

# ANALISIS TINGKAT PENCEMARAN LINGKUNGAN PADA KOTA/KABUPATEN DI JAWA TENGAH MENGGUNAKAN METODE K-MEANS

Achmad Resnu Maulana<sup>1</sup>

<sup>1,2</sup>Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Ampel Surabaya

*Riwayat artikel:*

*Received: 15 Juni 2023*

*Accepted: 10 Juli 2023*

*Published: 1 Agustus 2023*

## Keywords:

Clustering, K-Means,  
Pencemaran Lingkungan

## Correspondent Email:

resnum96@gmail.com

**Abstrak.** Kegiatan penambangan dapat menimbulkan dampak pencemaran lingkungan, seperti yang terjadi di desa Wadas dan Kendeng, Jawa tengah. Tujuan dari penelitian ini adalah untuk mengelompokkan tingkat pencemaran lingkungan yang berada di Jawa Tengah berdasarkan sejumlah parameter yaitu, jumlah desa yang terdampak pencemaran air, pencemaran udara dan tanah. Metode clustering pada penelitian ini menggunakan algoritma K-Means. Didapatkan *cluster* terbaik 0.51 dengan 2 cluster. Berdasarkan tingkat pencemaran lingkungan, temuan mengungkapkan bahwa Jawa Tengah terbagi menjadi dua *cluster* kota/kabupaten: *cluster* 1 (dengan tingkat pencemaran rendah) dengan 14 anggota, dan *cluster* 2 (dengan tingkat pencemaran rendah) dengan 24 anggota.

© 2023 JITET (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY NC)

**Abstract.** Mining activities can cause environmental pollution, as happened in the villages of Wadas and Kendeng, Central Java. The aim of this research is to classify the level of environmental pollution in Central Java based on a number of parameters, namely, the number of villages affected by water pollution, air pollution and soil pollution. The clustering method in this study uses the K-Means algorithm. The best cluster is 0.51 with 2 clusters. Based on the level of environmental pollution, the findings reveal that Central Java is divided into two city/district clusters: cluster 1 (with a low pollution level) with 14 members, and cluster 2 (with a low pollution level) with 24 members.

## 1. PENDAHULUAN

Pertambangan adalah rangkaian kegiatan dalam rangka upaya pencarian, penambangan (penggalian), pengolahan, pemanfaatan dan penjualan bahan galian (mineral, batubara, panas bumi, migas)[1]. Akan dampak ekologis yang timbul akibat pertambangan juga sangat nampak, dari pencemaran air, penurunan muka tanah, erosi dan lain-lain. Seperti halnya yang terjadi pada desa Wadas dan Kendeng. Terdapat proyek pembangunan pabrik semen di Kendeng dan Penambangan Batu Andesit di Wadas yang menyebabkan pencemaran lingkungan[2], [3].

Dari contoh di atas peneliti ingin mengelompokkan tingkat pencemaran di Jawa Tengah. Oleh karena itu diperlukan analisis pengelompokan (*clustering*) untuk mengetahui kota/kabupaten mana saja yang tingkat pencemarannya tinggi dan rendah. Proses *clustering* kualitas pendidikan dibutuhkan sebuah metode yaitu salah satunya K-Means Clustering. K - Means merupakan metode analisis *clustering* yang memakai data mining pada proses pengelompokannya[4]. Algoritma K - Means Clustering memiliki cara kerja mengelompokkan data berdasarkan jarak data dengan pusat cluster yang didapat dari iterasi

berulang[5]. Adapun kelebihan K-Means *clustering* dengan dibandingkan dengan *Hierarchy Clustering* adalah seperti relatif sederhana, mudah melakukan analisis sampel dalam ukuran besar dengan lebih efektif[6], [7]. Beberapa penelitian terdahulu menggunakan K-Means, seperti penelitian yang dilakukan oleh Andreas, Angraini, Heri, dkk mengelompokkan data kasus COVID-19 berdasarkan provinsi yang menghasilkan nilai *silhouette* 0.8572. Gede dan Agung mengelompokkan data kualitas pembelajaran di sekolah dengan nilai *silhouette* 0.534. Penelitian terdahulu menunjukkan bahwa K-Means menghasilkan struktur *cluster* yang baik.

Dari penyampaian di atas, peneliti bermaksud untuk membuat pengelompokkan kota/kabupaten berdasarkan tingkat pencemaran lingkungan menggunakan metode K-Means. Penelitian ini bertujuan untuk mengelompokkan daerah-daerah di Jawa Tengah mana saja yang tergolong dalam pencemaran lingkungan tinggi dan rendah. Dari hasil *clustering* tersebut maka akan diketahui kota/kabupaten manakah di Jawa Tengah yang harus mendapat perhatian dan prioritas lebih dari dinas lingkungan hidup.

## 2. TINJAUAN PUSTAKA

### 2.1. Silhouette Coefficient

Silhouette Coefficient adalah matriks yang digunakan untuk menghitung baik atau tidaknya sebuah teknik clustering[8]. Kualitas cluster dapat dievaluasi dengan menggunakan rata-rata Silhouette Coefficient dengan range nilai antara -1 dan 1. Menurut tabe Kaufman dan Rousseeuw struktur cluster dikategorikan kuat jika nilai *silhouette* berada pada 0.71 – 1.00. Struktur baik/sesuai jika nilai *silhouette* antara 0,51 dan 0,7. Struktur dikatakan lemah jika nilai *silhouette* antara 0.26 – 0.50. Jika nilai *silhouette* dibawah 0.25 maka tidak terdapat struktur cluster[9].

Berikut adalah persamaan dari Silhouette Coefficient yang ditunjukkan pada persamaan (1):

$$S_i = \frac{b_i - a_i}{\max(b_i, a_i)} \quad (1)$$

Dimana :

$S_i$  = *Silhouette Score*

$b_i$  = Jarak rata-rata minimum antara objek dalam cluster yang berbeda

$a_i$  = Jarak rata-rata minimum antara objek dalam cluster yang sama

### 2.2. K-Mean Cluster

Analisis cluster adalah suatu metode yang digunakan untuk mengklasifikasi objek kedalam kelompok yang karakteristiknya relatif sama. Metode cluster memperlihatkan keunggulan dalam hal konsistensi, keakuratan, dan waktu komputasi[10], [11]. Tujuan utama dari metode pengelompokan K-Means adalah untuk mengurangi jumlah kuadrat jarak antara titik data dan pusat klaster yang sesuai dengan setiap data. [12]. Langkah pertama dalam proses K-Means clustering adalah menetapkan sembarang pusat cluster. Kedua menghitung jarak setiap data pada setiap pusat cluster. Perhitungan jarak menggunakan Euclidean distance ditunjukkan pada persamaan(2):

$$D(x_p, Z_j) = \sqrt{\sum_{i=1}^d (x_p - Z_j)^2} \quad (2)$$

Dimana :

$D$  = *Euclidean Distance*

$d$  = Jumlah atribut dari tiap *cluster*

$Z_j$  = titik pusat *cluster* ke- $j$

$x_p$  = titik data ke- $p$

Suatu data akan menjadi anggota dari suatu cluster yang memiliki jarak terkecil dari pusat cluster. Ketiga, menghitung pusat cluster baru dengan menggunakan persamaan (3) :

$$Z_j = \frac{1}{n_j} \sum \forall x_p \in c_j x_p \quad (3)$$

Dimana :

$Z_j$  = titik pusat cluster ke- $j$

$N_j$  = Jumlah titik data dalam cluster  $k$

$C_j$  = himpunan titik data dalam cluster  $k$

$x_p$  = titik data ke- $p$

Ulangi proses iterasi pada tahap kedua dan ketiga hingga data tidak lagi mengalami perubahan.

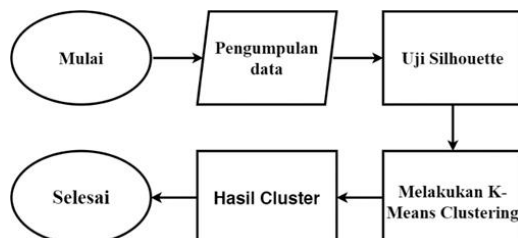
### 3. METODE PENELITIAN

Metode yang dipakai pada penelitian berikut adalah Algoritma K-Means. K-Means merupakan salah satu algoritma clustering non-hirarki yang menggolongkan data dalam konstruksi satu atau lebih cluster. Data yang digunakan dalam penelitian ini adalah jumlah desa yang terdampak pencemaran air, tanah, dan udara di Provinsi Jawa Tengah pada tahun 2021 yang ditunjukkan pada Tabel 1.

Tabel 1. Jumlah Desa yang terdampak Pencemaran di Jawa Tengah

Kota/kabupaten	Pencemaran Air	Pencemaran Tanah	Pencemaran Udara
Cilacap	22	3	15
Banyumas	68	6	39
Purbalingga	45	5	25
...	...	...	...
Kota Tegal	3	0	0

Berikut ini adalah flowchart yang dilakukan untuk mencapai tujuan penelitian :



Gambar 1. Flowchart Penelitian

## 4. HASIL DAN PEMBAHASAN

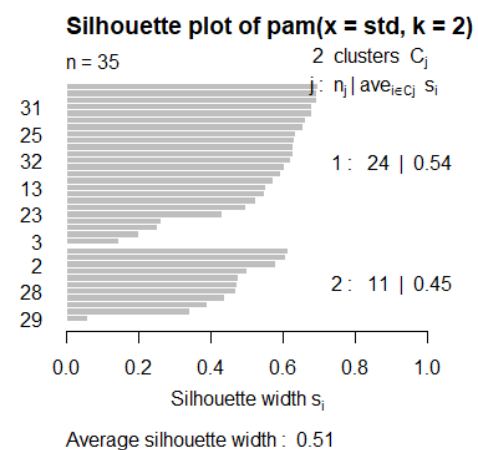
### 4.1. Silhouette Coefficient

Tabel 2. Nilai Silhouette untuk setiap jumlah cluster

	Pencemaran Air	Pencemaran Tanah	Pencemaran Udara
Cluster 1	23.75	3.46	15.38
Cluster 2	67.27	12.82	37.45

Jumlah K- Cluster	Nilai <i>Silhouette</i>
2	0.514
3	0.372
4	0.353
5	0.329

Dari hasil pengujian jumlah k clustering pada Tabel 2, didapat nilai silhouette terbaik pada cluster 2 yakni 0.541. Didasari pada Tabel 1 Kaufman dapat dikatakan bahwa cluster 2 dengan nilai silhouette 0.514 merupakan cluster dengan struktur yang baik.



Gambar 2. Silhouette Coefficient

Dari Gambar 2 terlihat bahwa tidak terdapat nilai negatif, yang artinya semua anggota cluster tepat masuk pada cluster. Pada label X mempresentasikan nilai cluster pada setiap anggota, sedangkan label Y mempresentasikan tiap anggota pada cluster.

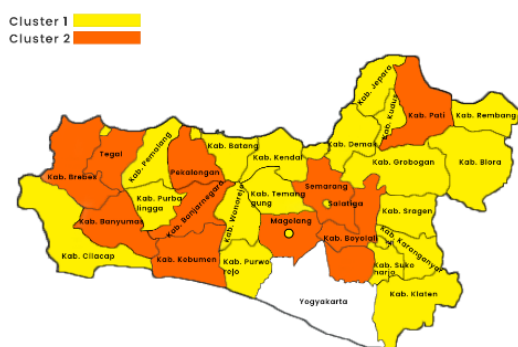
### 4.2. Analisis K-Means Clustering

Tabel 3. Tabel Cluster Center

Tabel 4. Tabel Cluster Membership

Cluster	Anggota
1	Cilacap, Purbalingga, Purworejo, Wonosobo, Kudus, Batang, Karang Anyar, Pemalang, Wonogiri, Kota Semarang, Sragen, Temanggung, Kendal, Kota Pekalongan, Kota Salatiga, Demak, Grobogan, Sukoharjo, Jepara, Kota Magelang, Blora, Kota Surakarta, Rembang, Tegal.
2	Banyumas, Banjarnegara, Kebumen, Kab. Magelang, Boyolali, Klaten, Pati, Kab. Semarang, Kab. Pekalongan, Kab. Tegal, Brebes,

Setelah dilakukan proses clustering didapat nilai final cluster seperti pada Tabel 4. Pada variabel Pencemaran Air, nilai cluster 2 (67.27) > cluster 1 (23.75), hal tersebut berarti bahwa cluster 2 merupakan daerah tingkat pencemaran air yang lebih tinggi dari pada cluster 1. Pada variabel Pencemaran Air, nilai cluster 2 (12.82) > cluster 1 (3.46), hal tersebut berarti bahwa cluster 2 merupakan daerah tingkat pencemaran tanah yang lebih tinggi dari pada cluster 1. Pada variabel Pencemaran Udara, nilai cluster 2 (37.45) > cluster 1 (15.38), hal tersebut berarti bahwa cluster 2 merupakan daerah tingkat pencemaran udara yang lebih tinggi



**Gambar 3.** Gambar 3. Visualisasi K-Means Clustering

Dari Analisa diatas terlihat bahwa cluster 2 merupakan kelompok dengan Kota/Kabupaten dengan tingkat pendemaran tinggi. Anggota cluster dapat dilihat pada tabel 5. yang dimana cluster 1 dengan 24 cluster membership merupakan kota/kabupaten dengan tingkat

pencemaran lingkungan yang rendah. Sedangkan cluster 2 dengan 12 cluster membership merupakan kota/kabupaten dengan tingkat pencemaran tinggi.

Dari hasil visualisasi gambar 2, didapat interpretasi bahwa cluster dengan pencemaran lingkungan yang tinggi masih berkuat di Jawa Tengah bagian barat(Kab. Brebes, Kab. Banyumas, Tegal, dll) dan bagian pusat (Magelang, Semarang, Boyolali, dll).

## 5. KESIMPULAN

- Metode Kmeans dapat digunakan dengan baik pada kota/kabupaten di Jawa Tengah dengan ditunjukkannya hasil plot silhouette yang tidak terdapat nilai yang negative. Hasil uji Silhouette menunjukkan bahwa nilai 0.54 dengan struktur yang baik dengan jumlah cluster 2.
- Cluster 1 (tingkat pencemaran rendah) sebanyak 14 anggota dan Cluster 2 (tingkat pencemaran tinggi) dengan 24 anggota.

## DAFTAR PUSTAKA

- D. L. H. dan Kehutanan, "Kerusakan Lingkungan Akibat Usaha / Kegiatan Pertambangan." [https://dlhk.bantenprov.go.id/upload/article/K](https://dlhk.bantenprov.go.id/upload/article/Kerusakan%20Lingkungan%20Akibat%20Pertambangan) erusakan Lingkungan Akibat Pertambangan.
- B. I. News, "Ganjar Pranowo jadi capres PDIP, rekam jejaknya terkait lingkungan 'masih jauh dari harapan,'" 24 April, 2023. <https://www.bbc.com/indonesia/articles/cn01znvd15no>
- Farisa, "https://nasional.kompas.com/read/2022/02/09/18264541/duduk-perkara-konflik-di-desa-wadas-yang-sebabkan-warga-dikepung-dan," 24 April, 2021. <https://nasional.kompas.com/read/2022/02/09/18264541/duduk-perkara-konflik-di-desa-wadas-yang-sebabkan-warga-dikepung-dan>
- A. BASTIAN, "Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)," *J. Sist. Inf.*, vol. 14, no. 1, pp. 28–34, 2018, doi: 10.21609/jsi.v14i1.566.
- Ediyanto, N. Mara, and N. Satyahadewi, "Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis," *Bul. Ilm. Mat. Stat. dan Ter.*, vol. 02, no. 2, pp. 133–136, 2013.
- M. Misdiyanto, Y. S. T, and I. Aprilia, "Identifikasi Jenis-Jenis Burung Lovebird Menggunakan Pengolahan Citra Digital Dengan Metode K-Means Clustering," *J. Sains*

- Komput. Inform.*, vol. 4, no. 2, pp. 445–456, 2020.
- [7] I. Rahma, P. Prima Arhandi, and A. Tufika Firdausi, “Penerapan Metode Hierarchical Clustering Dan K-Means Clustering Untuk Mengelompokkan Potensi Lokasi Penjualan Linkaja,” *J. Inform. Polinema*, vol. 6, no. 1, pp. 15–22, 2020, doi: 10.33795/jip.v6i1.287.
- [8] S. Paembonan and H. Abduh, “Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat,” *PENA Tek. J. Ilm. Ilmu-Ilmu Tek.*, vol. 6, no. 2, p. 48, 2021, doi: 10.51557/pt\_jiit.v6i2.659.
- [9] A. Supandi, A. Saefuddin, and I. D. Sulvianti, “Two step Cluster Application to Classify Villages in Kabupaten Madiun Based on Village Potential Data,” *Xplore J. Stat.*, vol. 10, no. 1, pp. 12–26, 2020, doi: 10.29244/xplore.v10i1.272.
- [10] H. Sulistiani, K. Muludi, and A. Syarif, “Implementation of Dynamic Mutual Information and Support Vector Machine for Customer Loyalty Classification,” *J. Phys. Conf. Ser.*, vol. 1338, no. 1, 2019, doi: 10.1088/1742-6596/1338/1/012050.
- [11] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, “Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia,” *Eduatic - Sci. J. Informatics Educ.*, vol. 7, no. 1, pp. 1–11, 2020, doi: 10.21107/edutic.v7i1.8779.
- [12] S. Naeem and A. Wumaier, “Study and Implementing K-mean Clustering Algorithm on English Text and Techniques to Find the Optimal Value of K,” *Int. J. Comput. Appl.*, vol. 182, no. 31, pp. 7–14, 2018, doi: 10.5120/ijca2018918234.