

Penerapan Metode K-Nearest Neighbor (K-NN) Dalam Klasifikasi Kualitas Udara di Provinsi DKI Jakarta

Kevin Yoga Ananta¹, Ester Anugrayningtyas^{2*}, Fitroh Ahmad Abdul Aziz³, Dwi Hartanti⁴

^{1,2,3,4}Informatika

Univesitas Duta Bangsa Surakarta

¹anantayoga04@gmail.com, ^{2*}202030324@mhs.udb.ac.id, ³202020685@mhs.udb.ac.id, ⁴dwhartanti@udb.ac.id

Abstrak— Udara merupakan bagian penting dalam kehidupan makhluk hidup, tetapi dengan berkembangnya kota, pusat industri dan transportasi mengakibatkan perubahan kualitas udara. Pencemaran udara adalah masalah lingkungan terpenting di Indonesia saat ini, seiring dengan pertumbuhan kendaraan bermotor dan ekonomi transportasi. Sebagai kota berkembang DKI Jakarta menyumbang emisi udara yang memperburuk kualitas udara. Oleh karena itu, Dinas Lingkungan Hidup (DLH) Jakarta memantau tingkat pencemaran udara yang disebabkan oleh beberapa zat lain seperti nitrogen dioksida, sulfur dioksida, ozon, karbon monoksida, dan partikulat. Algoritma k-nearest neighbor (KNN) adalah sekumpulan data training yang sebelumnya telah diklasifikasikan dengan metode data mining. Pada penelitian ini, penulis memilih menggunakan metode K-nearest neighbor dalam pengklasifikasian kualitas udara berdasarkan Standard Air Pollution Index (ISPU). Studi ini memberikan K-nearest neighbor akurasi 100%.

Kata kunci— KNN, Udara, Pencemaran Udara, Data Mining

Abstract— Air is an important part of the life of organisms, but the growth of cities, industrial centers and transportation leads to changes in air quality. Air pollution is the most serious environmental problem in Indonesia today, along with the development of motor vehicles and transportation economy. As a developing city, DKI Jakarta contributes to emissions that lead to a decrease in air quality. Therefore, the Jakarta Environment Agency (DLH) monitors the level of air pollution caused by a number of other substances such as nitrogen dioxide, sulfur dioxide, ozone, carbon monoxide and particulate matter. The k-nearest neighbor (KNN) algorithm is a training dataset that has been previously classified by data mining methods. In this study, the authors chose to use the K nearest neighbor method to rank air quality based on the Standard Air Pollution Index (ISPU). This study provides K nearest neighbors with 100% accuracy.

Keywords— KNN, Air, Air Pollution, Data Mining

I. PENDAHULUAN

Perkembangan teknologi informasi telah mempengaruhi berbagai bidang kehidupan baik ekonomi, politik, seni budaya maupun dalam dunia pendidikan. Lingkungan yang baik merupakan kebutuhan dasar manusia untuk mempertahankan kehidupan, sehingga seseorang sangat membutuhkan udara yang baik. Kota yang tumbuh dan berkembang dengan pesat dan padat merupakan penyebab polusi udara[1].

Udara yang tercemar diakibatkan oleh asap kendaraan bermotor, pabrik industry, pabrik kimia, hingga limbah domestik. Semakin besar jumlah penduduk, semakin banyak kita menggunakan alat transportasi. Semua orang menganggap bahwa memiliki alat transportasi akan lebih mudah. Banyaknya kendaraan bermotor menyebabkan asap akan mencemari udara membuat udara tercemar [2]. Selama polusi udara, zat lain bercampur dengan udara. Zat lain tersebut antara lain PM10, SO₂, CO, O₃ dan NO₂ dapat menjadi indikator parameter

kualitas udara di suatu wilayah termasuk DKI Jakarta [3].

Data hasil pendataan beberapa Stasiun Pemantau Kualitas Udara DKI Jakarta pada Desember 2021 dapat memberikan informasi klasifikasi udara sehat, sedang, dan tidak sehat. Namun data tersebut tidak memberikan informasi berupa klasifikasi udara untuk indikator pencemaran udara[7].

Berdasarkan permasalahan di atas, rumusan masalah yang dihasilkan klasifikasi pencemaran udara didasarkan pada indikator parameter pencemaran udara dan pemilihan metode yang cocok digunakan untuk memperoleh hasil tersebut [3].

Berdasarkan permasalahan di atas, maka Batasan masalah untuk klasifikasi kualitas udara pada penerapan metode K-Nearest Neighbor (K-NN) dalam klasifikasi kualitas udara DKI Jakarta adalah berdasarkan indikator parameter pencemaran udara.

II. TINJAUAN PUSTAKA

A. Data Mining

Data Mining dapat diartikan sebagai tahap pemrosesan pengetahuan informasi berdasarkan model Big Data. Data mining dapat digunakan di banyak bidang dan untuk tujuan yang berbeda yaitu untuk menambah pengetahuan, untuk beberapa daerah dapat digunakan untuk meningkatkan penjualan dengan cara pengklasifikasian. Salah satu bagian dari data mining adalah klasifikasi yang digunakan untuk menemukan pola dengan menganalisis data pelatihan yang menjelaskan kategori label atau konsep pengetahuan [5][6].

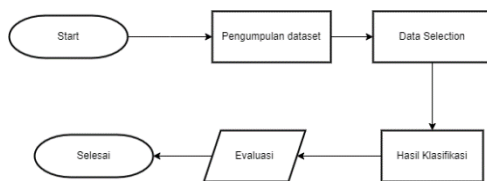
B. K-Nearest Neighbor (K-NN)

K-Nearest Neighbors merupakan salah satu metode data mining yang sering dipakai dalam pengklasifikasi data. Tujuan algoritma KNN untuk mengklasifikasikan objek ke dalam kelas-kelas yang telah ada dalam data sampel yang diberikan.[4].

III. METODOLOGI PENELITIAN

A. Tahapan Umum Penelitian

Langkah klasifikasi pada data mining dengan menggunakan metode KNN memiliki beberapa langkah, gambaran langkah klasifikasi disajikan di Gambar 1.



Gambar 1. tahapan klasifikasi metode KNN

B. Tahap Pengumpulan Data Set

Dalam penelitian ini peneliti menggunakan kumpulan data publik yang diperoleh dari situs resmi pemerintah yaitu Jakarta Open Data <https://data.jakarta.go.id/>. Data Indeks Standar Pencemaran Udara Provinsi DKI Jakarta Desember 2021. Data yang digunakan adalah 31 dan 11 atribut.

C. Data Selection

Proses penelitian diawali dengan pencarian informasi melalui kajian pustaka. Kemudian dilanjutkan dengan menggunakan knowledge discovery in database (KDD) yang digunakan untuk

mendapatkan informasi dari tabel yang ditautkan ke database sehingga dapat digunakan sebagai pengambilan keputusan [4].

Metode ini digunakan untuk memilih atribut, peneliti mempertimbangkan atribut mana yang peneliti gunakan dalam proses data mining. Data ini memiliki 10 atribut yaitu tanggal, O3, PM10, CO dan NO2, SO2, maks, kritis, kategori, lokasi. Peneliti memilih lima atribut diantaranya tanggal, co, o3, no2, kategori.

D. Tahapan Algoritma K-Nearest Neighbor

Langkah-langkah yang terlibat dalam menghitung menggunakan metode KNN antara lain:

- Penentuan jumlah K (jumlah tetangga terdekat).
- Mencari jarak antar data.
- Urutkan benda-benda ini berdasarkan jarak terpendek.
- Mengklasifikasikan data menggunakan ukuran jarak.

Persamaan (a) merupakan persamaan perhitungan untuk KNN.

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2}$$

Keterangan:

$d(x_i, x_j)$ = Jarak
 (x_i) = nilai ke-i
 (x_j) = nilai ke-j
 (a_r) = data ke-r
 $i, j = 1, 2, 3, \dots, n$

E. Tahapan Klasifikasi

Hasil klasifikasi ini memberikan kategori baru pada kualitas udara menurut data training yang ada dan data uji tambahan.

F. Tahapan Evaluasi

Tahap evaluasi adalah tahap merepresentasikan pengetahuan yang diperoleh. Hasil evaluasi dapat digunakan sebagai acuan untuk penelitian selanjutnya.

IV. HASIL DAN PEMBAHASAN

A. Implementasi KNN dengan Perhitungan Manual

Dalam penelitian ini, peneliti menggunakan data dari Jakarta Open Data. Selanjutnya data dapat diolah menggunakan metode Algoritma K-NN.

Tabel 1. Data Penelitian

No	tanggal	co	o3	no2	kategori
1	2021-12-01	13	41	30	SEDANG
2	2021-12-02	7	40	14	SEDANG
3	2021-12-03	9	48	17	SEDANG
4	2021-12-04	13	43	16	SEDANG
5	2021-12-05	17	39	29	SEDANG
6	2021-12-06	23	45	57	SEDANG TIDAK
7	2021-12-07	12	43	21	SEHAT
8	2021-12-08	17	51	30	SEDANG
9	2021-12-09	11	51	20	SEDANG
10	2021-12-10	14	41	30	SEDANG
.
.
25	2021-12-25	23	35	62	SEDANG
26	2021-12-26	15	37	34	SEDANG TIDAK
27	2021-12-27	23	40	47	SEHAT
28	2021-12-28	16	34	33	SEDANG
29	2021-12-29	15	37	29	SEDANG TIDAK
30	2021-12-30	17	38	44	SEHAT
31	2021-12-31	44	37	53	SEDANG

. Kemudian, membuat data uji untuk menentukan klasifikasi baru CO = 68, O3 = 70, NO2 = 65.

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2}$$

$$d = \sqrt{(13 - 68)^2 + (41 - 70)^2 + (65 - 30)^2}$$

$$d = 71,35124386$$

(seterusnya hingga data ke-31), kemudian menentukan ranking setiap data berdasarkan jarak yang paling dekat.

Tabel 2. Data Perhitungan secara manual

No	tanggal	co	o3	no2	kategori	Jarak
1	2021-12-01	13	41	30	SEDANG	71,35124
2	2021-12-02	7	40	14	SEDANG	84,98235
3	2021-12-03	9	48	17	SEDANG	79,17702
4	2021-12-04	13	43	16	SEDANG	78,45381
5	2021-12-05	17	39	29	SEDANG	69,69935
6	2021-12-06	23	45	57	SEDANG TIDAK	52,09607
7	2021-12-07	12	43	21	SEHAT	76,1643
8	2021-12-08	17	51	30	SEDANG	64,70703
9	2021-12-09	11	51	20	SEDANG	75,06664
10	2021-12-10	14	41	30	SEDANG	70,58328
.
.
25	2021-12-25	23	35	62	SEDANG	57,08765

26	2021-12-26	15	37	34	SEDANG TIDAK	69,70653
27	2021-12-27	23	40	47	SEHAT	57
28	2021-12-28	16	34	33	SEDANG	70,88018
29	2021-12-29	15	37	29	SEDANG TIDAK	72,06941
30	2021-12-30	17	38	44	SEHAT	63,76519
31	2021-12-31	44	37	53	SEDANG	42,53234

B. Mengurutkan Data Mulai Dari Jarak Terdekat

Setelah mendapatkan nilai jarak antar objek selanjutnya menentukan ranking data yang didasarkan pada jarak terdekat.

Tabel 3. Urutan 5 jarak tetangga terdekat

No	tanggal	co	o3	no2	kategori	Jarak	rank
31	2021-12-31	44	37	53	SEDANG	42,53234	1
22	2021-12-22	28	54	65	SEDANG TIDAK	43,08132	2
19	2021-12-19	24	78	45	SEHAT	48,98979	3
6	2021-12-06	23	45	57	SEDANG TIDAK	52,09607	4
27	2021-12-27	23	40	47	SEHAT	57	5

C. Klasifikasi Pencemaran Udara

Dari perhitungan klasifikasi Tabel 3 berdasarkan data uji CO = 68, O3 = 70, NO2 = 65. Maka data klasifikasi pencemaran udara termasuk kategori SEDANG.

V. KESIMPULAN

Berdasarkan perhitungan di atas dengan menggunakan perhitungan manual di excel, ternyata diagnosa testing data training "SEDANG" dengan akurasi 100%. Oleh karena itu dapat diambil kesimpulan bahwa hasil prediksi dengan memakai metode K-NN cukup dapat diandalkan sebagai alat pendukung bagi pengambil keputusan.

REFERENSI

- [1] A. I. Sang, E. Sutoyo, and I. Darmawan, "Analisis Data Mining Untuk Klasifikasi Data Kualitas Udara DKI Jakarta Menggunakan Algoritma Decision Tree Dan Support Vector Machine," *e-Proceeding Eng.*, vol. 8, no. 5, pp. 8954–8963, 2021.
- [2] M. Yasir, "Pencemaran Udara Di Perkotaan Berdampak Bahaya Bagi Manusia, Hewan, Tumbuhan dan Bangunan," *J. OSF.Oi*, pp. 1–10, 2021, [Online]. Available: <https://doi.org/10.31219/osf.io/nc5rg>
- [3] A. Wicahyo, A. Pudoli, and D. Kusumaningsih, "Penggunaan Algoritma Naive Bayes dalam klasifikasi Pengaruh Pencemaran Udara," *J. ICT Inf. Commun. Technol.*, vol. 20, no. 1, pp. 103–108, 2021, [Online]. Available: <https://ejournal.ikmi.ac.id/index.php/jict-ikmi/article/view/332>
- [4] A. Yoga Pratama *et al.*, "Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja)," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 897–910, 2021.
- [5] S. H. Wijoyo *et al.*, "MAHASISWA MENGGUNAKAN

- ALGORITME NAÏVE BAYES PADA JURUSAN ANALYSIS OF ATTRIBUTES SELECTION IN PREDICTING STUDENT ' S THESIS FAILURE USING NAÏVE BAYES ALGORITHM IN THE DEPARTMENT OF INFORMATION SYSTEMS UNIVERSITY OF BRAWIJAYA," vol. 10, no. 2, pp. 379–384, 2023, doi: 10.25126/jtiik.2023106426. [6]
- S. S. A. Umri *et al.*, "Analysis and Comparison of Classification Algorithm in Air," vol. 4, no. 2, pp. 98–104, 2021, doi: 10.33387/jiko. Afif Budiyo,"Pencemaran Udara, Dampak Pencemaran Udara Pada Lingkungan," BeiKa Ditgantata Vol.2 No 1 Mmt 2001, [Online]. Available: <https://jurnal.batan.go.id/index.php/bl/article/view/785> [7]