https://jurnal.pustakagalerimandiri.co.id/index.php/pustakadata
DOI: https://doi.org/10.55382/jurnalpustakadata.v6i3.1900

Jurnal Pustaka Data

Pusat Akses Kajian Database, Analisa Teknologi, dan Arsitektur Komputer

Vol. 6 No. 3 (2026) 311-318

E ISSN: 2809-7483

# Analisis Kinerja K-Means dan DBSCAN dalam Pengelompokan Kepadatan Kendaraan Bermotor Tingkat Provinsi

Syafiq Hafizh Farizi[1], Muhammad Rasyid[2], Masna Wati[3], Joan Angelina Widians[4]
[1,2,3]Informatika, Universitas Mulawarman
[4]Magister Informatika, Universitas Mulawarman
[1]syafiqhafizh06@gmail.com, [2]mrasyid18102005@gmail.com, [3]masnawati@fkti.unmul.ac.id, [4]angelwidians@unmul.ac.id

***Abstract***

*The growth of motor vehicles in Indonesia has increased significantly, accompanied by a highly uneven distribution across provinces. Data from Statistics Indonesia (BPS) in 2024 recorded an extreme disparity: West Java had 27,104,924 units, while North Kalimantan only had 305,187 units. This necessitates an analytical approach to systematically cluster vehicle density. This study aims to compare the performance of the K-Means and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithms in clustering vehicle density across 34 provinces in Indonesia. The research stages include secondary data collection, preprocessing (data cleaning, logarithmic transformation, and standardization), and optimal parameter determination using the Elbow and Silhouette methods (K-Means) and the k-distance plot (DBSCAN). Performance evaluation utilizes the Silhouette Coefficient, Davies-Bouldin Index, and Calinski-Harabasz Index. The results indicate that K-Means (K=4) generates four interpretable density levels with a Silhouette score of 0.537, a Davies-Bouldin Index of 0.560, and a Calinski-Harabasz Index of 98.00. Conversely, DBSCAN (ε=0.8; MinPts=5) only forms 2 clusters with 5 noise points and a Calinski-Harabasz Index of 42.67. In conclusion, K-Means is proven superior in producing granular and informative cluster separation for categorizing vehicle density levels in Indonesia.*

*Keywords: K-Means, DBSCAN, Clustering, Vehicle Density, Silhouette Coefficient.*

**Abstrak**

Pertumbuhan kendaraan bermotor di Indonesia meningkat signifikan dengan distribusi antarprovinsi yang sangat timpang. Data BPS (2024) mencatat ketimpangan ekstrem: Jawa Barat memiliki 27.104.924 unit, sedangkan Kalimantan Utara hanya 305.187 unit. Hal ini menuntut pendekatan analisis untuk mengelompokkan kepadatan kendaraan secara sistematis. Penelitian ini bertujuan membandingkan kinerja algoritma K-Means dan *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) dalam mengelompokkan kepadatan kendaraan pada 34 provinsi di Indonesia. Tahapan penelitian meliputi pengumpulan data sekunder, *preprocessing* (pembersihan, transformasi logaritmik, dan standardisasi), serta penentuan parameter optimal melalui metode *Elbow* dan *Silhouette* (K-Means) dan *k-distance plot* (DBSCAN). Evaluasi performa menggunakan *Silhouette Coefficient*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Index*. Hasilnya, K-Means (K=4) menghasilkan empat tingkat kepadatan yang *interpretable* dengan nilai *Silhouette* 0,537, *Davies-Bouldin* 0,560, dan *Calinski-Harabasz* 98,00. Sebaliknya, DBSCAN (ε=0,8; MinPts=5) hanya membentuk 2 *cluster* dengan 5 titik *noise* dan *Calinski-Harabasz* 42,67. Kesimpulannya, K-Means terbukti lebih unggul dalam menghasilkan separasi *cluster* yang granular dan informatif untuk pengelompokan tingkat kepadatan kendaraan di Indonesia.

Kata kunci: K-Means, DBSCAN, *Clustering*, Kepadatan Kendaraan, *Silhouette Coefficient.*

Submitted : 11-05-2026 | Reviewed : 14-06-2026 | Accepted : 15-06-2026

© 2026 Author
Creative Commons Attribution 4.0 International License

## 1. Pendahuluan

Pertumbuhan jumlah kendaraan bermotor di Indonesia mengalami peningkatan yang cukup signifikan dari tahun ke tahun, seiring dengan pertumbuhan penduduk dan peningkatan aktivitas ekonomi di berbagai provinsi di Indonesia. Data Badan Pusat Statistik (BPS) tahun 2024 menunjukkan total kendaraan bermotor nasional telah mencapai lebih dari 166 juta unit, dengan persebaran yang bervariasi antarprovinsi. Kepadatan kendaraan bermotor pada suatu wilayah mencerminkan intensitas mobilitas masyarakat dan volume aktivitas transportasi yang berlangsung di daerah tersebut[1].

Berdasarkan data BPS tahun 2024, Provinsi Jawa Barat memiliki jumlah total kendaraan bermotor tertinggi dengan 27.104.924 unit, jauh melampaui Provinsi Kalimantan Utara yang berada pada kisaran 305.187 unit. Ketimpangan distribusi ekstrem tersebut menggambarkan bahwa kepadatan kendaraan antarprovinsi tidak dapat dianalisis secara seragam, sehingga diperlukan pendekatan yang mampu mengelompokkan provinsi berdasarkan profil kendaraannya secara sistematis dan objektif. Salah satu pendekatan yang sesuai adalah teknik data mining, khususnya clustering. Clustering merupakan proses pengelompokan sekumpulan data ke dalam beberapa kelompok berdasarkan kesamaan karakteristik yang dimiliki[2]

Algoritma K-Means merupakan salah satu metode clustering yang paling banyak diterapkan karena efisiensi komputasionalnya. Algoritma ini menggunakan pendekatan berbasis jarak untuk menempatkan setiap titik data ke dalam cluster dengan pusat (centroid) terdekat secara iteratif hingga konvergen[3]. Akan tetapi, K-Means mengharuskan pengguna menentukan jumlah cluster (K) di awal proses serta sensitif terhadap data pencilan (outlier).

Sebagai alternatif, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) merupakan metode clustering berbasis kepadatan yang membangun cluster dari area dengan kepadatan titik yang saling terkoneksi (density connected)[4]. DBSCAN mampu mengidentifikasi jumlah cluster secara otomatis tanpa perlu menetapkan jumlah kelompok di awal, serta memiliki kemampuan menandai titik pencilan sebagai noise. Karakteristik ini menjadikan DBSCAN potensial untuk dataset dengan distribusi yang tidak merata, namun kinerjanya bergantung pada pemilihan parameter Epsilon (ε) dan Minimum Points (MinPts) [5].

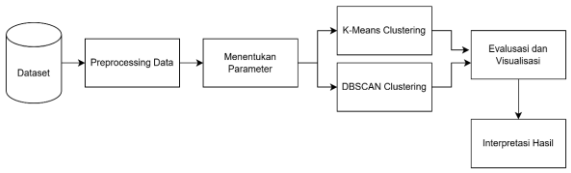
Beberapa penelitian terdahulu telah menerapkan kedua algoritma ini pada konteks yang berbeda. membandingkan K-Means dan Hierarchical Clustering pada data prestasi akademik siswa dan menemukan bahwa pemilihan algoritma sangat dipengaruhi oleh karakteristik distribusi data. Penerapan DBSCAN pada data tuberkulosis dan titik api di Indonesia menunjukkan pentingnya kalibrasi parameter agar hasil cluster informatif[6]. Namun, perbandingan langsung kinerja K-Means dan DBSCAN pada data kepadatan kendaraan tingkat provinsi dengan distribusi skewed yang ekstrem masih relatif terbatas dalam literatur, padahal evaluasi pada konteks tersebut penting untuk merekomendasikan metode clustering yang sesuai bagi kebijakan transportasi.

Berdasarkan latar belakang tersebut, dapat diidentifikasi tiga masalah utama, yaitu: (1) bagaimana pola distribusi kepadatan kendaraan bermotor di tiap provinsi di Indonesia, (2) bagaimana hasil pengelompokan provinsi yang dihasilkan oleh algoritma K-Means dan DBSCAN, serta (3) bagaimana perbandingan kinerja kedua algoritma tersebut berdasarkan metrik evaluasi clustering. Tujuan penelitian ini adalah menganalisis pola kepadatan kendaraan bermotor di tingkat provinsi, menerapkan algoritma K-Means dan DBSCAN untuk mengelompokkan provinsi berdasarkan profil kendaraannya, dan membandingkan kinerja kedua algoritma menggunakan tiga metrik yaitu Silhouette Coefficient, Davies-Bouldin Index, dan Calinski-Harabasz Index. Hasil penelitian diharapkan dapat memberikan rekomendasi metode clustering yang lebih sesuai untuk data spasial dengan distribusi sangat tidak merata.

## 2. Metode Penelitian

Penelitian ini menerapkan pendekatan Data Mining dengan paradigma pembelajaran tak terawasi (unsupervised learning) untuk melakukan pengelompokan provinsi berdasarkan profil kendaraan bermotor. Tahapan penelitian secara keseluruhan ditampilkan pada Gambar 1, yang dimulai dari pengumpulan data, preprocessing, penentuan parameter masing-masing algoritma, proses clustering secara paralel, hingga evaluasi multi-metrik dan analisis komparatif.

Submitted : 11-05-2026 | Reviewed : 14-06-2026 | Accepted : 15-06-2026




Gambar 1. Alur Metode Penelitian

2.1. Pengumpulan Data

Tahap pengumpulan data dilakukan dengan mengakses data statistik resmi dari laman Badan Pusat Statistik (BPS). Dataset yang digunakan adalah data Jumlah Kendaraan Bermotor Menurut Provinsi dan Jenis Kendaraan (Unit) tahun 2024, yang mencakup persebaran kendaraan di seluruh provinsi di Indonesia. Data diunduh dalam format CSV dan terdiri dari empat kategori jenis kendaraan, yaitu Mobil Penumpang, Bus, Truk, dan Sepeda Motor, beserta nilai total agregatnya per provinsi. Dataset dipilih karena mencerminkan ketimpangan mobilitas yang signifikan, sebagaimana terlihat pada perbedaan jumlah kendaraan antara Provinsi Jawa Barat (27.104.924 unit) dan Kalimantan Utara (305.187 unit), yang menjadikannya kasus yang representatif untuk mengevaluasi ketahanan algoritma clustering terhadap distribusi data yang sangat skewed.

2.2. Preprocessing Data

Tahap preprocessing bertujuan mengubah data mentah menjadi data yang siap diolah algoritma clustering [6]. Tahapan yang dilakukan adalah sebagai berikut:

1. Data Cleaning dan Filtering. Baris agregat "Indonesia" dihapus karena merupakan total nasional dan tidak mewakili unit observasi provinsi. Empat baris provinsi pemekaran baru (Papua Barat Daya, Papua Selatan, Papua Tengah, dan Papua Pegunungan) yang seluruh nilainya kosong (null values) juga dihilangkan, sehingga jumlah observasi akhir menjadi 34 provinsi.
2. Penghapusan Kolom Redundan. Kolom "Jumlah" yang merupakan penjumlahan linier dari empat jenis kendaraan dihapus karena menyebabkan multikolinearitas sempurna yang dapat mendistorsi proses clustering.
3. Konversi Tipe Data. Manipulasi string dilakukan untuk menghapus tanda titik sebagai pemisah ribuan, kemudian dilakukan casting dari tipe teks menjadi numerik (integer) agar dapat diproses oleh algoritma berbasis jarak.
4. Transformasi Logaritmik. Karena rentang nilai kendaraan antarprovinsi mencapai dua orde besaran, transformasi log(1+x) diterapkan untuk mengurangi skewness distribusi sehingga setiap fitur memberikan kontribusi yang lebih seimbang dalam perhitungan jarak.
5. Standardisasi. StandardScaler diterapkan untuk menstandarkan setiap fitur menjadi distribusi dengan rerata 0 dan simpangan baku 1, sehingga menghindari dominasi fitur dengan magnitudo lebih besar (Sepeda Motor) terhadap fitur lain (Bus).

2.3. Penentuan Parameter

2.3.1. K-Means

Pada algoritma K-Means, parameter utama yang harus ditentukan adalah jumlah cluster (K). Penentuan K yang optimal bertujuan meminimalkan jarak antara titik data dengan centroid-nya[7], yang secara matematis dievaluasi menggunakan Within-Cluster Sum of Squares (WCSS).

Untuk memilih K optimal, penelitian ini mengombinasikan dua metode. Pertama, Metode Elbow yang mencari titik di mana penurunan WCSS mulai melandai sehingga membentuk siku pada kurva[8]. Kedua, Metode Silhouette yang mencari nilai K dengan rerata Silhouette Coefficient tertinggi [9]. Kombinasi keduanya digunakan agar pemilihan K tidak hanya bergantung pada interpretasi visual subjektif kurva Elbow, tetapi juga divalidasi secara kuantitatif oleh metrik separasi cluster[10].

2.3.2. DBSCAN

Berbeda dengan K-Means, algoritma DBSCAN memerlukan dua parameter utama, yaitu Epsilon (ε) dan Minimum Points (MinPts). Parameter ε mendefinisikan radius lingkungan di sekitar suatu titik data, sedangkan MinPts menentukan jumlah minimum titik yang dibutuhkan untuk membentuk area padat (core point)[11]. Penentuan parameter ε yang tepat sangat krusial agar algoritma dapat memisahkan area padat dari titik noise yang muncul akibat ketimpangan distribusi kendaraan.

Pada penelitian ini, MinPts ditetapkan sebesar 5 dengan mengikuti aturan praktis MinPts >= D + 1 untuk data berdimensi rendah. Pemilihan ε dilakukan secara objektif menggunakan k-distance plot, yaitu mengurutkan jarak setiap titik ke tetangga ke-k terdekatnya dan memilih ε pada titik di mana grafik mulai menanjak tajam (knee point)[12]. Pendekatan ini menghindari pemilihan ε secara coba-coba yang dapat menghasilkan jumlah cluster dan noise yang tidak stabil[13].

2.4 Clustering

Tahap clustering merupakan inti dari proses data mining dalam penelitian ini, dilakukan secara paralel

menggunakan dua algoritma untuk membandingkan efektivitasnya.

K-Means mendefinisikan prototipe dalam bentuk pusat massa (centroid), yang biasanya adalah rerata dari sekelompok titik data, dan umumnya digunakan untuk objek dalam ruang n-dimensi yang kontinu[14]. K-means mampu membagi data ke dalam kelompok homogen berdasarkan kemiripan, sehingga hasil cluster dapat dengan mudah diinterpretasikan[15]. Algoritma menempatkan setiap titik data kemudian memperbarui posisi c

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

dengan a(i) sebagai rerata jarak titik i terhadap seluruh titik di cluster yang sama, dan b(i) sebagai rerata jarak terendah titik i terhadap titik-titik pada cluster terdekat lainnya. Nilai Silhouette Coefficient berada pada rentang –1 hingga 1; nilai yang mendekati 1 menunjukkan cluster yang baik dan terpisah, nilai mendekati 0 menunjukkan titik berada di perbatasan cluster, sedangkan nilai negatif mengindikasikan kesalahan penempatan.

b. Davies-Bouldin Index (DBI). Metrik ini mengukur rerata rasio sebaran intra-cluster terhadap jarak antar-cluster. Semakin kecil nilai DBI, semakin baik kualitas cluster karena menunjukkan cluster yang kompak dan terpisah jauh.

c. Calinski-Harabasz Index (CHI). Metrik ini juga dikenal sebagai Variance Ratio Criterion dan menghitung rasio antara dispersi antar-cluster terhadap dispersi intra-cluster. Nilai CHI yang lebih tinggi menunjukkan cluster yang lebih baik separasinya.

Penggunaan tiga metrik secara bersamaan bertujuan untuk memberikan evaluasi yang komprehensif. Sebuah algoritma dianggap unggul apabila konsisten memperoleh nilai yang lebih baik pada minimal dua dari tiga metrik tersebut.

## 3. Hasil dan Pembahasan

Bab ini memaparkan hasil setiap tahapan penelitian secara berurutan, mulai dari karakteristik dataset hasil preprocessing, penentuan parameter optimal masing-masing algoritma, hasil pengelompokan provinsi, hingga perbandingan kinerja kedua algoritma berdasarkan tiga metrik evaluasi internal clustering.

### 3.1. Karakteristik Dataset

Setelah tahap *data cleaning*, dataset akhir terdiri dari 34 provinsi dengan empat fitur jenis kendaraan, yaitu Mobil Penumpang, Bus, Truk, dan Sepeda Motor. Statistik deskriptif dataset sebelum transformasi ditampilkan pada Tabel 1.

Tabel 1.Statistik Deskriptif Jumlah Kendaraa Bermotor

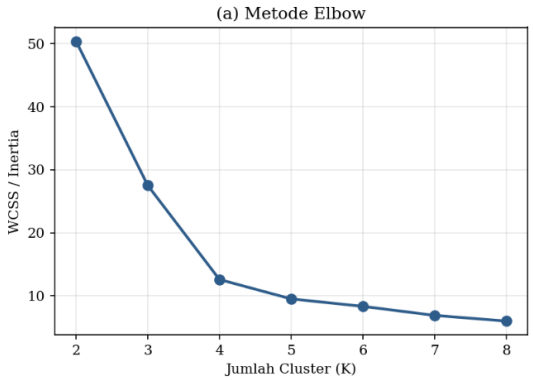
| Statistik | Mobil | Bus | Truk | Motor |
|---|---|---|---|---|
| Mean | 601.309 | 8.647 | 184.630 | 4.101.471 |
| Std. Deviasi | 1.092.636 | 12.800 | 210.605 | 5.670.688 |
| Minimum | 23.544 | 94 | 16.154 | 265.242 |
| Kuartil 1 | 108.897 | 1.154 | 53.890 | 998.028 |
| Median | 213.099 | 3.222 | 92.407 | 2.463.152 |
| Kuartil 3 | 455.668 | 7.423 | 235.931 | 3.974.449 |
| Maksimum | 5.544.750 | 45.193 | 797.226 | 23.348.676 |

Tabel 1 memperlihatkan ketimpangan distribusi yang sangat ekstrem. Pada fitur Sepeda Motor, rasio antara nilai maksimum dan minimum mencapai 88 kali lipat, sedangkan pada Mobil Penumpang mencapai 235 kali lipat. Nilai simpangan baku yang lebih besar dari rerata pada seluruh fitur juga menunjukkan distribusi yang sangat right-skewed. Kondisi ini menegaskan kebutuhan transformasi logaritmik sebelum proses standardisasi, sebagaimana telah dijelaskan pada bagian 2.2.

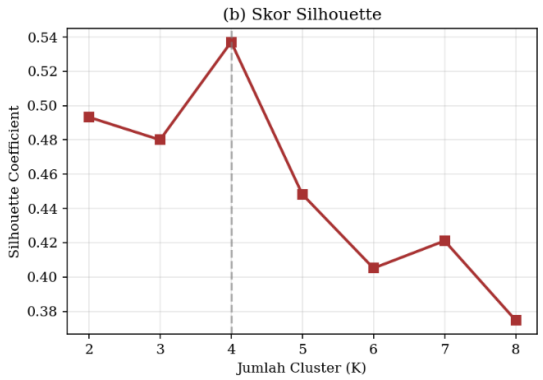
### 3.2. Penentuan Parameter Optimal

#### 3.2.1. K-Means

Penentuan jumlah cluster (K) optimal dilakukan dengan mengevaluasi nilai K = 2 sampai 8 menggunakan kombinasi metode Elbow dan Silhouette. Hasil pengujian disajikan pada Tabel 2 dan divisualisasikan pada Gambar 2.




Gambar 2. Kurva Kurva Elbow

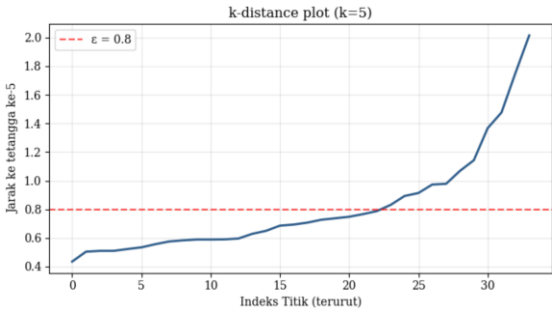



Gambar 3. Skor Silhouette pada K-Means

Berdasarkan Gambar 2(a), kurva Elbow menunjukkan penurunan WCSS yang signifikan dari K=2 hingga K=4, kemudian melandai pada K=5 dan seterusnya. Titik siku terlihat jelas pada K=4. Hasil ini dikonfirmasi oleh kurva *Silhouette* pada Gambar 2(b), di mana skor tertinggi sebesar 0,5371 juga dicapai pada K=4. Kesepakatan kedua metode tersebut memberikan justifikasi yang kuat untuk memilih K=4 sebagai jumlah cluster optimal. Selain itu, K=4 juga menghasilkan nilai *Davies-Bouldin* Index terendah (0,5600) dan *Calinski-Harabasz* Index tertinggi (98,00) di antara seluruh nilai K yang diuji, sehingga keempat metrik bersama-sama mengonfirmasi pemilihan ini.

3.2.2. DBSCAN

Penentuan parameter ε pada DBSCAN dilakukan menggunakan k-distance plot dengan MinPts = 5. Grafik jarak terurut titik ke tetangga ke-4 ditampilkan pada Gambar 3, dan hasil pengujian beberapa nilai ε ditampilkan pada Tabel 3.

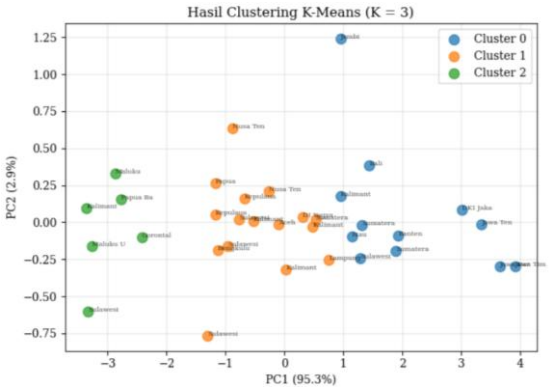



Gambar 4. k-distance plot untuk Penentuan ε pada DBSCAN (k=4)

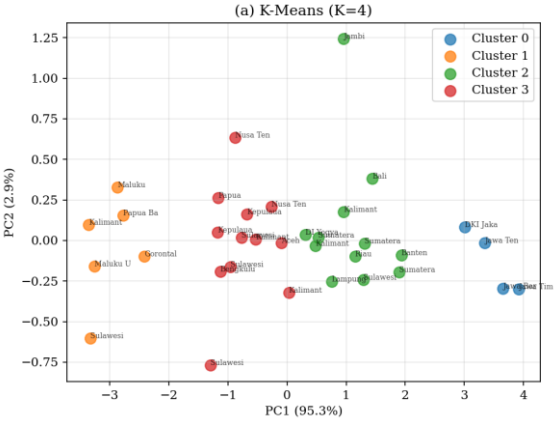
Berdasarkan *k-distance plot*, *knee point* teridentifikasi pada nilai ε ≈ 0,8. Nilai ini juga memberikan *Silhouette Coefficient* tertinggi sebesar 0,5618 di antara seluruh ε yang diuji. Tabel 3 memperlihatkan sensitivitas DBSCAN yang cukup tinggi terhadap perubahan ε. Pada ε=0,60, sebanyak 16 dari 34 provinsi (47%) ditandai sebagai *noise*, yang berarti hampir separuh data diabaikan dari analisis. Sebaliknya, pada ε ≥ 1,40, seluruh data terkumpul ke dalam satu cluster sehingga metrik evaluasi tidak dapat dihitung. Rentang ε yang menghasilkan *cluster* bermakna sangat sempit, yaitu hanya antara 0,8 hingga 1,2. Karakteristik ini mengindikasikan bahwa distribusi data kepadatan kendaraan provinsi cenderung kontinu tanpa pemisah kepadatan yang tegas, sehingga kurang sesuai dengan asumsi DBSCAN yang mensyaratkan adanya area kepadatan terkoneksi yang jelas berbeda dari *background*.

3.3. Hasil Pengelompokan

Hasil clustering kedua algoritma divisualisasikan pada ruang dua dimensi menggunakan reduksi dimensi Principal Component Analysis (PCA) sebagaimana ditampilkan pada Gambar 4. Komponen utama pertama (PC1) menjelaskan 95,3% variansi data, sedangkan komponen kedua (PC2) menjelaskan 2,9%, sehingga total kumulatif variansi yang dijelaskan oleh dua komponen ini mencapai 98,2%.

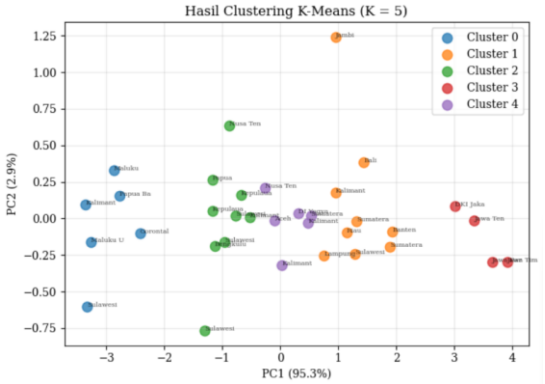



Gambar 5. Alur Visualisasi Hasil *Clustering* pada Ruang PCA Dua Dimensi: K-Means dengan K=3

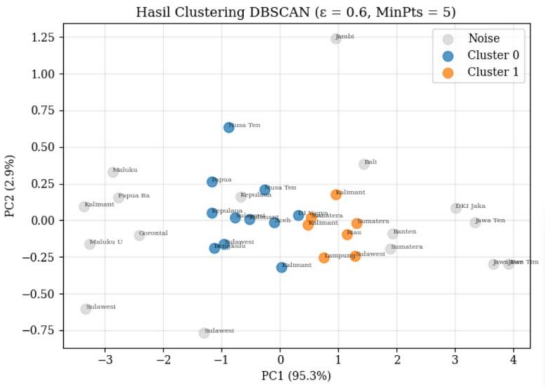



Gambar 6. Alur Visualisasi Hasil *Clustering* pada Ruang PCA Dua Dimensi: K-Means dengan K=4

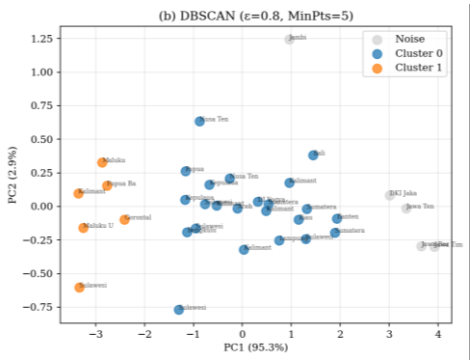
Submitted : 11-05-2026 | Reviewed : 14-06-2026 | Accepted : 15-06-2026



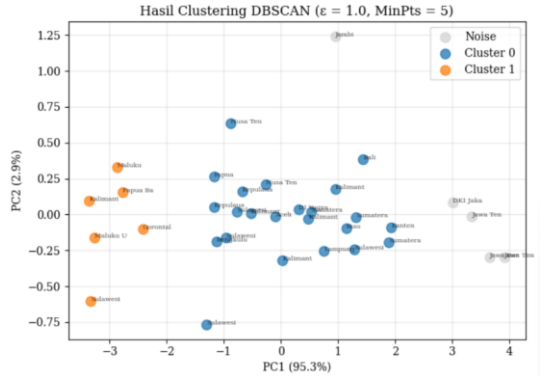

Gambar 7. Alur Visualisasi Hasil *Clustering* pada Ruang PCA Dua Dimensi: K-Means dengan K=5




Gambar 8. Visualisasi Hasil *Clustering* pada Ruang PCA Dua Dimensi: DBSCAN dengan ε=0,6 dan MinPts=5 *(figure caption)*




Gambar 9. Visualisasi Hasil *Clustering* pada Ruang PCA Dua Dimensi: DBSCAN dengan ε=0,8 dan MinPts=5 *(figure caption)*




Gambar 10. Visualisasi Hasil *Clustering* pada Ruang PCA Dua Dimensi: DBSCAN dengan ε=0,8 dan MinPts=5 *(figure caption)*

Tabel 2. Tabel Perbandingan Pelatihan dan Pengujian

| Inertia (WCSS) | Silhouette | Davies-Bouldin | Calinski-Harabasz |
|---|---|---|---|
| 50,400 | 0,4933 | 0,6881 | 54,35 |
| 27,525 | 0,4800 | 0,6362 | 61,08 |
| 12,593 | 0,5371 | 0,5600 | 98,00 |
| 9,522 | 0,4483 | 0,6967 | 96,30 |
| 8,335 | 0,4052 | 0,8096 | 85,77 |
| 6,902 | 0,4211 | 0,5982 | 84,17 |
| 5,999 | 0,3747 | 0,6998 | 80,49 |

Tabel 3. Hasil Pemindaian Nilai ε pada Algoritma DBSCAN (MinPts = 5)

| ε | Jumlah Cluster | Noise | Silhouette | DBI | Calinzki |
|---|---|---|---|---|---|
| 0,60 | 2 | 16 | 0,5011 | 0,6254 | 30,44 |
| 0,80 | 2 | 5 | 0,5618 | 0,4728 | 42,67 |
| 1,00 | 3 | 1 | 0,4891 | 0,4628 | 50,60 |
| 1,20 | 2 | 0 | 0,4792 | 0,4965 | 29,81 |
| 1,40 | 1 | 0 | – | – | – |
| 1,60 | 1 | 0 | – | – | – |
| 1,80 | 1 | 0 | – | – | – |

### 3.3.1. Profile Cluster K-Means

Algoritma K-Means menghasilkan empat cluster yang merepresentasikan tingkat kepadatan kendaraan secara berjenjang. Profil rerata fitur pada setiap cluster beserta anggotanya ditampilkan pada Tabel 4.

Tabel 4. Profil Cluster K-Means (Rerata Jumlah Kendaraan per Cluster)

| Cluster | Mobil | Bus | Truk | Motor | n |
|---|---|---|---|---|---|
| 0 (Sangat Tinggi) | 3.131.778 | 38.530 | 689.717 | 17.810.794 | 4 |
| 2 (Tinggi-Menengah) | 494.496 | 9.611 | 211.624 | 4.019.943 | 12 |
| 3 (Menengah-Rendah) | 149.284 | 1.890 | 71.534 | 1.476.854 | 12 |
| 1 (Rendah) | 32.007 | 312 | 20.107 | 374.212 | 6 |

Hasil pengelompokan pada Tabel 4 sangat interpretable dan sesuai dengan realitas geografis ekonomi Indonesia. Cluster 0 yang menghasilkan tingkat kepadatan paling tinggi seluruhnya dihuni oleh empat provinsi di Pulau Jawa yang merupakan pusat aktivitas ekonomi nasional. Cluster 2 didominasi oleh provinsi dengan kapasitas ekonomi menengah-besar di Sumatera, Bali, dan beberapa wilayah Kalimantan serta Sulawesi. Cluster 3 berisi provinsi pendukung di luar Jawa dengan profil mobilitas sedang. Sementara itu, Cluster 1 mengelompokkan provinsi terluar dan provinsi pemekaran dengan jumlah penduduk dan aktivitas ekonomi yang relatif rendah. Pola berjenjang yang konsisten pada keempat fitur (rerata Mobil, Bus, Truk, dan Motor menurun secara monotonik dari Cluster 0 ke Cluster 1) menunjukkan bahwa K-Means berhasil menangkap struktur hierarki kepadatan secara baik.

3.3.2. Profile Cluster DBSAN

Algoritma DBSCAN dengan parameter optimal ε=0,8 dan MinPts=5 hanya menghasilkan dua cluster dengan 5 titik noise. Kelima provinsi yang ditandai sebagai noise adalah DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, dan Banten. Provinsi-provinsi ini secara konsisten merupakan provinsi dengan kepadatan kendaraan tertinggi sehingga terpisah jauh dari pusat kepadatan data lainnya pada ruang fitur.

Mayoritas provinsi (24 provinsi) tergabung dalam Cluster 0, sedangkan Cluster 1 hanya terdiri dari lima provinsi dengan kepadatan terendah (Maluku, Maluku Utara, Papua Barat, Gorontalo, dan Sulawesi Barat). Granularitas yang rendah ini menyebabkan DBSCAN gagal membedakan provinsi tingkat menengah dari provinsi tingkat menengah-tinggi, yang merupakan informasi penting bagi pengambil kebijakan transportasi untuk perencanaan tipologi yang berjenjang.2. Profile Cluster DBSCAN

Tabel 5. Profil Cluster DBSCAN (Rerata Jumlah Kendaraan per Cluster)

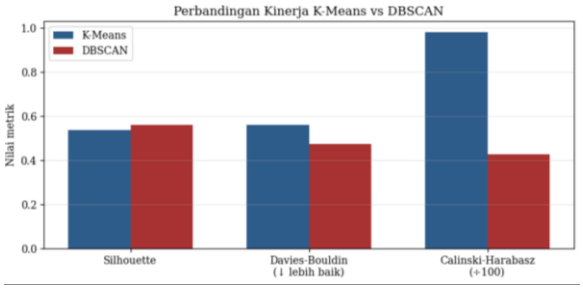
| Cluster / Label | Karakteristik | n | Anggota Provinsi |
|---|---|---|---|
| Noise (-1) | Kepadatan sangat tinggi yang terpisah jauh dari pusat data (*outlier*). | 5 | DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Banten. |
| Cluster 0 | Kepadatan menengah hingga cukup tinggi (membentuk satu kepadatan besar). | 24 | Mayoritas provinsi di luar anggota Noise dan Cluster 1 (24 provinsi). |
| Cluster 1 | Kepadatan paling rendah. | 5 | Maluku, Maluku Utara, Papua Barat, Gorontalo, Sulawesi Barat. |

3.4 Perbandingan Kinerja Algoritma

Ringkasan perbandingan kinerja kedua algoritma pada parameter optimal masing-masing disajikan pada Tabel 6 dan divisualisasikan pada Gambar 5.

Tabel 6.Tabel Perbandingan Pelatihan dan Pengujian

| Aspek | K-Means | DBSCAN |
|---|---|---|
| Parameter optimal | K = 4 | ε = 0,8; MinPts = 5 |
| Jumlah *cluster* | 4 | 2 |
| Titik *noise* | 0 | 5 |
| Silhouette Coefficient (↑) | 0,5371 | 0,5618 |
| Davies-Bouldin Index (↓) | 0,5600 | 0,4728 |
| Calinski-Harabasz Index (↑) | 98,00 | 42,67 |
| Granularitas pengelompokan | Tinggi (4 tingkat) | Rendah (2 tingkat) |
| *Interpretability* | Tinggi | Sedang |




Gambar 10. Perbandingan Kinerja K-Means dan DBSCAN Berdasarkan Tiga Metrik

Berdasarkan Tabel 5, hasil evaluasi multi-metrik menunjukkan keunggulan yang berbeda antara kedua algoritma. DBSCAN sedikit unggul pada Silhouette Coefficient (0,5618 vs 0,5371) dan Davies-Bouldin Index (0,4728 vs 0,5600). Akan tetapi, kedua metrik tersebut dihitung hanya pada 29 titik non-noise, sehingga keunggulan numerik DBSCAN harus diinterpretasikan dengan kehati-hatian karena sebagian data telah dieliminasi dari perhitungan.

Pada Calinski-Harabasz Index, K-Means menunjukkan keunggulan yang sangat signifikan dengan nilai 98,00 dibandingkan DBSCAN yang hanya 42,67. Selisih sekitar 2,3 kali lipat ini mengindikasikan bahwa K-Means menghasilkan separasi antar cluster yang jauh lebih jelas relatif terhadap dispersi internal cluster. CHI dianggap sebagai metrik yang sensitif terhadap kualitas separasi global cluster, sehingga keunggulan K-Means pada metrik ini menjadi argumen kuat untuk superioritasnya pada konteks penelitian ini.

3.5 Diskusi

Hasil penelitian menunjukkan bahwa pemilihan clustering yang tepat sangat dipengaruhi oleh karakteristik distribusi data, sebagaimana juga ditemukan oleh Sampurno dkk. [18] pada konteks pengelompokan prestasi akademik. Pada dataset kepadatan kendaraan provinsi, terdapat tiga temuan utama yang patut diperhatikan.

Pertama, K-Means terbukti lebih unggul untuk dataset ini meskipun Silhouette Coefficient DBSCAN sedikit lebih tinggi. Keunggulan ini disebabkan oleh kemampuan K-Means menghasilkan empat tingkat kepadatan berjenjang yang interpretable, sementara DBSCAN hanya menghasilkan dua cluster dengan 5 noise.

Pengelompokan berjenjang K-Means lebih bermanfaat bagi pengambil kebijakan transportasi yang membutuhkan tipologi multi-tingkat untuk perencanaan infrastruktur dan distribusi anggaran.

Kedua, kelemahan DBSCAN pada konteks ini bersumber dari distribusi data yang relatif kontinu pada ruang fitur setelah transformasi logaritmik. Schubert dkk. [16] mencatat bahwa DBSCAN bekerja optimal ketika data membentuk area padat yang terisolasi oleh ruang berkepadatan rendah, dan kurang sesuai untuk data dengan gradient kepadatan yang halus. Sensitivitas DBSCAN terhadap parameter ε—di mana rentang ε yang menghasilkan cluster bermakna hanya 0,8–1,2—menegaskan keterbatasan ini.

Ketiga, lima provinsi yang ditandai sebagai noise oleh DBSCAN (DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Banten) tepatnya adalah provinsi-provinsi yang oleh K-Means dikelompokkan sebagai Cluster 0 (Sangat Tinggi) dan sebagian Cluster 2. Hal ini menunjukkan bahwa DBSCAN tidak memandang provinsi-provinsi tersebut sebagai kelompok dengan profil tersendiri, melainkan sebagai pencilan. Dari perspektif analisis kebijakan, perlakuan ini tidak diinginkan karena justru provinsi tersebut yang paling membutuhkan perhatian khusus dalam perencanaan transportasi

Dengan demikian, penelitian pengelompokan ini direkomendasikan menggunakan metode K-Means sebagai metode yang lebih sesuai dibandingkan DBSCAN. Temuan ini melengkapi penelitian [4], [12] dengan menyediakan bukti komparatif yang spesifik untuk data spasial-administratif dengan distribusi skewed yang ekstrem.

## 4. Kesimpulan

Penelitian ini berhasil menerapkan dan membandingkan kinerja algoritma K-Means dan DBSCAN dalam pengelompokan kepadatan kendaraan bermotor pada 34 provinsi di Indonesia menggunakan data BPS tahun 2024. Berdasarkan pengujian dan analisis yang dilakukan, dapat ditarik tiga kesimpulan utama. Pertama, kombinasi metode Elbow dan Silhouette berhasil menetapkan jumlah cluster optimal pada algoritma K-Means sebesar K = 4, sedangkan k-distance plot menetapkan parameter optimal DBSCAN pada ε = 0,8 dengan MinPts = 5. Kedua, K-Means menghasilkan empat tingkat kepadatan yang interpretable, yaitu Sangat Tinggi (4 provinsi Pulau Jawa), Tinggi-Menengah (12 provinsi), Menengah-Rendah (12 provinsi), dan Rendah (6 provinsi terluar), sedangkan DBSCAN hanya menghasilkan 2 cluster dengan 5 titik noise. Ketiga, berdasarkan evaluasi multi-metrik, K-Means lebih unggul dengan Calinski-Harabasz Index sebesar 98,00 yang 2,3 kali lebih besar dari DBSCAN (42,67), meskipun DBSCAN sedikit unggul pada Silhouette (0,5618 vs 0,5371) dan Davies-Bouldin (0,4728 vs 0,5600) yang dihitung pada subset data non-noise. K-Means direkomendasikan sebagai metode yang lebih sesuai untuk pengelompokan data kepadatan kendaraan tingkat provinsi karena menghasilkan struktur hierarki yang jelas dan interpretable.

Penelitian ini memiliki keterbatasan pada penggunaan data tahun 2024 saja, sehingga belum menggambarkan dinamika temporal kepadatan kendaraan. Untuk pengembangan selanjutnya, dapat dilakukan analisis time series clustering dengan data multi-tahun untuk melihat evolusi tipologi kepadatan provinsi, penambahan variabel pendukung seperti luas wilayah, jumlah penduduk, serta penerapan algoritma clustering alternatif seperti Hierarchical Clustering atau Gaussian Mixture Model untuk perbandingan yang lebih komprehensif.

## Ucapan Terimakasih

Penulis mengucapkan terima kasih kepada Badan Pusat Statistik (BPS) Republik Indonesia yang telah menyediakan data pada penelitian ini secara terbuka, serta kepada Program Studi Informatika Fakultas Teknik Universitas Mulawarman atas dukungan akademik selama penelitian ini berlangsug.

## Daftar Rujukan

[1] M. G. Pongilatan, S. Pangerapan, and S. Mintalangi, "Pengaruh kepadatan kendaraan bermotor dan Produk Domestik Regional Bruto (PDRB) sektor transportasi terhadap pajak kendaraan bermotor Provinsi Sulawesi Utara tahun 2020-2024," Ris. Akunt. dan Portofolio Investasi, vol. 3, no. 2, pp. 378–386, 2025, doi: 10.58784/rapi.348.

[2] T. Lidia Putri and R. Danar Dana, "Penerapan Data Mining Pada Clustering Data Harga Rumah Dki Jakarta Menggunakan Algoritmak-Means," JATI (Jurnal Mhs. Tek. Inform., vol. 8, no. 1, pp. 1174–1179, 2024, doi: 10.36040/jati.v8i1.8957.

[3] A. Rahmadani and Nursyahira, "IJIRSE: Indonesian Journal of Informatic Research and Software Engineering Implementation of the K-Means Algorithm for Inventory Data Clustering Implementasi Algoritma K-Means Untuk Clustering Data Inventori," vol. 5, no. 1, pp. 1–11, 2025.

[4] F. Supriadi, "Penerapan Metode Dbscan Dalam Penentuan Mahasiswa Yang Layak Memperoleh Bidikmisi," Manag. Inf. Syst. J., vol. 3, no. 2, pp. 46–60, 2025, doi: 10.47065/mis.v3i2.1943.

[5] M. Y. Zidane, B. Nurina Sari, I. Maulana, A. Primaya, and G. Garno, "Penerapan Data Mining Dalam Klasifikasi Data Transaksi Produk Koperasi Di Smk Pgri 2 Karawang," JATI (Jurnal Mhs. Tek. Inform., vol. 9, no. 1, pp. 263–269, 2024, doi: 10.36040/jati.v9i1.12196.

[6] B. Driyandita, I. P. E. N. Kencana, and I. G. N. L. Wijayakusuma, "Analisis Pemilihan Parameter pada Algoritma DBSCAN untuk Pengelompokan Titik Api di Indonesia," J. Pendidik. dan Teknol. Indones., vol. 5, no. 3, pp. 793–803, 2025, doi: 10.52436/1.jpti.703.

[7] W. Mega, "Prinsip dan Cara Kerja K-Means," J. Inform., vol. 15, no. 2, pp. 160–174, 2015.

[8] F. Fauji and L. Farokhah, "Penerapan Data Mining Menggunakan K-Means Clustering Dalam Mengelompokkan Tingkat Kesulitan Mata Pelajaran," J.

Inf. Syst. Res., vol. 6, no. 3, pp. 1705–1714, 2025, doi: 10.47065/josh.v6i3.6959.

[9] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," J. Comput. Appl. Math., vol. 20, no. C, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.

[10] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," Proc. Annu. ACM-SIAM Symp. Discret. Algorithms, vol. 07-09-Janu, pp. 1027–1035, 2007.

[11] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN," ACM Trans. Database Syst., vol. 42, no. 3, 2017, doi: 10.1145/3068335.

[12] M. A. Idris, A. Apriyanto, and Rahmawati, "Pemetaan Produksi Perikanan Tangkap di Indonesia dengan Menggunakan Metode DBSCAN," J. Math. Theory Appl., vol. 5, no. 2, pp. 80–86, 2023, doi: 10.31605/jomta.v5i2.2930.

[13] S. Vijayalaksmi and M. Punithavalli, "A Fast Approach to Clustering Datasets using DBSCAN and Pruning Algorithms," Int. J. Comput. Appl., vol. 60, no. 14, pp. 1–7, 2012, doi: 10.5120/9757-8924.

[14] A. Khalif, A. N. Hasanah, M. H. Ridwan, and B. N. Sari, "Klasterisasi Tingkat Kemiskinan di Indonesia menggunakan Algoritma K-Means," Gener. J., vol. 8, no. 1, pp. 54–62, 2024, doi: 10.29407/gj.v8i1.21470.

[15] A. N. Fauzan, F. A. Wandi, A. Z. N. Aiman, M. Wati, and H. Haviluddin, "Pengelompokan Minat Akademik Siswa SMA Negeri 1 Loa Janan Menggunakan Metode Clustering K-means," J. Rekayasa Teknol. Inf., vol. 9, no. 2, p. 165, 2025, doi: 10.30872/jurti.v9i2.19673.

Submitted : 11-05-2026 | Reviewed : 14-06-2026 | Accepted : 15-06-2026