

ANALISIS SPASIAL KUALITAS AIR TELUK JAKARTA BERBASIS K-MEANS DAN SIG

Spatial analysis of Jakarta Bay water quality based on K-Means and GIS

Zaskia Dwiki Utami^{1*}, Muhammad Alif Akbar² dan Malaika Zayda Nuha³

^{1,2,3}Program Studi Sistem Informasi Kelautan, Universitas Pendidikan Indonesia, Jl. Ciracas No. 38, Serang, Banten 42116, Indonesia

**Corresponding author, e-mail: zaskiadwiki@upi.edu*

ABSTRACT

Jakarta Bay receives environmental pressure from urban activities, industrial areas, port operations, and river discharge that carries pollutants into coastal waters. This condition causes water quality to vary across locations, so point-based analysis needs spatial visualization. This study analyzed Jakarta Bay water quality in 2023 using K-Means Clustering and Geographic Information Systems (GIS). The dataset consisted of 53 sampling stations with nine parameters: BOD₅, DO, nitrate, orthophosphate, pH, salinity, temperature, ammonia, and Total Suspended Solids (TSS). The analysis included data checking, data transformation, Z-Score standardization, optimal cluster selection using the Elbow Method and Silhouette Score, K-Means clustering, and visualization in ArcGIS Pro. The evaluation selected four clusters, with a Silhouette Score of 0.4731. The clusters were interpreted as Relatively Good Water Zone, Transitional Water Zone, High Nutrient Pressure Zone, and High Sediment Pressure Zone. Spatial maps showed that water quality was not evenly distributed. Open-water areas and parts of the Seribu Islands tended to have better conditions, while several coastal locations near river mouths and drainage systems showed nutrient or sediment pressure. These results provide an initial basis for monitoring, location prioritization, and decision support in Jakarta Bay water management.

Keywords: Geographic Information Systems, Jakarta Bay, K-Means Clustering, spatial analysis, water quality

ABSTRAK

Teluk Jakarta menghadapi tekanan lingkungan dari aktivitas perkotaan, industri, pelabuhan, serta aliran sungai yang membawa bahan pencemar ke pesisir. Tekanan tersebut membuat kualitas air berbeda antar lokasi, sehingga analisis berbasis titik pengamatan perlu dipadukan dengan pemetaan spasial. Penelitian ini menganalisis kualitas air Teluk Jakarta tahun 2023 dengan K-Means Clustering dan Sistem Informasi Geografis (SIG). Data berasal dari 53 titik sampling dengan sembilan parameter, yaitu BOD₅, DO, nitrat, ortofosfat, pH, salinitas, suhu, amonia, dan Total Suspended Solids (TSS). Tahap analisis mencakup pemeriksaan data, perubahan format data, standarisasi Z-Score, penentuan jumlah cluster melalui Elbow Method dan Silhouette Score, pengelompokan K-Means, dan visualisasi menggunakan ArcGIS Pro. Hasil evaluasi menunjukkan empat cluster sebagai jumlah paling sesuai dengan Silhouette Score 0,4731. Empat cluster tersebut diinterpretasikan sebagai Zona Perairan Relatif Baik, Zona Perairan Transisi, Zona Tekanan Nutrien Tinggi, dan Zona Tekanan Sedimen Tinggi. Peta

hasil clustering memperlihatkan bahwa kondisi kualitas air tidak merata. Wilayah laut terbuka dan sebagian Kepulauan Seribu cenderung lebih baik, sedangkan beberapa lokasi pesisir dekat muara dan sistem drainase menunjukkan tekanan nutrien atau sedimen. Hasil ini dapat menjadi bahan awal untuk pemantauan, penentuan prioritas lokasi, dan dukungan keputusan pengelolaan perairan Teluk Jakarta.

Kata kunci: Analisis spasial, K-Means Clustering, kualitas air, Sistem Informasi Geografis, Teluk Jakarta

PENDAHULUAN

Wilayah pesisir menerima pengaruh langsung dari kegiatan di daratan, terutama pada kawasan padat penduduk, pusat industri, dan area pelabuhan. Teluk Jakarta menghadapi tekanan tersebut karena menerima aliran sungai yang melintasi permukiman dan kawasan perkotaan. Aliran ini membawa limbah organik, nutrien, sampah, serta material tersuspensi ke perairan (Cordova & Nurhati, 2019).

Sungai menjadi jalur penting masuknya bahan pencemar dari wilayah hulu menuju pesisir. Setelah mencapai Teluk Jakarta, material pencemar dapat berpindah mengikuti arus, pasang surut, dan perubahan musim (Iskandar et al., 2021). Akibatnya, kualitas air tidak seragam dan perlu dibaca berdasarkan karakteristik parameter sekaligus posisi geografis setiap titik pengamatan.

Data kualitas air memuat banyak variabel dengan satuan berbeda. K-Means Clustering dapat merangkum data tersebut menjadi kelompok yang mempunyai karakteristik serupa (Handayani et al., 2021). Hasil pengelompokan akan lebih mudah dipahami jika dihubungkan dengan Sistem Informasi Geografis karena SIG dapat menunjukkan posisi titik sampling dan pola persebaran cluster (Pratama et al., 2020). Penerapan SIG untuk memetakan kualitas air juga telah dilakukan pada Teluk Jakarta, Sungai Brantas, dan air sumur perkotaan (de Prima et al., 2016; Kissan et al., 2021; Yoviandianto et al., 2019).

Kajian kualitas perairan Teluk Jakarta umumnya membahas parameter secara deskriptif. Pendekatan tersebut belum selalu memperlihatkan kelompok kondisi perairan dan sebarannya dalam satu analisis. Kebaruan penelitian ini terletak pada penggabungan K-Means dan SIG untuk mengolah sembilan parameter dari 53 titik sampling menjadi zona kualitas air. Penelitian ini bertujuan menganalisis karakteristik kualitas air Teluk

Jakarta, mengelompokkan titik pengamatan dengan K-Means, dan memetakan hasilnya menggunakan SIG.

METODE PENELITIAN

Waktu dan Tempat Penelitian

Penelitian dilaksanakan pada Mei sampai Juni 2026 dengan wilayah kajian perairan Teluk Jakarta dan Kepulauan Seribu, Provinsi DKI Jakarta. Data yang dianalisis berasal dari pemantauan Januari sampai Desember 2023 pada 53 lokasi sampling. Setiap lokasi memiliki koordinat lintang dan bujur untuk mendukung pemetaan.

Bahan dan Alat

Bahan penelitian berupa dataset “Data Kualitas Air Teluk dan Laut Tahun 2023” yang diunduh dari portal Satu Data Jakarta pada 28 Mei 2026 (Pemerintah Provinsi DKI Jakarta 2023). Parameter yang digunakan meliputi Biological Oxygen Demand (BOD₅), Dissolved Oxygen (DO), nitrat (NO₃-N), ortofosfat (PO₄-P), pH, salinitas, suhu, total amonia (NH₃-N), dan Total Suspended Solids (TSS). Parameter fisika dan kimia tersebut umum digunakan untuk menggambarkan kondisi perairan, sedangkan nitrat, ortofosfat, dan amonia berkaitan dengan masukan nutrisi (Jaffar et al., 2020). Pengolahan data dilakukan melalui Google Colab, sedangkan pemetaan menggunakan ArcGIS Pro.

Rancangan dan Pengumpulan Data

Penelitian menggunakan pendekatan kuantitatif deskriptif dengan analisis kluster dan analisis spasial. Data dikumpulkan melalui dokumentasi dataset sekunder. Informasi yang dipakai mencakup nama lokasi, waktu pengamatan, koordinat geografis, dan nilai sembilan parameter kualitas air. K-Means, yang diperkenalkan oleh MacQueen (1967), dipilih karena mampu mengelompokkan data tanpa label awal berdasarkan kemiripan karakteristiknya. (Handayani et al., 2021) juga menerapkan K-Means pada data kualitas air.

Pengolahan dan Analisis Data

Data diperiksa untuk menemukan nilai kosong, data ganda, dan perbedaan format. Pra-pengolahan diperlukan karena kualitas data awal memengaruhi hasil clustering (Setyaningtyas et al., 2022). Data yang semula berbentuk long format diubah menjadi wide format melalui pivot table, sehingga setiap lokasi menjadi satu observasi dan setiap parameter menjadi variabel. Seluruh variabel kemudian distandardisasi dengan Z-Score agar perbedaan satuan tidak membuat variabel tertentu mendominasi perhitungan jarak (Hidayat & Fitriana, 2022)

$$Z = (x - \mu) / \sigma$$

Keterangan:

z = nilai hasil standardisasi

x = nilai observasi

μ = rata-rata data

σ = standar deviasi

Jumlah cluster ditentukan melalui Elbow Method dan Silhouette Score. Elbow Method membaca perubahan inertia pada $K = 1$ sampai $K = 10$, sedangkan Silhouette Score menilai kedekatan data dalam cluster dan jaraknya terhadap cluster lain sebagaimana dikembangkan oleh Rousseeuw (1987). Kombinasi kedua ukuran tersebut membantu memilih jumlah cluster yang tetap informatif dan mudah diinterpretasikan (Fatkuroji et al., 2022).

$$J = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$$

Keterangan:

J = nilai *inertia*

k = jumlah cluster

x = data pengamatan

C_i = cluster ke- i

μ_i = centroid cluster ke- i

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

Keterangan:

$S(i)$ = nilai silhouette untuk objek ke- i

$a(i)$ = rata-rata jarak objek ke- i terhadap seluruh anggota dalam cluster yang sama

$b(i)$ = rata-rata jarak objek ke- i terhadap anggota pada cluster terdekat lainnya

K-Means dijalankan pada data yang telah distandardisasi. Setiap titik ditempatkan pada centroid terdekat berdasarkan Euclidean Distance, kemudian centroid dihitung ulang dari rata-rata anggotanya. Proses berulang sampai keanggotaan cluster tidak berubah. Nilai rata-rata setiap parameter selanjutnya digunakan untuk membaca karakteristik masing-masing kelompok.

$$d(x, y) = \sqrt{\sum (x_i - y_i)^2}$$

Keterangan:

$d(x, y)$ = jarak antar objek

x_i = nilai variabel ke- i pada objek pertama

y_i = nilai variabel ke- i pada objek kedua

Label cluster digabungkan dengan koordinat di ArcGIS Pro. Bujur ditempatkan sebagai koordinat X dan lintang sebagai koordinat Y. Titik spasial dibuat melalui fitur XY Table to Point dengan sistem koordinat GCS WGS 1984 (EPSG:4326). Peta menampilkan simbol berbeda untuk setiap cluster dan grid koordinat agar posisi lokasi dapat dibaca dengan jelas. Penggunaan SIG membantu menghubungkan variasi kualitas air dengan letak geografisnya (Pratama et al., 2020).

Prosedur Penelitian

Alur penelitian mencakup pengunduhan dan seleksi data, pemeriksaan kelengkapan, perubahan format, standardisasi, penentuan jumlah cluster, K-Means, analisis karakteristik cluster, dan pemetaan spasial. Rangkaian tahap tersebut disajikan pada Gambar 1.

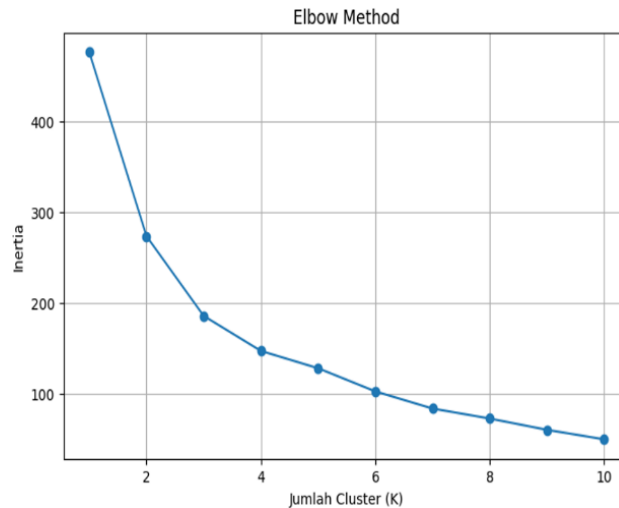


Gambar 1. Alur penelitian

HASIL DAN PEMBAHASAN

Penentuan Jumlah Cluster Optimal

Kurva Elbow pada Gambar 2 menunjukkan penurunan inerti yang tajam hingga $K = 4$. Setelah titik tersebut, kurva mulai melandai sehingga penambahan cluster tidak lagi memberikan penurunan inerti yang besar.



Gambar 2. Grafik Elbow Method

Nilai Silhouette Score untuk beberapa jumlah cluster disajikan pada Tabel 1. $K = 4$ menghasilkan nilai 0,4731. Nilai tertinggi memang terdapat pada $K = 2$, tetapi dua kelompok belum mampu membedakan kondisi perairan relatif baik, transisi, tekanan nutrien, dan tekanan sedimen. Karena itu, $K = 4$ dipilih dengan mempertimbangkan kurva Elbow, kualitas pemisahan, dan kebutuhan interpretasi spasial. Pendekatan yang menggabungkan beberapa ukuran evaluasi (Fatkuroji et al., 2022) juga menggunakan pendekatan tersebut.

Tabel 1. Nilai Silhouette Score pada berbagai jumlah cluster

Jumlah cluster (K)	Silhouette Score
2	0,4862
3	0,4592
4	0,4731
5	0,4552
6	0,4712

Hasil Pengelompokan dan Karakteristik Cluster

K-Means membagi 53 titik sampling ke dalam empat cluster. Cluster 0 berisi 31 titik, Cluster 1 sebanyak 15 titik, Cluster 2 sebanyak 4 titik, dan Cluster 3 sebanyak 3 titik (Tabel 2). Jumlah anggota yang tidak sama bukan merupakan kesalahan karena K-Means mengelompokkan data berdasarkan kemiripan, bukan membagi anggota secara merata.

Pola tersebut menunjukkan bahwa sebagian besar lokasi mempunyai karakteristik umum, sedangkan beberapa titik memiliki tekanan lingkungan yang lebih khas. Temuan ini sesuai dengan fungsi K-Means dalam merangkum data kualitas air multidimensi (Handayani et al., 2021).

Tabel 2. Jumlah anggota setiap cluster

Cluster	Jumlah anggota
0	31
1	15
2	4
3	3

Perbedaan karakteristik antarkelompok paling jelas terlihat pada DO, salinitas, nitrat, ortofosfat, amonia, dan TSS (Tabel 3). Cluster 0 mempunyai DO 6,263 mg/L dan salinitas 28,368 ppt, sedangkan nilai nutrisi serta TSS relatif rendah. Kondisi ini menunjukkan pengaruh laut yang masih kuat, sehingga kelompok tersebut disebut Zona Perairan Relatif Baik. Istilah relatif baik digunakan untuk membandingkan cluster dalam dataset dan bukan pernyataan bahwa seluruh nilainya telah memenuhi baku mutu.

Tabel 3. Nilai rata-rata parameter kualitas air pada setiap cluster

Parameter	Cluster 0	Cluster 1	Cluster 2	Cluster 3
BOD ₅	1,515	1,266	1,262	1,230
DO	6,263	4,742	4,144	4,600
Nitrat	0,071	0,136	1,042	0,471
Ortofosfat	0,008	0,044	0,200	0,047
pH	8,226	7,810	7,361	7,609
Salinitas	28,368	25,508	20,831	19,842
Suhu	28,585	29,950	29,975	29,442

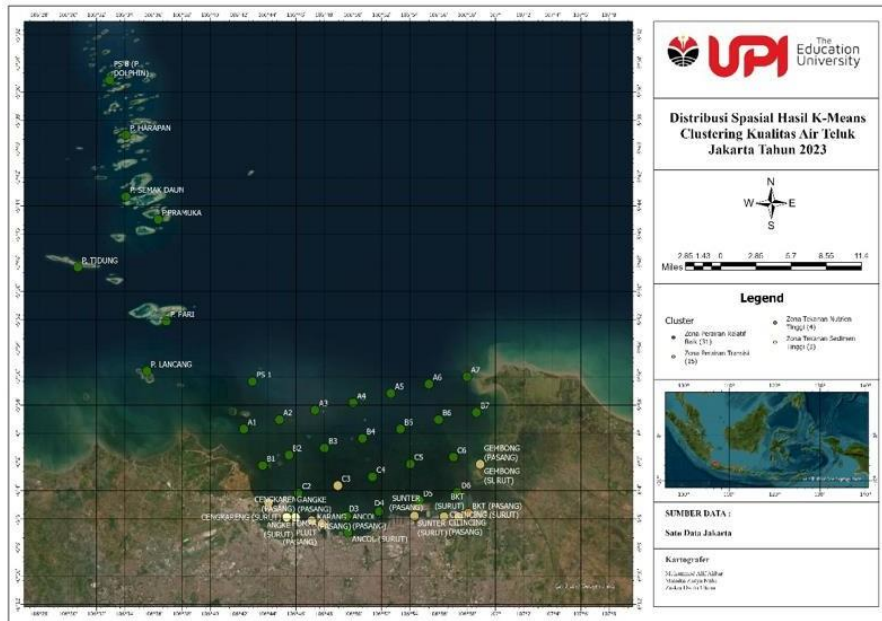
Parameter	Cluster 0	Cluster 1	Cluster 2	Cluster 3
NH ₃ -N	0,048	0,332	1,698	0,563
TSS	8,435	13,267	33,125	98,500

Cluster 1 menunjukkan kondisi peralihan. DO dan salinitas lebih rendah daripada Cluster 0, sedangkan nitrat, ortofosfat, amonia, dan TSS mulai meningkat. Cluster 2 mempunyai konsentrasi nutrisi tertinggi, yaitu nitrat 1,042 mg/L, ortofosfat 0,200 mg/L, dan amonia 1,698 mg/L, disertai DO terendah sebesar 4,144 mg/L. Masukan nitrogen, fosfor, dan bahan organik dari aktivitas manusia dapat meningkatkan risiko eutrofikasi di wilayah pesisir (Adyasari et al., 2021). Karena itu, Cluster 2 disebut Zona Tekanan Nutrien Tinggi.

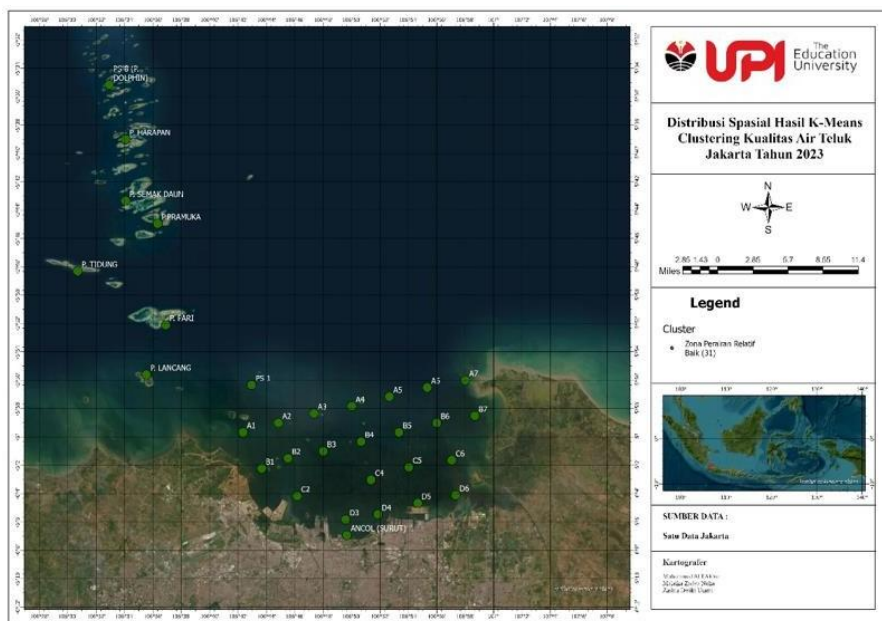
Cluster 3 dibedakan oleh TSS sebesar 98,500 mg/L, jauh di atas cluster lain. Nilai tersebut menunjukkan banyaknya material tersuspensi yang dapat berasal dari limpasan permukaan, pengadukan sedimen, dan aktivitas di sekitar muara. Berdasarkan ciri tersebut, Cluster 3 disebut Zona Tekanan Sedimen Tinggi.

Distribusi Spasial Zona Kualitas Air Teluk Jakarta

Peta keseluruhan pada Gambar 3 menunjukkan bahwa kualitas air Teluk Jakarta tersebar secara heterogen. Zona Perairan Relatif Baik mendominasi wilayah laut terbuka dan beberapa lokasi di Kepulauan Seribu. Sebaran khusus Cluster 0 disajikan pada Gambar 4.

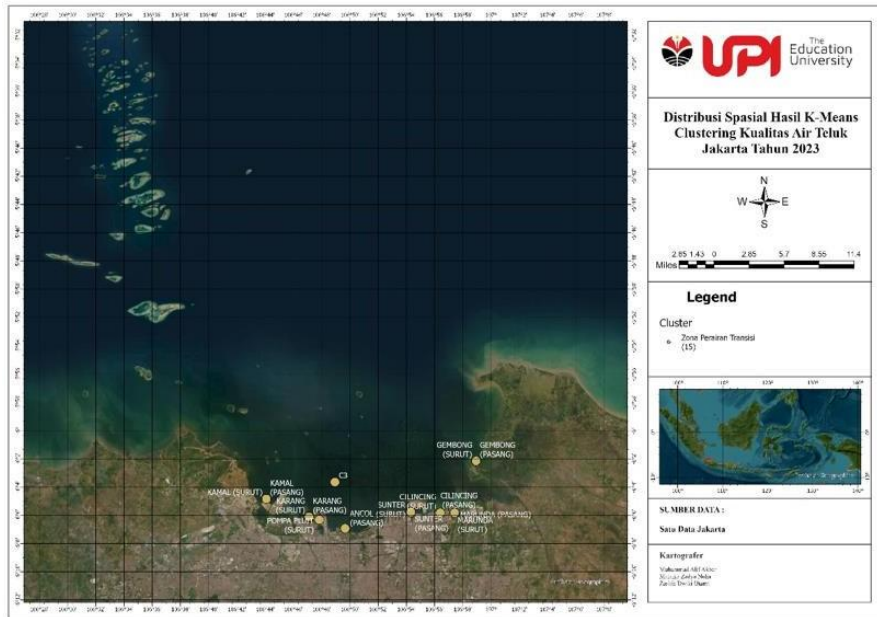


Gambar 3. Peta zonasi kualitas air Teluk Jakarta berdasarkan K-Means Clustering tahun 2023



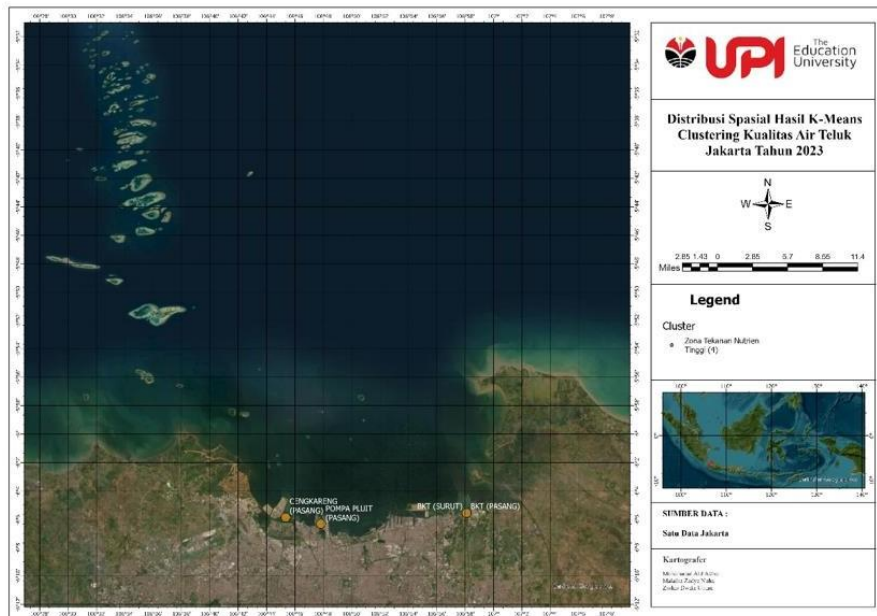
Gambar 4. Sebaran Cluster 0 atau Zona Perairan Relatif Baik

Wilayah laut terbuka cenderung memiliki sirkulasi dan pengenceran yang lebih kuat serta berjarak lebih jauh dari sumber pencemar daratan. Sebaliknya, Cluster 1 banyak muncul di pesisir sebagai zona peralihan antara perairan terbuka dan kawasan yang menerima masukan dari daratan (Gambar 5).



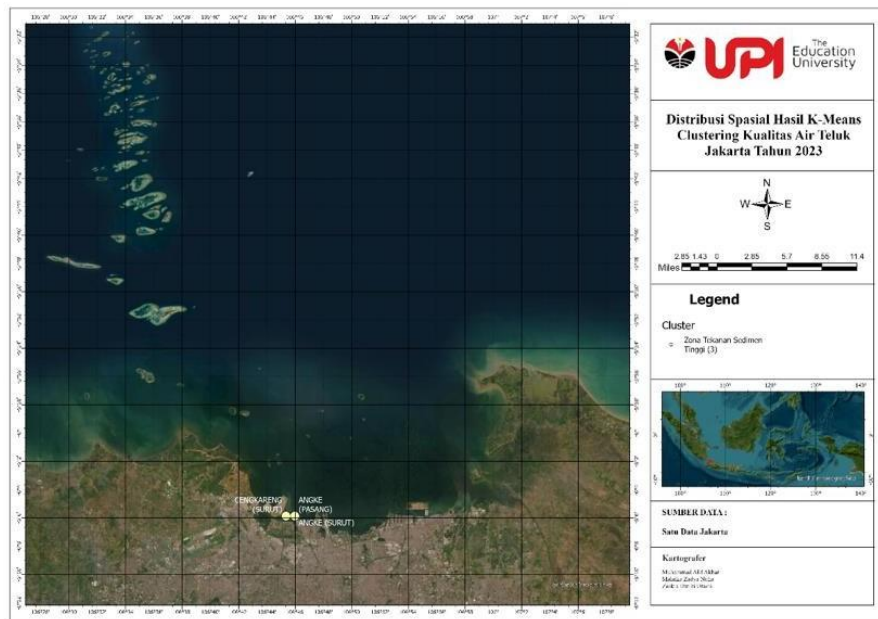
Gambar 5. Sebaran Cluster 1 atau Zona Perairan Transisi

Zona Tekanan Nutrien Tinggi ditemukan pada BKT saat pasang dan surut, Cengkareng saat pasang, serta Pompa Pluit saat pasang (Gambar 6). Lokasi tersebut berhubungan dengan drainase, muara, dan aliran sungai. Sungai merupakan jalur utama masuknya bahan pencemar dari daratan ke Teluk Jakarta (Cordova & Nurhati, 2019). Kondisi perairan di kawasan ini juga dipengaruhi aktivitas pesisir (Akbar et al., 2024).



Gambar 6. Sebaran Cluster 2 atau Zona Tekanan Nutrien Tinggi

Zona Tekanan Sedimen Tinggi terdapat di Angke saat pasang dan surut serta Cengkareng saat surut (Gambar 7). Perbedaan kondisi pasang dan surut memperlihatkan bahwa konsentrasi material tidak hanya ditentukan oleh sumber dari daratan, tetapi juga oleh dinamika perairan. (Iskandar et al., 2021) menjelaskan bahwa arus dapat memindahkan material pencemar sehingga sebarannya berubah antarlokasi.



Gambar 7. Sebaran Cluster 3 atau Zona Tekanan Sedimen Tinggi

Secara keseluruhan, integrasi K-Means dan SIG memperlihatkan hubungan antara karakteristik kualitas air dan posisi geografis titik sampling. Hasil zonasi dapat digunakan sebagai informasi awal untuk memilih lokasi yang memerlukan pemantauan lebih intensif.

KESIMPULAN

Kualitas air Teluk Jakarta tahun 2023 terbagi menjadi empat cluster berdasarkan sembilan parameter, dengan Silhouette Score 0,4731. Cluster tersebut membentuk Zona Perairan Relatif Baik, Zona Perairan Transisi, Zona Tekanan Nutrien Tinggi, dan Zona Tekanan Sedimen Tinggi. Perbedaan antarzona terutama terlihat pada DO, salinitas, nitrat, ortofosfat, amonia, dan TSS. Pemetaan menunjukkan bahwa wilayah laut terbuka dan sebagian Kepulauan Seribu cenderung berada pada zona relatif baik, sedangkan

tekanan nutrien dan sedimen lebih banyak muncul di pesisir dekat muara, drainase, dan aktivitas daratan.

Kombinasi K-Means dan SIG mampu menyederhanakan data kualitas air menjadi zonasi yang mudah dibaca dan dapat mendukung penentuan prioritas pemantauan. Penelitian berikutnya perlu menggunakan data beberapa musim atau tahun, menambahkan arus, pasang surut, curah hujan, dan debit sungai, serta membandingkan K-Means dengan DBSCAN, Hierarchical Clustering, atau Fuzzy C-Means.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Program Studi Sistem Informasi Kelautan Universitas Pendidikan Indonesia atas dukungan selama penelitian. Terima kasih juga disampaikan kepada Pemerintah Provinsi DKI Jakarta melalui portal Satu Data Jakarta yang menyediakan data kualitas air Teluk dan Laut tahun 2023, serta kepada seluruh pihak yang membantu pengolahan data, pemetaan, dan penyusunan artikel.

DAFTAR PUSTAKA

- Adyasari, D., Pratama, M. A., Teguh, N. A., Sabdaningsih, A., Kusumaningtyas, M. A., & Dimova, N. (2021). Anthropogenic impact on Indonesian coastal water and ecosystems: Current status and future opportunities. *Marine Pollution Bulletin*, *171*, 112689. <https://doi.org/10.1016/j.marpolbul.2021.112689>
- Akbar, H. I., Sapanli, K., Kusumastanto, T., Damar, A., Ambarwulan, W., & Rosalia, A. A. (2024). The role of water conditions in Jakarta Bay on the sustainability of capture fisheries production. *AACL Bioflux*, *17*(6), 2765–2774.
- Cordova, M. R., & Nurhati, I. S. (2019). Major sources and monthly variations in the release of land-derived marine debris from the Greater Jakarta area, Indonesia. *Scientific Reports*, *9*, 18730. <https://doi.org/10.1038/s41598-019-55065-2>
- de Prima, C., Hartoko, A., & Muskananfolo, M. R. (2016). Analisis sebaran spasial kualitas perairan Teluk Jakarta. *Diponegoro Journal of Maquares*, *5*(2), 51–60.
- Fatkuroji, Moh., Fajrizal, Taslim, Sabna, E., & Ningsih, K. W. (2022). Optimasi nilai K pada algoritma K-Means untuk klusterisasi data pasien Covid-19. *Indonesian Journal of Computer Science*, *11*(2), 697–707.

- Handayani, Y., Syafrudin, S., & Suherman, S. (2021). Analisis kluster K-Means terhadap kualitas air sungai di Kota Surakarta, Jawa Tengah. *Seminar Nasional Official Statistics, 2021*(1), 139–145. <https://doi.org/10.34123/semnasoffstat.v2021i1.800>
- Hidayat, M. K., & Fitriana, R. (2022). Implementasi K-Means dan K-Medoids dalam pengelompokan wilayah potensial produksi daging ayam. *Jurnal Teknologi Industri Pertanian, 32*(3), 239–247. <https://doi.org/10.24961/j.tek.ind.pert.2022.32.3.239>
- Iskandar, M. R., Surinati, D., Cordova, M. R., & Siong, K. (2021). Pathways of floating marine debris in Jakarta Bay, Indonesia. *Marine Pollution Bulletin, 169*, 112511. <https://doi.org/10.1016/j.marpolbul.2021.112511>
- Jaffar, A., Thamrin, N. M., Megat Ali, M. S. A., Misnan, M. F., & Mohd Yassin, A. I. (2020). The influence of physico-chemical parameters to determine water quality: A review. *Journal of Electrical and Electronic Systems Research, 17*, 116–121. <https://doi.org/10.24191/jeesr.v17i1.016>
- Kissan, Rauf, M., Selintung, M., & Bakri, B. (2021). Sistem Informasi Geografis kualitas air sumur di Kota Makassar. *JACEE, 1*(1), 78–85.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1*, 281–297.
- Pratama, M. A., Immanuel, Y. D., & Marthanty, D. R. (2020). A multivariate and spatiotemporal analysis of water quality in Code River, Indonesia. *The Scientific World Journal, 2020*, 8897029. <https://doi.org/10.1155/2020/8897029>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Setyaningtyas, S., Nugroho, B. I., & Arif, Z. (2022). Tinjauan pustaka sistematis pada data mining: Studi kasus algoritma K-Means Clustering. *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang, 10*(2), 52–61. <https://doi.org/10.21063/jtif.2022.V10.2.52-61>
- Yoviandianto, I. A., Mahmudi, M., & Darmawan, A. (2019). Pemetaan distribusi kualitas air untuk mendukung pengelolaan sumberdaya perairan dengan Sistem Informasi Geografis, kasus di Sungai Brantas, Kecamatan Bumiaji. *Journal of Fisheries and Marine Research, 3*(3), 372–380.