



Department of Digital Business

Journal of Artificial Intelligence and Digital Business (RIGGS)

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 4 No. 2 (2025) pp: 3804-3811

P-ISSN: 2963-9298, e-ISSN: 2963-914X

Clustering Pencemaran Udara Yogyakarta Menggunakan Algoritma K-Means dan Orange Data Mining

Aldyansyah Tarigan¹, Bella Azka Oktavia², Clara Pratiwi³, Fachrizky Catur Putra⁴, Giatika Chrisnawati⁵, Yayan Hendrian⁶

^{1,2,3,4,5,6} Informatika, Teknik dan Informatika, Universitas Bina Sarana Informatika

¹aldytrgn@gmail.com, ²bellaazka61@gmail.com*, ³clarapratiiiii2@gmail.com, ⁴caturputra70@gmail.com,

⁵giatika.gcw@bsi.ac.id, ⁶yayan.yhn@bsi.ac.id

Abstrak

Pencemaran udara menjadi salah satu permasalahan lingkungan yang perlu mendapatkan perhatian khusus, terutama di wilayah perkotaan seperti Daerah Istimewa Yogyakarta. Udara yang tercemar dapat berdampak buruk bagi kesehatan manusia, hewan, dan lingkungan secara keseluruhan. Penelitian ini bertujuan untuk menganalisis dan mengelompokkan kualitas udara di Yogyakarta berdasarkan parameter Indeks Standar Pencemar Udara (ISPU), yaitu PM10, PM2.5, SO₂, CO, O₃, NO₂, dan HC. Metode yang digunakan dalam penelitian ini adalah K-Means Clustering, yang diterapkan untuk mengelompokkan data kualitas udara ke dalam beberapa kluster berdasarkan tingkat pencemaran. Proses pengolahan dan visualisasi data dilakukan menggunakan aplikasi Orange Data Mining, dengan data yang diperoleh dari Dinas Lingkungan Hidup Kota Yogyakarta selama periode tahun 2022 hingga 2024. Hasil dari proses klusterisasi menunjukkan adanya pengelompokan yang jelas terhadap tingkat pencemaran udara di wilayah tersebut, mulai dari kluster dengan tingkat pencemaran rendah hingga tinggi. Temuan ini diharapkan dapat menjadi bahan pertimbangan bagi pemerintah daerah dalam merancang kebijakan pengendalian polusi udara secara lebih efektif dan berbasis data.

Kata kunci: Kualitas Udara, ISPU, K-Means Clustering, Orange Data Mining, Yogyakarta

1. Latar Belakang

Udara dan pencemaran udara merupakan sepasang hal yang kerap menjadi topik kekhawatiran di lingkup kehidupan sehari-hari. [1]. Udara memegang peranan yang sangat penting dalam kehidupan manusia dan makhluk hidup lainnya [2]. Namun di samping itu udara erat hubungannya dengan pencemaran udara. Udara merupakan salah satu zat penting yang menyusun atmosfer dan berada di sekeliling bumi. Udara dapat didefinisikan sebagai campuran gas yang terdapat pada lapisan yang mengelilingi bumi (atmosfer) [3]. Udara terdiri dari nitrogen, oksigen, dan beberapa gas lainnya. udara sangat penting bagi kehidupan kita karena kita membutuhkan oksigen yang ada di udara untuk bernapas.

Udara juga membantu menciptakan iklim dan membawa suara serta bau dan tanaman juga membutuhkan udara untuk tumbuh dan menghasilkan makanan. Penting bagi kita untuk menjaga udara tetap bersih agar kita bisa hidup sehat dan lingkungan kita terjaga. Namun, banyak dari manusia yang tidak menyadari tentang pentingnya menjaga udara agar tetap baik sehingga secara sadar maupun tidak sadar mereka menimbulkan pencemaran udara[4]. Udara juga dapat mengalami pencemaran ketika terdapat sejumlah konsentrasi emisi di dalamnya [5]. Lebih dari 70% pencemaran udara di kota besar Indonesia berasal dari emisi kendaraan bermotor dan aktivitas domestik yang tidak terkontrol, terutama di wilayah urban [6]. Dampak negatif dari penurunan kualitas udara dapat berakibat buruk terhadap kesehatan manusia dan merusak lingkungan [7].

Kesehatan manusia dapat terpengaruh oleh udara yang tercemar, yang berpotensi menyebabkan penyakit seperti infeksi saluran pernapasan, gangguan paru-paru, penyakit jantung, dan juga dapat memicu terjadinya kanker yang berbahaya. Penting bagi masyarakat, terutama yang tinggal di wilayah perkotaan, untuk lebih sadar akan bahaya polusi udara. Sayangnya, masih banyak orang yang tetap beraktivitas di luar rumah meskipun sudah mengetahui bahwa tingkat pencemaran udara berada pada level yang membahayakan kesehatan [8]. Hal ini diperkuat oleh laporan Oktavia yang mencatat konsentrasi PM2.5 mencapai 258 µg/m³ di wilayah Sayidan Yogyakarta, termasuk dalam kategori "beracun", dan menjadikan Yogyakarta sebagai salah satu kota dengan kualitas udara terburuk di Asia Tenggara [9].

Di wilayah DI Yogyakarta sendiri, kualitas udara diukur oleh Dinas Lingkungan Hidup Kota Yogyakarta dan didokumentasikan kedalam suatu indeks yang disebut Indeks Standar Pencemar Udara (ISPU). ISPU adalah suatu laporan hasil pemantauan kualitas udara yang memberikan informasi mengenai kualitas udara yang bersih ataupun udara tercemar serta dampak yang ditimbulkan terhadap kesehatan di lokasi tertentu. ISPU juga berfungsi sebagai sistem informasi peringatan dini terhadap potensi gangguan kesehatan akibat paparan polusi jangka pendek maupun panjang [10]. ISPU disajikan dalam format numerik tanpa satuan spesifik [11]. ISPU atau penamaan secara internasional biasa dikenal dengan Air Quality Index (AQI) adalah indeks yang dipakai untuk mengukur kualitas udara di suatu wilayah [12]. Adapun tujuh parameter yang digunakan yakni PM10, PM2.5, NO₂, SO₂, CO, O₃, dan HC. Oleh karena itu, ISPU dapat digunakan untuk membantu pemerintah dalam mengambil langkah dan kebijakan terhadap pengendalian kualitas udara serta memprediksi kualitas udara di masa yang akan datang [9].

Beberapa penelitian mengenai kualitas udara di DI Yogyakarta cukup banyak dilakukan namun hanya sebatas gambaran dan evaluasi. Oleh karena itu diperlukan penelitian lanjutan yang dapat memprediksi kualitas udara di wilayah Yogyakarta, berdasarkan dataset yang diperoleh dari Dinas Lingkungan Hidup Kota Yogyakarta. Dengan sumber terpercaya dan terbaru. Penelitian ini menggunakan metodologi Cross Industry Standard Process for Data Mining (CRISP DM) untuk mendukung sebuah proses clustering [13], dengan algoritma K-Means Clustering, yang diharapkan dapat memberikan informasi akurat mengenai kondisi udara di suatu wilayah [14].

2. Metode Penelitian

2.1. Dataset Indeks Standar Pencemaran Udara (ISPU)

Dataset yang akan penulis gunakan dalam penelitian ini diperoleh dari Dinas Lingkungan Hidup Kota Yogyakarta. Data yang akan digunakan merupakan data harian kualitas udara berdasarkan parameter Indeks Standar Pencemar Udara (ISPU) yang terdiri dari PM10, PM2.5, SO₂, CO, O₃, NO₂, dan HC. Dataset mencakup periode waktu dari tahun 2022 hingga 2024, dan dikumpulkan dalam bentuk file PDF yang kemudian diubah ke format CSV untuk keperluan analisis. Metode pengumpulan data yang dilakukan dalam penelitian ini melalui dokumentasi dari laporan resmi mengenai kualitas udara di Kota Yogyakarta. Sumber data berasal dari tiga dokumen, yaitu “Data ISPU Rata-Rata Harian AQMS Tahun 2022”, “Data ISPU Rata-Rata Harian AQMS Tahun 2023”, dan “ISPU Kota Yogyakarta Tahun 2024”.

tanggal	pm10	pm2.5	no2	co	o3	no2	hc	catatan
2022-01-01	0.32355	0.531925	0.15933	0.48148	0.03247	0.26442	0.1341	co
2022-01-02	0.20758	0.531925	0.15723	0.42328	0.02427	0.26442	0.14751	co
2022-01-03	0.17137	0.531925	0.18029	0.44611	0.02427	0.2634	0.13648	co
2022-01-04	0.38533	0.531925	0.1782	0.5	0.02927	0.24818	0.13103	co
2022-01-05	0.23154	0.40282	0.48122	0.00646	0.34026	0.13221	0.22283	pm2.5
2022-01-06	0.33337	0.531925	0.00337	0.38895	0.07188	0.15385	0.03172	co
2022-01-07	0.33337	0.531925	0.19078	0.21194	0.37188	0.14423	0.04046	no2
2022-01-08	0.33337	0.531925	0.41635	0.21998	0.37188	0.18029	0.0913	no2
2022-01-09	0.33337	0.531925	0.39023	0.40212	0.37188	0.2478	0.09087	no2
2022-01-10	0.33337	0.531925	0.2691	0.44413	0.37188	0.21944	0.11111	no2
2022-01-11	0.33337	0.531925	0.79833	0.47197	0.37188	0.28363	0.09175	no2
2022-01-12	0.33337	0.531925	0.26305	0.51704	0.37188	0.25942	0.12066	co
2022-01-13	0.18078	0.531925	0.30996	0.32373	0.37188	0.29507	0.13799	co
2022-01-14	0.47495	0.531925	0.38784	0.44746	0.37188	0.32028	0.0747	pm10
2022-01-15	0.56088	0.531925	0.71384	0.53446	0.37188	0.29647	0.07663	pm10
2022-01-16	0.48489	0.531925	0.33227	0.60382	0.37188	0.29452	0.11111	pm10
2022-01-17	0.66667	0.531925	0.30998	0.57907	0.37188	0.28363	0.11111	pm10
2022-01-18	0.48931	0.531925	0.31656	0.57672	0.36091	0.28363	0.11484	pm10
2022-01-19	0.57485	0.531925	0.174	0.44942	0.18182	0.30529	0.11484	pm10
2022-01-20	0.501	0.018	0.30795	0.51658	0.25252	0.33773	0.11809	pm10
2022-01-21	0.21237	0.10795	0.27463	0.3182	0.16066	0.24971	0.12046	co
2022-01-22	0.16786	0.16085	0.13627	0.47884	0.16883	0.25	0.09195	co
2022-01-23	0.17985	0.16885	0.08615	0.45393	0.10714	0.27214	0.11484	co
2022-01-24	0.11589	0.14981	0.08644	0.39947	0.12526	0.24279	0.10728	co
2022-01-25	0.20559	0.21597	0.08366	0.39647	0.24251	0.22837	0.08428	pm2.5
2022-01-26	0.28643	0.31139	0.10063	0.47619	0.25225	0.23558	0.10151	pm2.5
2022-01-27	0.31917	0.34308	0.10327	0.5071	0.1039	0.29883	0.16475	pm2.5
2022-01-28	0.27746	0.34844	0.10023	0.51852	0.09446	0.27443	0.17963	pm2.5
2022-01-29	0.33928	0.41507	0.11133	0.51323	0.06484	0.25002	0.1341	pm2.5
2022-01-30	0.48933	0.54335	0.11111	0.50529	0.08766	0.30789	0.11484	pm2.5
2022-01-31	0.33333	0.38383	0.13627	0.44413	0.07792	0.28948	0.1092	pm2.5

Gambar 1. Dataset Indeks Standar Pencemaran Udara Tahun 2022

tanggal	pm10	pm2.5	so2	co	o3	no2	hc	catatan
2023-01-01	0.38987	0.60037	0.23298	0.41143	0.02421	0.03371	0.29499	pm2.5
2023-01-02	0.38153	0.510683	0.23638	0.41143	0.1386	0.03843	0.18449	pm2.5
2023-01-03	0.14634	0.28089	0.12321	0.22371	0.15429	0.73822	0.07818	no2
2023-01-04	0.14286	0.30238	0.18284	0.29714	0.21769	0.52449	0.07818	no2
2023-01-05	0.13837	0.348578	0.16664	0.33429	0.26647	0.52427	0.11669	no2
2023-01-06	0.13021	0.21322	0.13884	0.28288	0.19379	0.54719	0.10373	no2
2023-01-07	0.12321	0.16699	0.16878	0.37141	0.15179	0.81146	0.13138	no2
2023-01-08	0.10676	0.300975	0.17808	0.33429	0.32456	0.81146	0.11341	no2
2023-01-09	0.15585	0.279636	0.20561	0.37714	0.34561	0.82022	0.13814	no2
2023-01-10	0.21777	0.37054	0.25425	0.48371	0.42681	0.82146	0.15721	pm2.5
2023-01-11	0.14683	0.285959	0.24959	0.446	0.33333	0.81146	0.14811	no2
2023-01-12	0.21777	0.38842	0.27275	0.51429	0.33228	0.81124	0.16664	pm2.5
2023-01-13	0.24642	0.421985	0.27805	0.64429	0.33296	0.80989	0.16207	pm2.5
2023-01-14	0.18912	0.348916	0.17385	0.29143	0.38246	0.80987	0.20678	pm2.5
2023-01-15	0.17777	0.331366	0.14543	0.22386	0.41105	0.82022	0.08104	pm2.5
2023-01-16	0.20537	0.254132	0.1405	0.26271	0.4	0.81786	0.09724	pm2.5
2023-01-17	0.34146	0.488194	0.16228	0.52026	0.27902	0.80989	0.14425	pm2.5
2023-01-18	0.24811	0.429292	0.26448	0.56	0.33333	0.81146	0.16664	pm2.5
2023-01-19	0.37108	0.312488	0.24483	0.49143	0.33988	0.82146	0.14811	pm2.5
2023-01-20	0.28132	0.408481	0.2124	0.42837	0.32688	0.80989	0.13883	pm2.5
2023-01-21	0.18681	0.358347	0.20496	0.44286	0.32683	0.82943	0.13452	pm2.5
2023-01-22	0.21933	0.385339	0.19174	0.41429	0.37188	0.81393	0.18111	pm2.5
2023-01-23	0.12178	0.232383	0.1572	0.38	0.38421	0.81146	0.09712	no2
2023-01-24	0.24813	0.406408	0.1983	0.41429	0.42105	0.82119	0.11954	pm2.5
2023-01-25	0.18815	0.328207	0.18188	0.37714	0.42881	0.80944	0.07293	pm2.5
2023-01-26	0.12178	0.232383	0.1572	0.38	0.38421	0.82022	0.11664	no2
2023-01-27	0.12021	0.221754	0.16664	0.38371	0.19421	0.82022	0.11621	no2
2023-01-28	0.38885	0.513491	0.20982	0.48143	0.33333	0.82146	0.13778	pm2.5
2023-01-29	0.37932	0.50675	0.28119	0.62371	0.23996	0.82146	0.1718	pm2.5
2023-01-30	0.46992	0.50511	0.27834	0.62837	0.25088	0.80989	0.18683	pm2.5
2023-01-31	0.33824	0.473862	0.27828	0.53388	0.13228	0.80920	0.17866	pm2.5
2023-02-01	0.3446	0.39996	0.21157	0.47174	0.23884	0.81124	0.13128	no2
2023-02-02	0.26038	0.471938	0.23214	0.53714	0.17286	0.80944	0.1328	pm2.5
2023-02-03	0.3323	0.471919	0.13388	0.47174	0.37895	0.80764	0.08887	pm2.5

Gambar 2. Dataset Indeks Standar Pencemaran Udara Tahun 2023

	waktu	pm10	pm2.5	so2	co	o3	no2	hc	critical
1	2024-01-01	0.230373	0.4377	0.45195	0.89167	0.88626	0.4865	0.0124	g3
2	2024-01-02	0.27289	0.44188	0.45221	0.89167	0.8831	0.41832	0.0312	g3
3	2024-01-03	0.23352	0.41885	0.57829	0.89167	0.87994	0.659	0.03826	g3
4	2024-01-04	0.16725	0.55981	0.93961	0.3	0.63933	0.40795	0.00413	no2
5	2024-01-05	0.24472	0.38493	0.52135	0.30833	0.16838	0.39854	0.00826	no2
6	2024-01-06	0.51585	0.42513	0.5516	0.25	0.52891	0.57218	0.42555	no2
7	2024-01-07	0.58979	0.48482	0.45552	0.23333	0.64297	0.32427	0.00413	pm2.5
8	2024-01-08	0.28149	0.5178	0.83808	0.30833	0.4793	0.24959	0.03512	pm2.5
9	2024-01-09	0.26585	0.31812	0.45781	0.38133	0.53807	0.16855	0.0618	pm2.5
10	2024-01-10	0.40845	0.38325	0.5089	0.425	0.63507	0.44247	0.01653	no2
11	2024-01-11	0.22359	0.50576	0.33452	0.25	0.64297	0.25418	0.1343	pm2.5
12	2024-01-12	0.6257	0.29738	0.5427	0.40833	0.40136	0.553	0.0623	pm2.5
13	2024-01-13	0.19025	0.48257	0.66941	0.225	0.62875	0.31604	0.0186	pm2.5
14	2024-01-14	0.65845	0.51728	0.66726	0.46667	0.61295	0.99059	0.01033	pm2.5
15	2024-01-15	0.17077	0.48738	0.40214	0.24167	0.74566	0.4749	0.03826	pm2.5
16	2024-01-16	0.24472	0.36126	0.5805	0.30833	0.6404	0.5028	0.0413	no2
17	2024-01-17	0.15317	0.50966	0.44415	0.05	0.01106	0.96925	0.0042	pm2.5
18	2024-01-18	0.40669	0.50262	1	0.375	0.73776	0.6569	0.0124	pm2.5
19	2024-01-19	0.45775	0.4377	0.5089	0.175	0.19008	0.48477	0.03826	no2
20	2024-01-20	0.28972	0.50966	0.47987	0.20833	0.0079	0.38703	0.09504	pm2.5
21	2024-01-21	0.45775	0.34955	0.69781	0.225	0.01106	0.6046	0.5	pm2.5
22	2024-01-22	0.21127	0.51999	0.45552	0.25833	0.52765	0.42155	0.00207	no2
23	2024-01-23	0.383457	0.483402	0.52481	0.00833	0.01106	0.49833	0.110811	pm2.5
24	2024-01-24	0.15356	0.20942	0.45238	0.23333	0.5624	0.55146	0.05942	pm2.5
25	2024-01-25	0.78169	0.5388	0.34669	0.3	0.43033	0.26755	0.0186	pm2.5
26	2024-01-26	0.15317	0.4555	0.73843	0.13333	0.62717	0.2887	0.00413	pm2.5
27	2024-01-27	0.33035	0.5388	0.64591	0.16667	0.51817	0.29289	0.01653	no2
28	2024-01-28	0.25528	0.4042	0.8121	0.475	0.64297	0.11844	0.01833	pm2.5
29	2024-01-29	0.36444	0.71898	0.52481	0.5	0.55688	0.41318	0.99811	pm2.5
30	2024-01-30	0.46303	0.47958	0.63879	0.46667	0.63191	0.36402	0.21901	no2
31	2024-01-31	0.23768	0.49215	0.61744	0.425	0.69036	0.45586	0.01653	no2
32	2024-02-01	0.83275	0.4749	0.8119	0.79667	0.42486	0.18031	0.11157	pm2.5
33	2024-02-02	0.38838	0.29065	0.83836	0.29167	0.35345	0.13812	0.1343	pm2.5
34	2024-02-03	0.30106	0.44712	0.73665	0.50833	0.36651	0.14817	0.66942	pm2.5

Gambar 3. Dataset Indeks Standar Pencemaran Udara Tahun 2024

2.2. Preparasi Data

Tahap preparasi data dapat dilakukan melalui proses data cleansing (pembersihan data). Proses ini bertujuan untuk mendeteksi dan mengoreksi (atau menghilangkan) records yang rusak atau tidak akurat dari record set, tabel, atau database sehingga data tersebut bisa digunakan untuk proses permodelan data mining [15]. Setelah proses pembersihan, data kemudian dinormalisasi agar dapat digunakan secara optimal dalam algoritma K-Means Clustering menggunakan aplikasi Orange Data Mining. Data dikumpulkan untuk setiap hari sepanjang tahun 2022, 2023, dan 2024, meskipun terdapat beberapa hari yang tidak memiliki data lengkap. Setiap entri data menyertakan informasi mengenai komponen pencemar kritis, yaitu parameter yang paling dominan memengaruhi status ISPU harian. Hasil pengamatan menunjukkan bahwa PM2.5 merupakan komponen kritis paling dominan di ketiga tahun tersebut, khususnya pada bulan-bulan kemarau. Komponen lain seperti O₃ dan NO₂ juga muncul sebagai penyebab utama pencemaran udara pada hari-hari tertentu.

2.3. Proses Data

Penelitian ini menggunakan algoritma K-Means Clustering untuk mengelompokkan kualitas udara di Daerah Istimewa Yogyakarta berdasarkan nilai maksimum (max) harian dari tujuh parameter ISPU. Proses klusterisasi dilakukan di aplikasi Orange Data Mining. Nilai max digunakan sebagai indikator utama tingkat pencemaran udara. Algoritma tersebut dapat digunakan untuk menganalisis serta mengelompokkan data kualitas udara yang telah dikumpulkan dan diproses terlebih dahulu. Melalui proses clustering, data yang memiliki nilai parameter ISPU yang serupa dikelompokkan bersama, sehingga dapat diperoleh gambaran mengenai pola dan tren pencemaran udara di wilayah Daerah Istimewa Yogyakarta.

2.4. Hasil

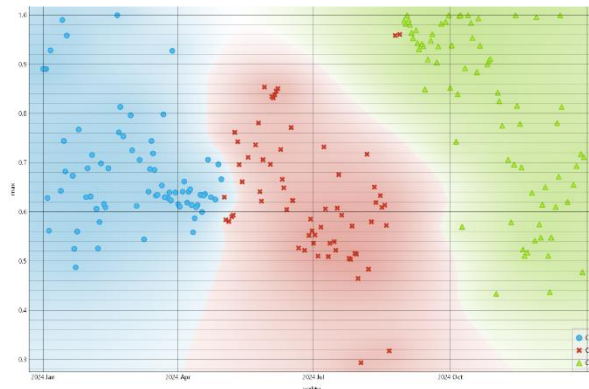
Dengan menggunakan data dari Indeks Standar Pencemaran Udara (ISPU) Kota Yogyakarta Tahun 2022–2024, dilakukan pengelompokan tingkat kualitas udara ke dalam beberapa kategori seperti baik, sedang, dan tidak sehat. Pengelompokan ini didasarkan pada parameter ISPU, yaitu PM10, PM2.5, SO₂, CO, O₃, NO₂, dan HC. Data yang telah diproses kemudian dianalisis menggunakan algoritma K-Means di aplikasi Orange. Hasil dari proses tersebut berupa pembentukan kluster dan nilai centroid dari masing-masing kluster, yang merepresentasikan karakteristik utama dari setiap kelompok kualitas udara.

3. Hasil dan Diskusi

3.1. Hasil Penelitian

Penelitian ini menggunakan algoritma K-Means Clustering untuk mengelompokkan kualitas udara di Daerah Istimewa Yogyakarta berdasarkan data Indeks Standar Pencemaran Udara (ISPU) dari tahun 2022 hingga 2024. Proses analisis dilakukan menggunakan perangkat lunak Orange Data Mining, dengan indikator utama berupa nilai maximum (max) dari tujuh parameter ISPU setiap harinya. Nilai max digunakan sebagai representasi tingkat pencemaran udara, dengan ketentuan yaitu nilai max < 0,6 dapat dikategorikan sebagai udara baik, dan nilai max ≥ 0,6 dapat dikategorikan sebagai buruk. Berikut proses pengujian yang dilakukan sebagai berikut:

Baru setelah memasuki bulan Oktober, ketika curah hujan mulai meningkat, nilai max menunjukkan penurunan bertahap dan kualitas udara mulai membaik.



Gambar 7. Visualisasi Kualitas Udara Tahun 2024 Berdasarkan Nilai Maksimum ISPU Harian

Secara umum, analisis tiga tahun terakhir menunjukkan bahwa fluktuasi kualitas udara di wilayah Yogyakarta sangat dipengaruhi oleh dinamika musim, khususnya curah hujan. Musim hujan secara konsisten berkontribusi pada peningkatan kualitas udara melalui proses pengendapan polutan di atmosfer, sedangkan musim kemarau memperburuk akumulasi pencemaran karena ketiadaan pembersih alami.

3.2. Pengujian Model

Pengujian model dilakukan dengan menggunakan algoritma Random Forest untuk mengevaluasi kemampuan model dalam mengklasifikasikan data berdasarkan hasil klusterisasi K-Means. Evaluasi dilakukan dengan teknik Stratified Shuffle Split sebanyak tiga kali dengan proporsi pelatihan sebesar 75%. Kinerja prediksi dari algoritma random forest dapat diukur menggunakan berbagai metrik, termasuk AUC (area under the curve), CA (akurasi), F1 Score, Precision, Recall, dan MCC [16].

Pada tahun 2022, performa model menunjukkan hasil yang cukup tinggi pada seluruh metrik evaluasi. Nilai AUC mencapai 0,936 pada kategori tertentu, dan akurasi klasifikasi rata-rata berada di atas 85%. Hal ini mengindikasikan bahwa distribusi data pada tahun tersebut relatif stabil dan mudah dipisahkan antar kluster, didukung oleh adanya perbedaan kualitas udara yang cukup kontras antara musim hujan dan kemarau.

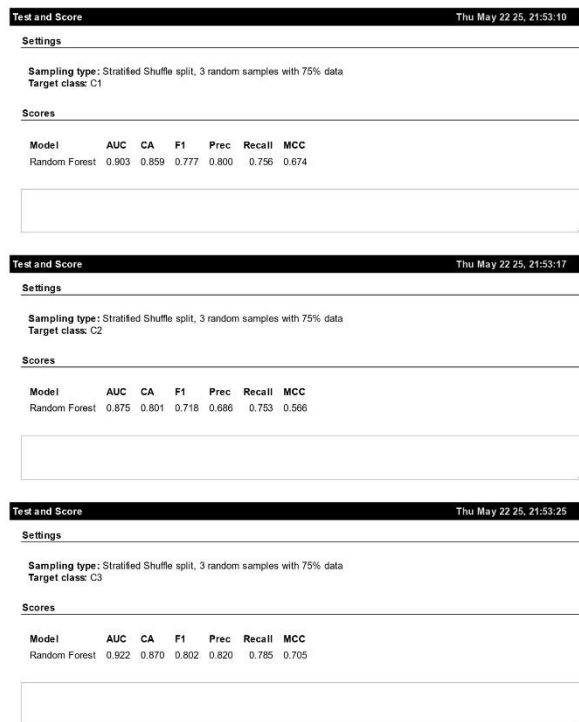
Test and Score		Thu May 22 25, 21:45:57					
Settings							
Sampling type: Stratified Shuffle split, 3 random samples with 75% data							
Target class: C1							
Scores							
Model	AUC	CA	F1	Prec	Recall	MCC	
Random Forest	0.936	0.851	0.794	0.745	0.849	0.682	
Thu May 22 25, 21:46:06							
Settings							
Sampling type: Stratified Shuffle split, 3 random samples with 75% data							
Target class: C2							
Scores							
Model	AUC	CA	F1	Prec	Recall	MCC	
Random Forest	0.906	0.880	0.814	0.857	0.774	0.728	
Thu May 22 25, 21:46:19							
Settings							
Sampling type: Stratified Shuffle split, 3 random samples with 75% data							
Target class: C3							
Scores							
Model	AUC	CA	F1	Prec	Recall	MCC	
Random Forest	0.964	0.870	0.795	0.814	0.778	0.700	

Gambar 8. Test and Score Hasil Kalsifikasi Kualitas Udara Tahun 2022

		Predicted			Σ
		C1	C2	C3	
Actual	C1	74.5 %	11.9 %	4.7 %	93
	C2	8.5 %	85.7 %	14.0 %	93
	C3	17.0 %	2.4 %	81.4 %	90
Σ		106	84	86	276

Gambar 9. Confusion Matrix Hasil Klasifikasi Kualitas Udara Tahun 2022

Tahun 2023 memperlihatkan performa model yang juga solid, dengan AUC mencapai 0,922 dan nilai F1 score serta akurasi berada di atas 80% untuk sebagian besar kategori. Meski terdapat pergeseran musiman yang lebih halus, model masih mampu mengidentifikasi perbedaan antar kluster dengan baik.



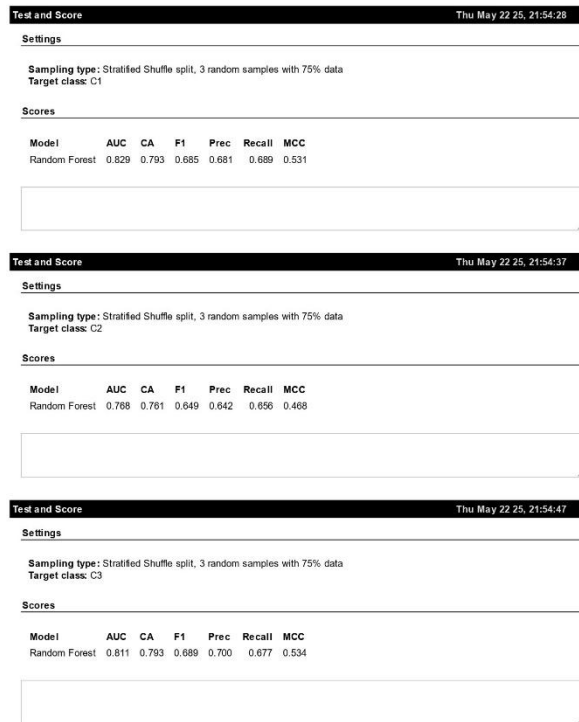
Gambar 10. Test and Score Hasil Kalsifikasi Kualitas Udara Tahun 2023

		Predicted			Σ
		C1	C2	C3	
Actual	C1	80.0 %	20.6 %	1.1 %	90
	C2	9.4 %	68.6 %	16.9 %	93
	C3	10.6 %	10.8 %	82.0 %	93
Σ		85	102	89	276

Gambar 11. Confusion Matrix Hasil Klasifikasi Kualitas Udara Tahun 2023

Sebaliknya, pada tahun 2024, performa model sedikit menurun. Nilai AUC lebih rendah dibandingkan tahun-tahun sebelumnya, dengan nilai terendah pada target klasifikasi C2 yaitu 0.768. Akurasi juga mengalami penurunan pada beberapa kategori. Penurunan performa model pada tahun 2024 diduga berasal dari karakteristik dataset yang kurang variatif. Rentang nilai parameter pencemar udara cenderung sempit dan terdistribusi secara merata, tanpa adanya nilai ekstrem yang menonjol. Selain itu, banyak nilai berada di sekitar batas antar kategori kualitas udara, sehingga menyulitkan model dalam mengelompokkan data secara akurat. Hal ini mengindikasikan bahwa dataset

tahun 2024 belum mencerminkan perbedaan pola pencemaran udara yang cukup kuat seperti pada tahun-tahun sebelumnya.



Gambar 12. Test and Score Hasil Kalsifikasi Kualitas Udara Tahun 2024

		Predicted			
		C1	C2	C3	Σ
Actual	C1	68.1 %	18.9 %	11.1 %	90
	C2	16.5 %	64.2 %	18.9 %	93
	C3	15.4 %	16.8 %	70.0 %	93
Σ		91	95	90	276

Gambar 13. Confusion Matrix Hasil Klasifikasi Kualitas Udara Tahun 2024

Secara keseluruhan, model Random Forest dapat bekerja cukup baik dalam memetakan hasil klusterisasi, terutama ketika distribusi data memiliki perbedaan yang cukup jelas. Fluktuasi performa model dari tahun ke tahun juga memperkuat temuan bahwa kualitas udara di Yogyakarta sangat dipengaruhi oleh siklus musiman, khususnya terkait curah hujan.

Hasil rata-rata curah hujan di Yogyakarta selama tiga tahun terakhir memperlihatkan tren penurunan yang cukup tajam. Berdasarkan data dari BMKG, rata-rata curah hujan bulanan tercatat sebesar 268,48 mm pada tahun 2022, menurun menjadi 157,70 mm pada tahun 2023, dan mencapai titik terendah pada 156,43 mm di tahun 2024.

Minimnya curah hujan pada tahun 2024 menyebabkan peningkatan akumulasi polutan di atmosfer, yang tidak hanya menurunkan kualitas udara, tetapi juga menghasilkan distribusi pencemar yang lebih merata. Hal ini berdampak pada performa model klasifikasi, di mana Random Forest mengalami kesulitan dalam membedakan kluster secara optimal karena hilangnya pola ekstrem yang biasa terjadi pada musim hujan.

Tabel 1. Rata-rata Curah Hujan di Daerah Istimewa Yogyakarta pada 2022-2024

Tahun	Curah Hujan/Bulan												Rata-rata
	Jan.	Feb.	Mar.	Apr.	Mei	Jun.	Jul.	Agu.	Sep.	Okt.	Nov.	Des.	
2022	244,00	386,00	349,00	308,60	169,20	176,80	22,50	74,70	163,50	412,50	375,50	339,50	268,48
2023	264,00	437,00	397,00	65,00	142,00	4,00	38,00	1,00	-	-	117,00	112,00	157,70
2024	498,00	235,00	210,00	113,00	13,00	12,00	14,00	-	-	-	-	-	156,43

Sumber: Badan Pusat Statistik (BPS) Yogyakarta.

DOI: <https://doi.org/10.31004/riggs.v4i2.1116>

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

4. Kesimpulan

Selama periode 2022 hingga 2024, kualitas udara di Daerah Istimewa Yogyakarta menunjukkan pola yang naik-turun, mengikuti siklus musim kemarau dan musim hujan. Dalam analisis ini, algoritma K-Means Clustering digunakan untuk mengelompokkan data ISPU berdasarkan nilai maksimum dari parameter pencemar udara harian. Hasilnya, tiap kelompok mencerminkan kondisi kualitas udara yang berbeda-beda sepanjang tahun. Di tahun 2022, terlihat adanya perbaikan kualitas udara menjelang akhir tahun. Kondisi ini kemungkinan besar dipengaruhi oleh meningkatnya curah hujan yang membantu menurunkan konsentrasi polutan di udara. Memasuki 2023, tren kualitas udara cenderung membaik setelah pertengahan tahun, seiring datangnya musim hujan. Namun, berbeda dengan tahun sebelumnya, tahun 2024 justru menunjukkan penurunan kualitas udara yang cukup tajam, terutama saat musim kemarau. Hal ini menguatkan dugaan bahwa musim kemarau berkontribusi besar terhadap peningkatan kadar polusi. Curah hujan terbukti memainkan peran penting dalam mengurangi polutan di atmosfer. Hal ini terlihat jelas dari perubahan pola pengelompokan data selama setahun penuh. Dari sisi teknis, K-Means terbukti cukup andal dalam membagi data berdasarkan karakteristik kualitas udara, sementara model Random Forest yang digunakan untuk memvalidasi hasil pengelompokan juga menunjukkan performa yang konsisten, terutama saat menangani data yang kontras antara musim kemarau dan musim hujan. Faktor curah hujan terbukti memiliki pengaruh besar terhadap penurunan kadar polutan di atmosfer, sebagaimana ditunjukkan dari perubahan pola kluster sepanjang tahun. Untuk penelitian lanjutan disarankan memasukkan hasil klusterisasi yang dapat dikembangkan menjadi sistem visualisasi atau dashboard yang lebih aplikatif, supaya informasi data terkait kualitas udara dapat diakses secara praktis oleh masyarakat dan pemangku kebijakan.

Referensi

- [1] S. A. Safitri, N. Hikmalia, D. Ulhaq, and A. Salim, "Upaya Peningkatan Kualitas Lingkungan Hidup di Kota Yogyakarta," *JISHUM (Jurnal Ilmu Sos. dan Humaniora)*, vol. 2, no. 4, pp. 527–546, 2024, [Online]. Available: <https://doi.org/10.57248/jishum.v2i4.409>
- [2] M. Hasanuddin, "Sistem Monitoring dan Deteksi Dini Pencemaran Udara Berbasis Internet of Things (IOT)," *J. Comput. Syst. Informatics*, vol. 4, no. 4, pp. 976–984, 2023, doi: 10.47065/josyc.v4i4.4034.
- [3] M. Andini, A. Budianto, L. Mardiana, and S. Rahayu, "Pengukuran Konsentrasi Emisi Partikulat di Ruang Tertutup Menggunakan Kit Pemantauan Kualitas Udara," *Lambda J. Pendidik. MIPA dan Apl. Lemb. "Bale Literasi"*, vol. 5, no. 1, pp. 133–139, 2025, [Online]. Available: <https://doi.org/10.58218/lambda.v5i1.1227>
- [4] A. Budianto *et al.*, "Identifikasi Pengaruh Laju Alir Udara Terhadap Performa Efisiensi Aliran PM dalam Sistem Blower ESP Berbasis IoT," *Lambda J. Pendidik. MIPA dan Apl. Lemb. "Bale Literasi"*, vol. 5, no. 1, pp. 162–168, 2025, [Online]. Available: <https://doi.org/10.58218/lambda.v5i1.1243>
- [5] F. D. C. W, R. Emilia, G. G. P, and F. Indrayatna, "KLASIFIKASI TINGKAT PENCEMARAN UDARA KOTA JAKARTA TAHUN 2021 MENGGUNAKAN ALGORITMA DECISION TREE," *Semin. Nas. Stat. AKTUARIA II*, 2023, [Online]. Available: <https://doi.org/10.1234/sns.v2i1.345>
- [6] I. Purnengsih, E. Nurcahyawati, A. G. Harianto, and D. K. Visual, "Perancangan Infografis Pencemaran Udara Berjudul Polutan Langit," *ASKARA*, vol. 3, no. 128, pp. 167–178, 2025, [Online]. Available: <https://doi.org/10.20895/askara.v3i2.1609>
- [7] P. Y. Utami, R. S. Surya, U. M. Pontianak, and E. Method, "Jurnal Pendidikan Integratif Jurnal Pendidikan Integratif," *J. Pendidik. Integr.*, vol. 6, no. 2, pp. 398–417, 2025, [Online]. Available: <https://ejournals.com/ojs/index.php/jpi/article/view/2270>
- [8] Michael Sitorus, Depriansa Fitron, and Carolus Agung Segara Wisesa, "Implementasi Algoritma K-Means Menggunakan Aplikasi Orange dalam Clustering Pencemaran Udara di DKI Jakarta Tahun 2021," *J. Informatics Adv. Comput.*, vol. 3, no. 2, pp. 161–164, 2022, [Online]. Available: <https://doi.org/10.35814/jiac.v3i2.4662>
- [9] F. Oktavia; and A. Witanti, "Implementasi Prophet Forecasting Model Dalam Prediksi Kualitas Udara Daerah Istimewa Yogyakarta," *J. Tek. Inform. dan Sist. Inf.*, vol. 11, no. 1, pp. 64–74, 2024, [Online]. Available: <https://doi.org/10.35957/jatisi.v11i1.6804>
- [10] Az Zahra Dwi Nur Adiya; Amanda Fitria Desvita; Anindya Fidela; Dwi Amelia; Tri Astuti., "Penerapan Data Mining untuk Klasifikasi Udara di Daerah Istimewa Yogyakarta Menggunakan Algoritma C4.5," *JDMIS J. Data Min. Inf. Syst. Vol.*, vol. 2, no. 2, pp. 59–65, 2024, doi: 10.54259/jdmis.v2i2.2800.
- [11] A. R. Kannajmi *et al.*, "PENENTUAN MODEL ALGORITMA KLASIFIKASI TERBAIK UNTUK KLASIFIKASI KUALITAS UDARA DI JAKARTA 2023," *JITET (Jurnal Inform. dan Tek. Elektro Ter.)*, vol. 13, no. 1, 2025, [Online]. Available: <https://doi.org/10.23960/jitet.v13i1.5664>
- [12] R. C. Irjayana, A. Fadlil, R. Umar, P. Studi, M. Informatika, and U. A. Dahlan, "Pengaruh Seleksi Fitur Terhadap Akurasi Klasifikasi Indeks Standar Pencemar Udara Menggunakan Naive Bayes," *INSECT*, vol. 11, no. 1, 2025, [Online]. Available: <https://doi.org/10.33506/insect.v11i1.4303>
- [13] R. Forest and T. P. Terdahulu, "CLUSTERING DAN KLASIFIKASI DATA CUACA KOTA CILACAP MENGGUNAKAN K-MEANS DAN RANDOM FOREST," *SINTA J. Sist. Inf. dan Teknol. Komputasi*, vol. 1, no. April, pp. 90–97, 2024, doi: 10.61124/sinta.v1i2.15.
- [14] A. Amalia, A. Zaidiah, and I. N. Isnainiyah, "Prediksi Kualitas Udara Menggunakan Algoritma K- Nearest Neighbor," *JIPi (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 07, no. April 2021, pp. 496–507, 2022, [Online]. Available: <https://doi.org/10.29100/jipi.v7i2.2843>
- [15] Fadhlan Sulistiyo Hidayat; Rizma Berliana Putri Affandi; Virgaria Zuliana; Tesa Nur Padilah, "Penerapan K-Means Clustering dalam Pengelompokan Kasus Tuberkulosis di Provinsi Jawa Barat," *J. Ilm. Wahana Pendidik.*, vol. 8, no. September, pp. 219–227, 2022, [Online]. Available: <https://doi.org/10.5281/zenodo.7049113>
- [16] D. B. Saputra *et al.*, "PENERAPAN MODEL CRISP-DM PADA PREDIKSI NASABAH KREDIT," *Idealis Indones. J. Inf. Syst. Vol.*, vol. 7, no. 2021, pp. 240–247, 2024, [Online]. Available: <https://doi.org/10.36080/idealisis.v7i2.3244>