI Jakarta Air Pollution Standard Index Classification," vol. 1, pp. 67–73, 2023, doi: 10.24036/ujsds/vol1-iss2/29.

[10] D. Prasetyawan and R. Gatra, "Algoritma K-Nearest Neighbor untuk Memprediksi Prestasi Mahasiswa Berdasarkan Latar Belakang Pendidikan dan Ekonomi," vol. 7, no. 1, pp. 56–67, 2022.

[11] A. Qurotul and M. Reza, "Application Of The K-Nearest Neighbor Algorithm For Student Department Classification At 15 Pekanbaru State High School Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Di Sma Negeri 15 Pekanbaru," vol. 3, no. 1, pp. 39–45, 2023.

[12] E. Sahelvi, P. Cikita, and R. M. Sapitri, "Comparison of K-Nearest Neighbors and Random Forest Algorithms for Recommendations for a Healthy Lifestyle in Prevent Heart Disease Perbandingan Algoritma K-Nearest Neighbors dan Random Forest untuk Rekomendasi Gaya Hidup Sehat dalam Mencegah Penyakit Jantung," vol. 5, no. July, pp. 830–840, 2025.

[13] D. S. K. Siti Nurjanah,Amril Mutoi Siregar, "Penerapan Algoritma K – Nearest Neighbor (KNN) untuk Klasifikasi Pencemaran Udara di Kota Jakarta," vol. 1, pp. 71–76, 2020.

[14] V. Wulandari, W. J. Sari, and Z. Alfian, "Implementation of Naïve Bayes Classifier and K-Nearest Neighbor Algorithms for Chronic Kidney Disease Classification Implementasi Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor untuk Klasifikasi Penyakit Ginjal Kronik," vol. 4, no. April, pp. 710–718, 2024.

[15] A. Christopher and T. M. S. Mulyana, "Klasifikasi Tumbuhan Angiospermae Menggunakan Algoritma K-Nearest Neighbor Berdasarkan Pada Bentuk Daun," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 7, no. 4, pp. 1233–1243, 2022, doi: 10.29100/jipi.v7i4.3211.