



Penerapan Teknik Clustering untuk Analisis Pemerataan Angkutan Umum di Kabupaten Bogor

Emir Muhammad Al-Fariq^{1*}, dan Safrizal²

¹ Universitas Pembangunan Jaya; Tangerang Selatan, Banten; e-mail : emir.muhammadal-fariq@student.upj.ac.id

² Universitas Pembangunan Jaya; Tangerang Selatan, Banten; e-mail : safrizal.abdurrahman@upj.ac.id

* Corresponding Author : Emir Muhammad Al-Fariq

Abstract: This research analyzes the distribution of public transport services from AKDP and AKDK units in 40 sub-districts in Bogor Regency, an area with rapid population growth that faces congestion due to dependence on private vehicles. The main problem is that service gaps have not been mapped quantitatively, so the aim of this research is to apply clustering techniques to identify underserved areas and provide an objective basis for data-based policy recommendations. The proposed method is a quantitative approach using the K-Means Clustering algorithm on secondary datasets from the Department of Transportation with three main features: number of licensed fleets, number of unlicensed fleets, and total fleet. The main findings succeeded in classifying sub-districts into three clusters, Cluster 0 (established services), Cluster 1 (static growth), and Cluster 2 (significant service gaps dominated by informal fleets). In conclusion, this research proves the existence of significant service disparities, and shows that the clustering method is effective in mapping priority zones for more equitable transportation policy interventions.

Keywords: Bogor Regency; Clustering; Data Mining; K-Means; Public Transportation; Service Equity

Abstrak: Penelitian ini menganalisis pemerataan layanan angkutan umum unit AKDP dan AKDK di 40 kecamatan di Kabupaten Bogor, wilayah dengan pertumbuhan penduduk pesat yang menghadapi kemacetan akibat ketergantungan pada kendaraan pribadi. Permasalahan utama adalah belum terpecahnya kesenjangan layanan secara kuantitatif, sehingga tujuan penelitian ini adalah menerapkan teknik clustering untuk mengidentifikasi wilayah yang kurang terlayani dan menyediakan dasar objektif untuk rekomendasi kebijakan berbasis data. Metode yang diusulkan adalah pendekatan kuantitatif menggunakan algoritma *K-Means Clustering* terhadap dataset sekunder dari Dinas Perhubungan dengan tiga fitur utama : jumlah armada berizin, jumlah armada tidak berizin, dan total armada. Temuan utama berhasil mengklasifikasikan kecamatan ke dalam tiga *cluster*, *Cluster 0* (layanan mapan), *Cluster 1* (pertumbuhan statis), dan *Cluster 2* (kesenjangan layanan signifikan yang didominasi armada informal). Kesimpulannya, penelitian ini membuktikan adanya disparitas layanan yang signifikan, serta menunjukkan bahwa metode clustering efektif dalam memetakan zona prioritas untuk intervensi kebijakan transporasi yang lebih merata.

Kata kunci: Clustering; Data Mining; K-Means; Kabupaten Bogor; Pemerataan Layanan; Transportasi Publik

Received: September 19, 2025

Revised: October 3, 2025

Accepted: November 26, 2025

Published: November 29, 2025

Curr. Ver.: November 29, 2025



Copyright: © 2025 by the authors.

Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>)

1. Pendahuluan

Kabupaten Bogor, yang terletak di Provinsi Jawa Barat, mengalami pertumbuhan jumlah penduduk yang pesat [1]. Peningkatan populasi ini berkontribusi pada meningkatnya kebutuhan akan sistem transportasi yang lebih efektif dan efisien. Salah satu solusi yang diusulkan oleh pemerintah Kabupaten Bogor adalah pengoperasian angkutan umum massal yang diharapkan dapat meningkatkan daya tarik angkutan umum dan menekan penggunaan kendaraan pribadi [2]. Hal ini bertujuan untuk mengurangi ketergantungan pada kendaraan

pribadi, yang sering kali dipilih masyarakat karena dianggap lebih efektif dan efisien [3], sehingga pada akhirnya dapat mengurangi kemacetan lalu lintas di pusat kota[4].

Penerapan angkutan umum memiliki permasalahan mengenai pemerataan distribusi di seluruh wilayah. Untuk mengatasi hal ini, diperlukan suatu pendekatan berbasis data mining, yang merupakan metode untuk memecahkan masalah dengan menganalisis data yang telah ada [5]. Salah satu teknik yang dapat digunakan adalah klasterisasi (*clustering*), yang bertujuan untuk mengelompokkan sejumlah data atau objek ke dalam sebuah cluster sehingga dalam setiap cluster akan berisi data yang semirip mungkin [6].

Teknik klasterisasi memungkinkan untuk mengelompokkan area berdasarkan karakteristik tertentu seperti kepadatan penduduk, tingkat arus lalu lintas, dan fitur jalan lainnya [7]. Dengan menggunakan teknik ini, dapat dihasilkan pemetaan yang lebih akurat mengenai distribusi angkutan umum yang ada dan area yang membutuhkan intervensi lebih lanjut. Informasi yang terkumpul dalam klaster-klaster tersebut sangat membantu bagi para pengambil kebijakan dalam proses pengambilan keputusan [8].

Beberapa penelitian sebelumnya menunjukkan bahwa meskipun angkutan umum telah diterapkan, efektivitasnya dalam mengurangi kemacetan sering kali belum sepenuhnya optimal [9]. Beberapa faktor yang memengaruhinya antara lain aksesibilitas rute yang terbatas sehingga tidak dapat menjangkau semua wilayah, lokasi halte dan shelter yang belum ideal, serta sosialisasi yang belum optimal kepada masyarakat [9]. Dalam hal ini, analisis klasterisasi dapat digunakan untuk menilai dan merancang ulang distribusi angkutan umum yang lebih efisien untuk memenuhi kebutuhan masyarakat.

Lebih jauh lagi, pemanfaatan teknologi *Big Data* dalam analisis klasterisasi dapat membuka peluang baru bagi pemerintah untuk menganalisis volume data yang besar secara efisien, yang pada gilirannya mendukung pengambilan keputusan berbasis data (*data-driven decision making*)[10]. Dengan data yang lebih akurat mengenai pola perjalanan dan preferensi masyarakat, pemerintah dapat merancang sistem transportasi yang lebih merata dan dapat diakses oleh seluruh lapisan masyarakat.

Penelitian ini bertujuan untuk menerapkan teknik klasterisasi dalam menganalisis pemerataan angkutan umum di Kabupaten Bogor. Dengan menggunakan pendekatan berbasis data, diharapkan penelitian ini dapat memberikan kontribusi dalam merancang kebijakan transportasi yang lebih baik dan lebih inklusif, yang pada akhirnya dapat meningkatkan kualitas hidup masyarakat dan mengurangi masalah kemacetan di Kabupaten Bogor.

Batasan Penelitian

- a. Batasan : Data yang digunakan hanya terbatas pada jumlah angkutan umum unit AKDP dan AKDK berizin maupun tidak berizin. Data ini belum mencakup variabel sisi permintaan seperti jumlah penumpang, frekuensi layanan, atau waktu tempuh , serta belum memasukkan pengaruh transportasi lain (*commuter line*, bus, dan layanan berbasis aplikasi).
- b. Justifikasi : Pembatasan ini bertujuan untuk mengisi kesenjangan literatur dengan memberikan fokus yang kuat pada analisis sisi penawaran dan legalitas armada. Hal ini bertujuan menghasilkan pemetaan kuantitatif atas disparitas penawaran di wilayah studi, yang merupakan target intervensi kebijakan yang paling mendesak.

2. Kajian Pustaka atau Penelitian Terkait

Perencanaan sistem transportasi perkotaan dan wilayah penyanga telah bertransformasi secara signifikan dengan adanya pemanfaatan analisis data. Teknik *data mining*, khususnya metode *unsupervised learning* seperti *clustering*, telah menjadi instrumen penting untuk mengidentifikasi pola tersembunyi dalam data transportasi yang kompleks [11]. Analisis *clustering* memungkinkan para perencana dan pembuat kebijakan untuk mengelompokkan wilayah berdasarkan karakteristik transportasi yang serupa, sehingga intervensi dapat dirancang secara lebih efektif dan tepat sasaran [12]. Pendekatan ini merupakan bagian dari implementasi *Business Intelligence* (BI) yang lebih luas, di mana data dari berbagai sumber diintegrasikan dan diolah

melalui *data warehouse* untuk mendukung pengambilan keputusan strategis dalam sebuah organisasi atau instansi pemerintah [13].

Pemanfaatan teknologi *big data* untuk memproses dan menganalisis data dalam jumlah besar, bahkan data tidak terstruktur sekalipun, terbukti dapat menghasilkan model dan Indikator Kinerja Utama (*Key Performance Indicators*) yang lebih akurat untuk mendukung pengambilan keputusan [14]. Dalam konteks clustering, salah satu langkah metodologis yang krusial adalah penentuan jumlah klaster (nilai K) yang optimal. Penggunaan teknik validasi seperti *Silhouette Score* dapat membantu memastikan bahwa hasil pengelompokan data. Misalnya, segmentasi pelanggan mal atau pola transportasi benar-benar valid dan menghasilkan klaster yang bermakna, sehingga kualitas analisis dapat dipertanggungjawabkan [15].

2.1. Penerapan Clustering dalam Analisis Transportasi Publik

Algoritma *clustering*, terutama K-Means, telah banyak diterapkan dalam berbagai studi transportasi publik. K-Means merupakan salah satu metode *clustering* non-hierarki yang bekerja dengan membagi data ke dalam beberapa kelompok (*cluster*) berdasarkan kesamaan karakteristiknya [16]. Metode ini efektif karena kemudahannya dalam implementasi dan kemampuannya untuk mencapai konvergensi dengan cepat [17].

Salah satu penerapan utamanya adalah untuk menganalisis pola pergerakan penumpang. Penelitian yang dilakukan oleh Reynaldi dkk. (2025) menggunakan K-Means untuk menganalisis 189.501 data transaksi pengguna Transjakarta. Dengan memanfaatkan informasi lokasi *tap-in* dan *tap-out* serta waktu perjalanan, penelitian tersebut berhasil mengidentifikasi pola penyebaran spasial dan temporal pengguna, yang menunjukkan konsentrasi aktivitas di wilayah seperti Jakarta Barat dan Jakarta Pusat pada jam-jam sibuk [18].

Selain dari sisi permintaan, K-Means juga efektif untuk menganalisis faktor penyebab masalah transportasi. Yudistira dkk. (2025) menerapkan K-Means untuk mengelompokkan kota dan kabupaten di Jawa Barat berdasarkan tingkat pertumbuhan kendaraan pribadi. Hasilnya mengidentifikasi lima klaster wilayah, di mana daerah urban seperti Bogor, Bekasi, dan Bandung menunjukkan aktivitas kendaraan tertinggi yang berkontribusi signifikan terhadap kemacetan [12].

Lebih lanjut, K-Means juga digunakan sebagai alat bantu pengambilan keputusan dalam perencanaan infrastruktur oleh instansi pemerintah. Studi oleh Kurniadi dkk. (2022) menerapkan algoritma ini untuk mengelompokkan data usulan pembangunan jalan di Dinas Pekerjaan Umum dan Penataan Ruang (PUPR) Kabupaten Garut. Pengelompokan berdasarkan atribut jenis, panjang, dan biaya pembangunan membantu dinas terkait dalam menentukan prioritas proyek secara lebih objektif dan sistematis [16]. Dalam konteks perencanaan wilayah yang lebih luas, penelitian oleh Rahmiyati dkk. (2025) menunjukkan bagaimana K-Means dapat digunakan untuk memetakan objek wisata di Bogor berdasarkan variabel seperti jumlah kunjungan, fasilitas, dan aksesibilitas. Meskipun domainnya berbeda, pendekatan ini relevan karena menunjukkan kapabilitas *clustering* dalam membantu pemerintah daerah mengalokasikan sumber daya dan menentukan prioritas pengembangan wilayah [11].

2.2 Kesenjangan Penelitian dan Kebaruan

Meskipun teknik Klustering lazim dalam analisis transportasi publik, *research gap* yang signifikan ditemukan : studi mayoritas berfokus pada sisi permintaan (*demand-side*) dan area metropolitan. Analisis dari sisi penawaran (*supply-side*), terutama yang mengukur kuantitas dan legalitas armada di wilayah penyangga yang kompleks dan memiliki masalah angkutan informal seperti Kabupaten Bogor, masih sangat terbatas. Oleh karena itu, penelitian ini memiliki kebaruan dengan menerapkan K-Means Clustering untuk memetakan pemerataan layanan angkutan umum berdasarkan data jumlah armada. Pendekatan kuantitatif ini menawarkan dasar yang objektif bagi pemerintah daerah untuk mengidentifikasi secara spesifik kecamatan yang kekurangan layanan dan memerlukan intervensi kebijakan prioritas.

3. Metode yang Diusulkan

Bab ini menguraikan secara rinci metodologi yang digunakan dalam penelitian untuk menganalisis pemerataan angkutan umum di Kabupaten Bogor. Pembahasan mencakup pendekatan penelitian, sumber dan proses pengumpulan data, serta tahapan analisis data yang meliputi pra-pemrosesan, implementasi algoritma *clustering*, dan teknik penentuan jumlah *cluster* yang optimal.

3.1. Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan teknik *data mining*. Secara spesifik, metode yang digunakan adalah *unsupervised learning* melalui analisis *clustering*. Pendekatan ini dipilih karena kemampuannya untuk mengidentifikasi pola dan struktur intrinsik dalam data tanpa memerlukan label atau kategori yang telah ditentukan sebelumnya. Tujuan utamanya adalah untuk mengelompokkan kecamatan-kecamatan di Kabupaten Bogor ke dalam beberapa *cluster* yang memiliki karakteristik serupa berdasarkan data jumlah armada angkutan umum, baik yang berizin maupun yang tidak berizin. Hasil pengelompokan ini akan menjadi dasar untuk memetakan tingkat pemerataan layanan angkutan umum di seluruh wilayah studi.

3.2. Sumber dan Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang bersumber dari Dinas Perhubungan Kabupaten Bogor[19], [20]. Data tersebut telah diolah dan disimpan dalam format file CSV. Dataset ini memuat informasi komprehensif mengenai distribusi angkutan umum di seluruh kecamatan di Kabupaten Bogor dalam rentang waktu beberapa tahun.

Variabel-variabel kunci yang terdapat dalam dataset meliputi:

- a. bps_nama_kecamatan : Nama kecamatan berdasarkan kode BPS (Badan Pusat Statistik)
- b. tahun : Tahun pencatatan data
- c. Angkutan Berizin : Jumlah angkutan umum yang memiliki izin operasional resmi
- d. Angkutan Tidak Berizin : Jumlah angkutan umum yang beroperasi tanpa izin resmi
- e. Total Angkutan : Jumlah keseluruhan angkutan umum (berizin dan tidak berizin)

3.3. Tahapan Analisis Data

Alur analisis data dalam penelitian ini dirancang secara sistematis yang terdiri dari tiga tahapan utama, yaitu pra-pemrosesan data, analisis *clustering* dengan K-Means, dan penentuan jumlah *cluster* optimal.

3.3.1. Pra-pemrosesan Data

Tahap pra-pemrosesan data merupakan langkah krusial untuk mempersiapkan data sebelum dilakukan analisis *clustering*. Tahap ini terdiri dari dua proses utama:

a. Agregasi Data

Data di-pivot dan dikelompokkan berdasarkan bps_nama_kecamatan untuk mengakumulasi total jumlah angkutan berizin, angkutan tidak berizin, dan total angkutan di setiap kecamatan selama periode observasi. Langkah ini menghasilkan satu set data unik untuk setiap kecamatan yang merepresentasikan kondisi armada secara keseluruhan.

b. Seleksi Fitur

Untuk analisis *clustering*, tiga fitur numerik dipilih sebagai variabel input, yaitu total angkutan berizin, total angkutan tidak berizin, dan total angkutan.

c. Standardisasi Data

Karena ketiga fitur memiliki rentang nilai yang berbeda, dilakukan standardisasi menggunakan StandardScaler dari library Scikit-learn. Proses ini mengubah distribusi setiap fitur sehingga memiliki rata-rata (μ) 0 dan deviasi standar (σ) 1. Standardisasi penting untuk algoritma berbasis jarak seperti K-Means agar setiap fitur memberikan

kontribusi yang seimbang dalam proses pengelompokan. Transformasi ini dilakukan menggunakan formula Z-score:

$$Z = \frac{(x - \mu)}{\sigma}$$

Gambar 1. Formula Z-Score

3.3.2. Analisis Clustering dengan K-Means

Metode *clustering* yang digunakan adalah K-Means, yang merupakan algoritma partisi non-hierarkis. Algoritma ini bertujuan untuk mempartisi n pengamatan ke dalam k *cluster* di mana setiap pengamatan menjadi bagian dari *cluster* dengan *mean* (centroid) terdekat. Proses ini diimplementasikan dengan parameter `init='k-means++'` untuk penempatan centroid awal yang lebih cerdas dan `random_state=42` untuk memastikan hasil analisis dapat direplikasi. Algoritma ini akan meminimalkan *Within-Cluster Sum of Squares* (WCSS) atau inersia, yang merupakan jumlah kuadrat jarak antara setiap titik data dengan centroid *clusternya*.

3.3.3. Penentuan Jumlah Cluster Optimal (K)

Salah satu tantangan utama dalam K-Means adalah menentukan jumlah *cluster* (K) yang paling optimal. Dalam penelitian ini, digunakan dua metode komplementer untuk menentukan nilai K yang paling sesuai dengan struktur data:

a. Metode Elbow

Metode Elbow digunakan sebagai pendekatan utama untuk menentukan jumlah *cluster* optimal. Metode ini bekerja dengan menghitung *Within-Cluster Sum of Squares* (WCSS) untuk rentang nilai K dari 1 hingga 10. WCSS mengukur kompaktness dari *cluster* yang terbentuk, di mana nilai yang lebih rendah mengindikasikan *cluster* yang lebih padat dan terpisah dengan baik. Grafik WCSS versus jumlah *cluster* kemudian dianalisis untuk mengidentifikasi titik "siku" (*elbow*), yaitu titik di mana penurunan WCSS mulai melambat secara signifikan. Titik ini mengindikasikan nilai K yang optimal karena menambahkan *cluster* setelah titik tersebut hanya memberikan perbaikan marginal.

b. Silhouette Score

Metode ini mengukur seberapa mirip sebuah objek dengan *clusternya* sendiri (kohesi) dibandingkan dengan *cluster* lain (separasi). Koefisien siluet dihitung untuk setiap sampel dengan formula :

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

Gambar 2. Formula Silhouette Score

$a(i)$ adalah jarak rata-rata dari sampel i ke titik lain di *cluster* yang sama, dan $b(i)$ adalah jarak rata-rata dari sampel i ke semua titik di *cluster* terdekat berikutnya. Nilai koefisien berkisar antara -1 hingga 1, di mana nilai yang mendekati 1 menunjukkan bahwa sampel tersebut sangat cocok dengan *clusternya*. Nilai K yang memberikan rata-rata koefisien siluet tertinggi dipilih sebagai kandidat K optimal.

Berdasarkan evaluasi dari kedua metode tersebut, jumlah *cluster* optimal (K) ditetapkan untuk analisis akhir.

3.3.4. Analisis Karakteristik dan Visualisasi Cluster

Berdasarkan hasil evaluasi menggunakan Metode Elbow dan Silhouette Score, jumlah *cluster* optimal ditetapkan (misalnya, K=3). Setelah model K-Means dijalankan dengan parameter K tersebut, dilakukan analisis mendalam terhadap karakteristik setiap *cluster* yang terbentuk.

a. Analisis Karakteristik *Cluster*

Analisis ini dilakukan dengan mengekstraksi pusat *cluster* (*centroids*) dari model K-Means. Nilai centroid, yang berada dalam skala standar, kemudian ditransformasi kembali ke skala data asli menggunakan fungsi *inverse_transform* untuk memudahkan interpretasi. Proses ini menghasilkan nilai rata-rata untuk setiap fitur (angkutan berizin, angkutan tidak berizin, total angkutan) pada masing-masing *cluster*, yang memungkinkan identifikasi profil dan karakteristik unik dari setiap kelompok kecamatan.

b. Visualisasi Hasil *Cluster*

Untuk memperjelas pemisahan antar *cluster* dan memfasilitasi interpretasi, dibuat beberapa jenis visualisasi data:

1) Pair Plot

Visualisasi ini menampilkan scatter plot untuk setiap pasangan fitur, dengan titik data diwarnai berdasarkan keanggotaan *clusternya*. Pair plot efektif untuk mengidentifikasi secara visual bagaimana *cluster* terpisah satu sama lain dalam ruang fitur multidimensi.

2) Box Plot

Visualisasi ini menyajikan distribusi nilai dari setiap fitur untuk masing-masing *cluster*. Box plot membantu dalam membandingkan statistik deskriptif (seperti median, kuartil, dan outlier) antar *cluster*, sehingga menonjolkan perbedaan signifikan dalam karakteristik armada angkutan di setiap kelompok.

3.4. Analisis Sekunder : Dinamika Pertumbuhan Armada

Sebagai analisis pelengkap, dilakukan evaluasi terhadap dinamika pertumbuhan jumlah angkutan umum di setiap kecamatan. Analisis ini bertujuan untuk menambahkan dimensi temporal pada hasil clustering yang bersifat statis, serta mengidentifikasi kecamatan-kecamatan yang mengalami perubahan jumlah armada paling signifikan selama periode observasi.

Proses analisis ini dilakukan melalui langkah-langkah berikut:

- a. Identifikasi Data Awal dan Akhir : Untuk setiap kecamatan, data pada tahun paling awal dan paling akhir dalam dataset diidentifikasi.
- b. Perhitungan Peningkatan Absolut : Dihitung selisih (peningkatan atau penurunan) antara data tahun akhir dan tahun awal untuk variabel angkutan berizin, angkutan tidak berizin, dan total angkutan.
- c. Pengurutan Kecamatan : Kecamatan diurutkan berdasarkan nilai peningkatan absolut tertinggi untuk setiap kategori armada.
- d. Visualisasi Data : Hasil analisis 10 kecamatan teratas dengan peningkatan tertinggi divisualisasikan menggunakan diagram batang untuk memberikan gambaran yang jelas mengenai wilayah dengan pertumbuhan armada paling pesat.

4. Hasil dan Pembahasan

Penyajian hasil akan diikuti dengan pembahasan mendalam yang menginterpretasikan temuan tersebut dalam konteks permasalahan pemerataan layanan angkutan umum di Kabupaten Bogor.

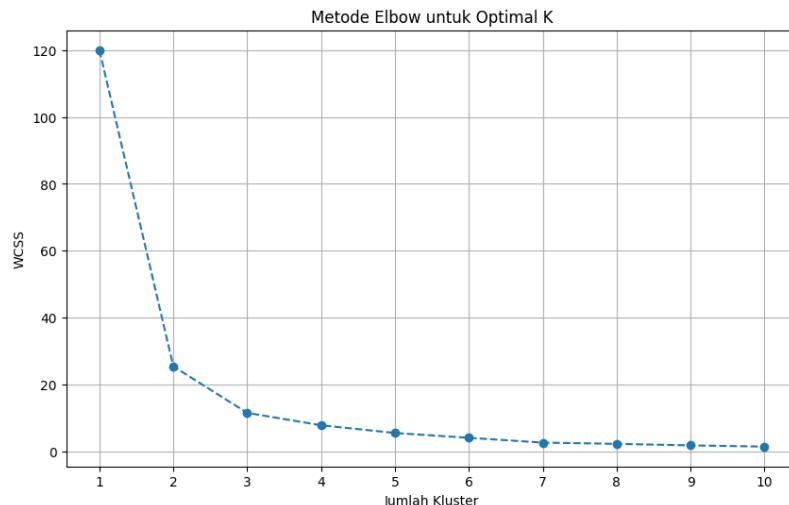
4.1. Hasil Analisis *Clustering*

Analisis clustering dilakukan untuk mengelompokkan kecamatan-kecamatan di Kabupaten Bogor ke dalam beberapa kategori berdasarkan karakteristik jumlah armada angkutan umumnya.

4.1.1. Penentuan Jumlah Cluster Optimal (K)

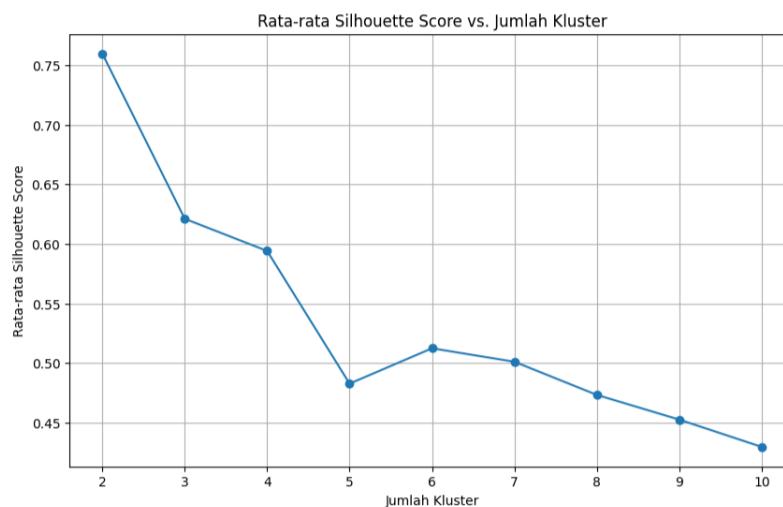
Langkah awal dalam analisis clustering adalah menentukan jumlah *cluster* (K) yang paling optimal untuk mempartisi data. Berdasarkan analisis yang dilakukan, jumlah *cluster* optimal ditetapkan sebanyak 3 *cluster*.

Justifikasi utama untuk pemilihan ini berasal dari Metode Elbow, seperti yang ditunjukkan pada visualisasi yang merujuk pada Grafik Metode Elbow. Grafik tersebut memetakan penurunan nilai *Within-Cluster Sum of Squares* (WCSS) seiring dengan penambahan jumlah *cluster*. Terlihat bahwa penurunan nilai WCSS mulai melandai secara signifikan setelah K=3. Titik ini diinterpretasikan sebagai "siku" (*elbow point*), yang mengindikasikan bahwa penambahan *cluster* lebih lanjut tidak memberikan peningkatan efisiensi yang sepadan dalam menjelaskan varian data.



Gambar 3. Metode Elbow

Keputusan ini juga didukung oleh hasil Silhouette Score. Meskipun nilai tertinggi tidak selalu mutlak pada K=3, nilai pada titik ini menunjukkan tingkat kepadatan dan pemisahan antar *cluster* yang baik, yang menegaskan bahwa pembagian data menjadi tiga kelompok sudah merupakan partisi yang solid dan dapat diinterpretasikan dengan jelas.



Gambar 4. Silhouette Score.

4.1.2. Profil dan Karakteristik Cluster

Sebanyak 40 kecamatan di Kabupaten Bogor berhasil dikelompokkan ke dalam 3 *cluster* yang berbeda. Setiap *cluster* menunjukkan karakteristik yang unik terkait tingkat layanan angkutan umumnya, yang diukur melalui variabel 'Angkutan Berizin', 'Angkutan Tidak Berizin', dan

"Total Angkutan'. Tabel berikut menyajikan nilai rata-rata untuk setiap fitur pada masing-masing *cluster* setelah ditransformasi kembali ke skala aslinya.

Tabel 1. Rata-Rata Peningkatan Angkutan Umum

Kluster	Rata-Rata Peningkatan Angkutan Umum Berizin	Rata-Rata Peningkatan Angkutan Umum Tidak Berizin	Rata-Rata Peningkatan Total Angkutan Umum
0	66.80	-47.80	19.00
1	7.97	3.18	11.15
2	5.00	122.00	127.00

Dari data di atas, dapat ditarik profil sebagai berikut:

- a. *Cluster 0* (Layanan Mapan & Formal) : Terdiri dari 10 kecamatan, *cluster* ini dicirikan oleh jumlah rata-rata angkutan umum berizin yang sangat tinggi dan diikuti oleh penurunan jumlah angkutan umum tidak berizin.

	bps_nama_kecamatan	angkutan Berizin	angkutan Tidak Berizin	angkutan total	Cluster
1	BOJONG GEDE	603	825	1428	0
2	CIBUNGULANG	362	648	1010	0
3	CIOMAS	554	936	1490	0
4	CITEUREUP	730	1174	1904	0
5	JONGGOL	39	794	833	0
6	KEMANG	552	706	1258	0
7	KLAPANUNGGAL	206	844	1050	0
8	MEGAMENDUNG	730	906	1636	0
9	PARUNG	492	992	1484	0
10	TAJUR HALANG	244	898	1142	0

Tabel 2. Kecamatan *Cluster 0*

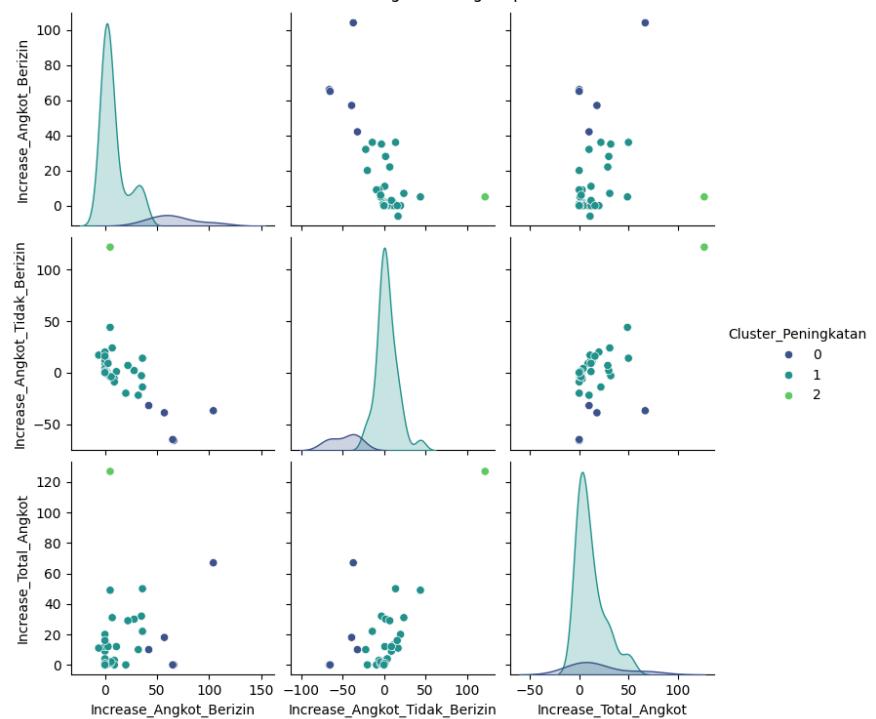
- b. *Cluster 1* (Pertumbuhan Seimbang & Rendah) : Terdiri dari 6 kecamatan, *cluster* ini menunjukkan pertumbuhan armada yang relatif kecil namun seimbang antara sektor formal dan informal.

	bps_nama_kecamatan	angkutan Berizin	angkutan Tidak Berizin	angkutan total	Cluster
1	CIAMPEA	1332	1842	3174	1
2	CIBINONG	1792	2142	3934	1
3	CILEUNGSI	1384	2257	3641	1
4	GUNUNG PUTRI	1251	1442	2693	1
5	PAMIJAHAN	1506	2018	3524	1
6	SUKARAJA	2176	2667	4843	1

Tabel 3. Kecamatan *Cluster 1*

- c. *Cluster 2* (Kekurangan Layanan & Dominasi Informal) : Merupakan kelompok terbesar dengan 24 kecamatan, *cluster* ini memiliki karakteristik jumlah angkutan umum berizin yang sangat rendah namun disertai dengan lonjakan signifikan pada jumlah angkutan umum tidak berizin.

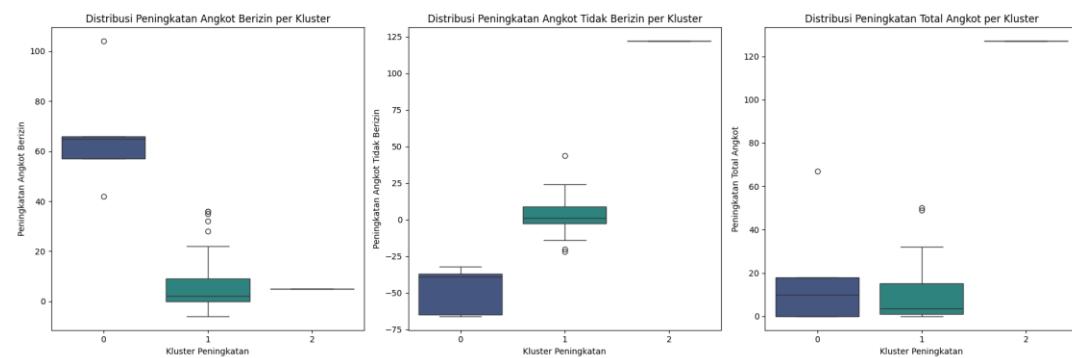
	bps_nama_kecamatan	angkutan Berizin	angkutan Tidak Berizin	angkutan total	Cluster
1	BABAKAN MADANG	278	302	580	2
2	CARINGIN	154	186	340	2
3	CARIU	136	223	359	2
4	CIAWI	204	412	616	2
5	CIGOMBONG	206	398	604	2
6	CIGUDEG	95	118	213	2
7	CIJERUK	305	407	712	2
8	CISARUA	188	328	516	2
9	CISEENG	228	460	688	2
10	DRAMAGA	510	148	658	2
11	GUNUNG SINDUR	154	213	367	2
12	JASINGA	253	306	559	2
13	LEUWILANG	202	172	374	2
14	LEUWISADENG	39	97	136	2
15	NANGGUNG	79	97	176	2
16	PARUNG PANJANG	0	0	0	2
17	RANCA BUNGUR	81	206	287	2
18	RUMPIN	44	234	278	2
19	SUKAJAYA	16	17	33	2
20	SUKAMAKMUR	44	39	83	2
21	TAMANSARI	126	176	302	2
22	TANJUNGSARI	8	10	18	2
23	TENJO	24	24	48	2
24	TENJOLAYA	84	190	274	2

Tabel 3. Kecamatan *Cluster 2***4.1.3. Visualisasi dan Interpretasi Cluster****a. Analisis Pair Plot****Gambar 5.** Pair Plot.

Visualisasi Pair Plot menunjukkan bahwa *Cluster 2* secara visual sangat terpisah dari dua *cluster* lainnya, terutama pada sumbu Peningkatan angkutan tidak berizin. Hal ini

mengonfirmasi bahwa pertumbuhan armada informal adalah pembeda utama dari kelompok kecamatan ini. Sementara itu, *Cluster 0* dan *Cluster 1* menunjukkan adanya tumpang tindih, namun *Cluster 0* secara konsisten menunjukkan nilai yang lebih tinggi pada sumbu peningkatan angkutan berizin.

b. Analisis Box Plot

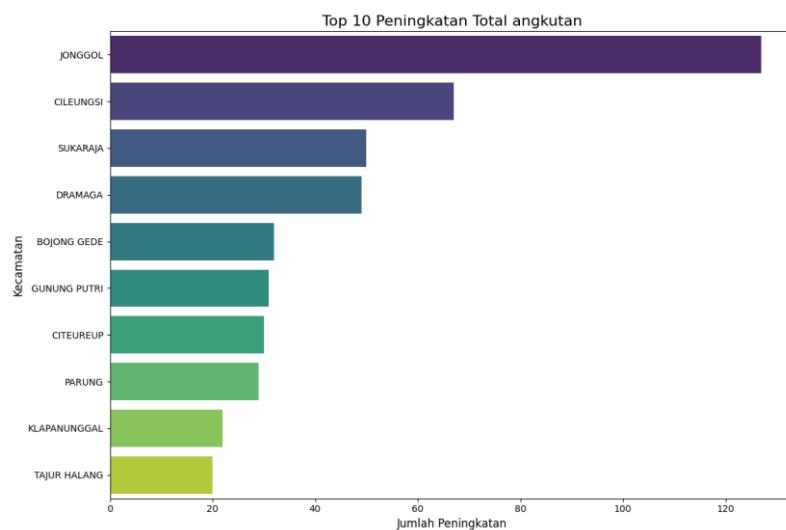


Gambar 6. Box Plot

Visualisasi Box Plot memperkuat temuan ini dengan menunjukkan distribusi data pada setiap *cluster*. *Cluster 2* menunjukkan rentang nilai yang ekstrem tinggi untuk peningkatan angkutan tidak berizin, sedangkan *Cluster 0* memiliki median dan rentang interkuartil (IQR) tertinggi untuk peningkatan angkutan berizin. Distribusi *Cluster 1* cenderung terkonsentrasi di dekat angka nol, menandakan kondisi yang lebih statis.

4.2. Analisis Sekunder : Dinamika Pertumbuhan Armada

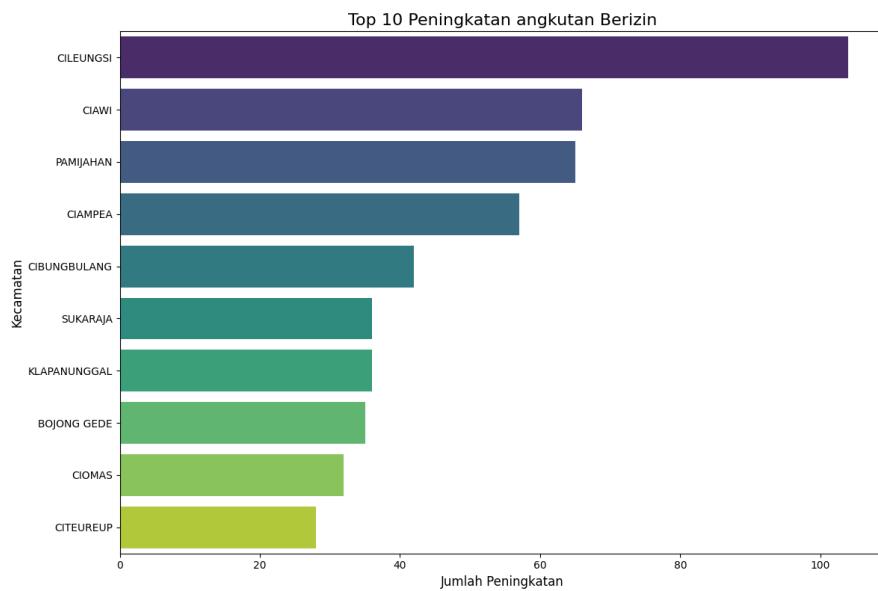
Untuk melengkapi analisis klasterisasi yang bersifat statis, dilakukan analisis temporal untuk mengidentifikasi kecamatan mana yang mengalami dinamika pertumbuhan armada paling signifikan selama periode observasi. Analisis ini dihitung berdasarkan selisih absolut jumlah armada antara tahun terakhir dan tahun awal data.



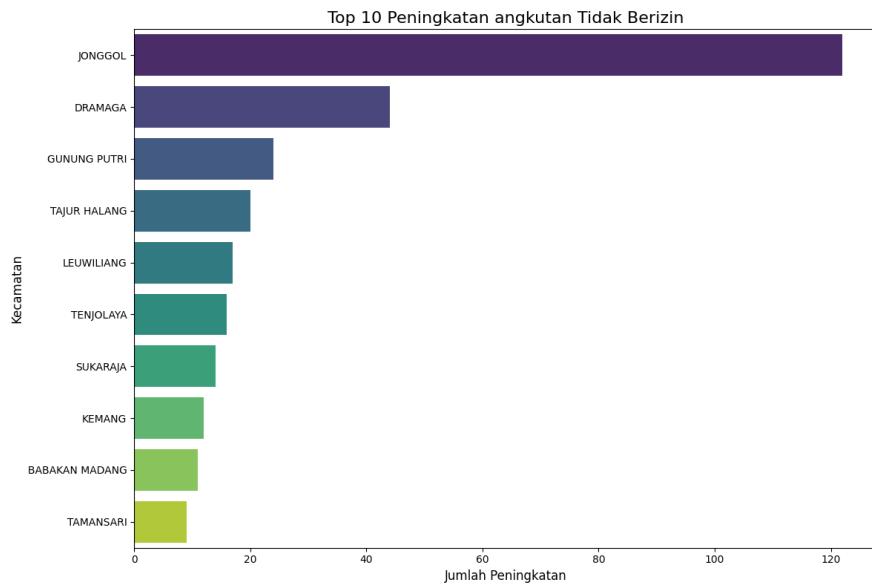
Gambar 7. Top 10 Kecamatan dengan Peningkatan Total Angkutan Umum Tertinggi

Temuan ini memberikan konteks tambahan bahwa beberapa kecamatan yang masuk dalam *Cluster 2* tidak hanya didominasi oleh armada informal, tetapi juga merupakan pusat pertumbuhan armada yang paling aktif. Kecamatan Jonggol, Cileungsi, dan Sukaraja menempati posisi tiga teratas, menunjukkan adanya lonjakan permintaan mobilitas yang sangat tinggi di wilayah tersebut.

Untuk melihat lebih detail komposisi pertumbuhan ini, berikut visualisasi 10 kecamatan teratas berdasarkan peningkatan armada berizin dan tidak berizin.



Gambar 8. Top 10 Kecamatan dengan Peningkatan Angkutan Umum Berizin Tertinggi



Gambar 9. Top 10 Kecamatan dengan Peningkatan Angkutan Umum Tidak Berizin Tertinggi

Dari visualisasi di atas, terlihat pola yang kontras. Peningkatan angkutan umum berizin didominasi oleh kecamatan-kecamatan seperti Cileungsi dan Ciawi. Sebaliknya, peningkatan angkutan tidak berizin secara signifikan terkonsentrasi di kecamatan-kecamatan seperti Jonggol, dan Dramaga yang mengkonfirmasi temuan dari analisis klasterisasi.

4.3. Pembahasan

Hasil analisis kuantitatif ini memberikan bukti empiris yang jelas mengenai adanya disparitas dalam pemerataan layanan angkutan umum di Kabupaten Bogor. Temuan ini berhasil mengisi kesenjangan penelitian yang telah diidentifikasi pada Bab 2. Jika studi sebelumnya cenderung berfokus pada analisis sisi permintaan penelitian ini membuktikan bahwa analisis

dari sisi penawaran dengan membedakan legalitas armada, mampu memetakan zona-zona kritis secara efektif.

Cluster 0 merepresentasikan wilayah pusat layanan seperti Cibinong, di mana sistem transportasi formal relatif berhasil memenuhi permintaan dan menekan pertumbuhan sektor informal. Hal ini sejalan dengan teori perencanaan kota di mana pusat administrasi dan ekonomi cenderung mendapatkan prioritas layanan yang lebih baik.

Cluster 2, yang mencakup mayoritas kecamatan, adalah manifestasi nyata dari kesenjangan layanan (service gap). Rendahnya jumlah angkutan umum berizin di wilayah ini menciptakan kekosongan yang direspon oleh pasar melalui ledakan armada informal. Ini menunjukkan bahwa permintaan masyarakat akan mobilitas sangat tinggi, namun tidak diakomodasi oleh layanan publik yang memadai. Fenomena di Jonggol dan Dramaga menjadi studi kasus yang sempurna, di mana pertumbuhan pesat wilayah tidak diimbangi dengan perencanaan transportasi publik yang proaktif.

Cluster 1 dapat diinterpretasikan sebagai wilayah penyangga atau satelit yang lebih statis, di mana keseimbangan antara penawaran dan permintaan mungkin telah tercapai pada level yang lebih rendah.

Dengan demikian, pendekatan clustering tidak hanya mengelompokkan data, tetapi juga berhasil memetakan geografi ketidakmerataan layanan, memberikan dasar yang objektif untuk perumusan kebijakan.

5. Kesimpulan dan Saran

Penelitian ini bertujuan untuk menganalisis tingkat pemerataan layanan angkutan umum di Kabupaten Bogor dengan menerapkan teknik *clustering K-Means* pada data jumlah armada angkutan umum berizin dan tidak berizin. Berdasarkan analisis yang telah dilakukan, dapat ditarik beberapa kesimpulan utama: 1) Identifikasi Ketidakmerataan Layanan : Penelitian ini berhasil membuktikan secara kuantitatif adanya disparitas yang signifikan dalam penyediaan layanan angkutan umum di seluruh kecamatan di Kabupaten Bogor. Pendekatan clustering secara efektif mengelompokkan kecamatan ke dalam tiga *cluster* yang berbeda, yang merepresentasikan tingkat layanan yang bervariasi. 2) Profil Tiga Kelompok Wilayah : a) *Cluster 0* (Layanan Mapan & Formal): Terdiri dari 10 kecamatan yang didominasi oleh angkutan umum berizin dalam jumlah besar, menunjukkan area dengan layanan transportasi formal yang relatif terpenuhi. b) *Cluster 1* (Pertumbuhan Seimbang & Rendah): Mencakup 6 kecamatan yang cenderung statis dengan pertumbuhan angkutan umum yang rendah, mengindikasikan keseimbangan pasokan dan permintaan pada level yang lebih rendah. c) *Cluster 2* (Kekurangan Layanan & Dominasi Informal): Merupakan kelompok terbesar (24 kecamatan) yang secara jelas mengalami kekosongan layanan (service gap). Rendahnya jumlah armada formal di *cluster* ini direspon oleh lonjakan signifikan armada informal, yang menandakan adanya permintaan mobilitas yang tinggi namun tidak terakomodasi oleh sistem transportasi publik yang resmi. 3) Kontribusi Metodologis : Penggunaan analisis clustering dari sisi penawaran (supply-side), khususnya dengan membedakan legalitas armada, terbukti menjadi pendekatan yang baru dan efektif untuk memetakan wilayah-wilayah yang kurang terlayani (*underserved*) di daerah penyangga yang kompleks seperti Kabupaten Bogor.

Saran

Saran untuk Kebijakan

- a. Intervensi Terfokus pada *Cluster 2* : Pemerintah Kabupaten Bogor disarankan untuk menjadikan 24 kecamatan dalam *Cluster 2* sebagai prioritas utama dalam perencanaan transportasi. Program yang dapat dirancang mencakup legalisasi rute informal yang strategis, pembukaan trayek baru yang terintegrasi, serta peningkatan infrastruktur pendukung.

- b. Pemanfaatan Data untuk Perencanaan : Hasil pemetaan *cluster* ini dapat digunakan sebagai alat bantu pengambilan keputusan yang objektif untuk alokasi anggaran dan sumber daya dalam pengembangan transportasi publik.
- c. Regulasi yang Adaptif : Merumuskan kebijakan yang tidak hanya bersifat restriktif terhadap angkutan informal, tetapi juga adaptif dengan mencari solusi integrasi yang dapat meningkatkan cakupan layanan secara keseluruhan.

Saran untuk Penelitian Selanjutnya

- a. Integrasi Data Kualitas Layanan : Penelitian di masa depan disarankan untuk mengintegrasikan variabel-variabel kualitas layanan ke dalam model analisis untuk mendapatkan gambaran pemerataan yang lebih holistik.
- b. Pemanfaatan Big Data : Menggunakan sumber data yang lebih dinamis seperti data GPS dari armada, data transaksi tiket elektronik, atau data pergerakan dari aplikasi seluler untuk menangkap pola mobilitas yang lebih akurat dan *real-time*.
- c. Analisis Multimoda : Memperluas cakupan analisis dengan memasukkan berbagai moda transportasi lain untuk memahami bagaimana interaksi dan kompetisi antarmoda mempengaruhi sistem transportasi secara keseluruhan.
- d. Pendekatan Kualitatif : Melakukan studi kualitatif melalui wawancara dengan operator angkutan umum (formal dan informal) serta survei kepada pengguna jasa untuk memahami faktor-faktor sosio-ekonomi yang mendorong dinamika yang teramati dalam data.

Ucapan Terima Kasih

Puji syukur kehadiran Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan penelitian dan menyusun naskah ilmiah ini. Penulis mengucapkan terima kasih yang tulus kepada kedua orang tua, Ibu Dewi Novianas dan Bapak Suhamren, serta seluruh keluarga besar atas doa, dukungan moral, dan material yang tak ternilai harganya selama proses penelitian ini.

Ucapan terima kasih ditujukan kepada Bapak Dr. Safrizal, S.T., M.M., M.Kom., selaku dosen pembimbing, atas bimbingan, arahan, dan masukan konstruktif yang telah diberikan sejak awal hingga akhir penyusunan penelitian. Penghargaan juga penulis sampaikan kepada sahabat Penulis, Fadhila Aditya, Muhammad Davin Hassan Ghiffari, dan semua teman-teman lain yang telah memberikan dukungan serta semangat dalam diskusi-diskusi yang membangun.

Daftar Pustaka

- [1] P. Kawasan *et al.*, “*Tataloka The Development of Settlement in the Transit Areas of Bogor Regency*,” vol. 24, pp. 349–365, 2022.
- [2] N. P. Azwanda, A. Kota Bogor, and J. Barat, “Efektivitas Bus Rapid Transit Biska Trans Pakuan dalam Mengurangi Kemacetan di Kota Bogor,” 2024.
- [3] V. Carolin and E. Kurniati, “Tantangan Pembangunan Perkotaan Terhadap Urbanisasi, Kemacetan di Jakarta : Analisis Permasalahan dan Solusi,” *Jurnal Ilmu Ekonomi*, vol. 4, no. 1, pp. 252–273, Apr. 2025, doi: 10.59827/jie.v4i1.222.
- [4] J. A. Ariesandi, R. Resita, and Z. Salsabila, “Kebijakan Transportasi Umum,” vol. 11, pp. 55–112, Oct. 2020.
- [5] A. wulandari, Irmayansyah, and L. T. Ningrum, “Penerapan Metode *K-Means Clustering* Untuk Pemetaan,” vol. 7, pp. 107–116, Feb. 2025.
- [6] M. Annas and S. N. Wahab, “Data Mining Methods: *K-Means Clustering* Algorithms,” *International Journal of Cyber and IT Service Management*, vol. 3, no. 1, pp. 40–47, Mar. 2023, doi: 10.34306/ijctsm.v3i1.122.
- [7] N. Selvia, E. W. Ambarsari, and N. Dwitiyanti, “Shortest Path Clustering Dalam Menyaring Tingkat Kepadatan Arus Lalu Lintas,” *JURIKOM (Jurnal Riset Komputer)*, vol. 10, no. 2, p. 396, Apr. 2023, doi: 10.30865/jurikom.v10i2.5979.
- [8] J. Manajemen *et al.*, “Pengaruh Data Warehouse Terhadap Pengambilan Keputusan Article History,” vol. 3, no. 1, 2024.
- [9] F. Dwi Ratnaningtyas and D. Widi Astuti, “Kurang Optimalnya Fungsi dari Terminal Cepu Type A,” 2022. [Online]. Available: <http://siar.ums.ac.id/>
- [10] I. hiqmah Turmuzi, D. Mellenia, and U. Ikhwana Fadhlilah, “Implementasi Teknologi *Big Data* di Pemerintahan Indonesia,” 2023.
- [11] D. Rahmiyati and E. Britanton Siswantoro, “Penerapan Metode *K-Means* untuk Pemetaan Objek Wisata Sebagai Rekomendasi Prioritas Pengembangan Pariwisata,” 2025. [Online]. Available: <http://creativecommons.org/licenses/by/4.0/>

- [12] A. R. Yudistira, H. A. Yasmin, and M. K. Pasya, "Upaya Mengatasi Kemacetan di Jawa Barat Melalui Pengelompokkan Kendaraan Pribadi," *Gunung Djati Conference Series*, vol. 50, 2025.
- [13] H. F. Ramadhan, A. Fauzi, C. N. Rupelu, D. P. Aprillia, N. D. Anjani, and Halimatusadiah, "Pengaruh *Business Intelligence* terhadap Perusahaan," Jun. 2022.
- [14] J. Zemnickis, "Data Warehouse Data Model Improvements from Customer Feedback," *Baltic Journal of Modern Computing*, vol. 11, no. 3, pp. 475–499, 2023, doi: 10.22364/bjmc.2023.11.3.08.
- [15] H. Mulyani, R. A. Setiawan, and H. Fathi, "Optimization of K Value in Clustering Using Silhouette Score (Case Study: Mall Customers Data)," *Journal of Information Technology and Its Utilization*, vol. 6, no. 2, pp. 45–50, Dec. 2023, doi: 10.56873/jitu.6.2.5243.
- [16] D. Kurniadi, Y. Handoko Agustin, H. Ilham Nur Akbar, and I. Farida, "Penerapan Algoritma *k-Means Clustering* untuk Pengelompokan Pembangunan Jalan pada Dinas Pekerjaan Umum dan Penataan Ruang," *ATII: Jurnal Teknologi Informasi*, vol. 20, no. 1, pp. 64–77, 2022.
- [17] S. Paembonan, H. Abduh, and K. Kunci, "Penerapan Metode *Silhouette Coeficient* Untuk Evaluasi *Clustering* Obat *Clustering*; K-means; *Silhouette coefficient*," 2021. [Online]. Available: <https://ojs.unanda.ac.id/index.php/jiit/index>
- [18] Reynaldi, R. J. F. Djarot, M. Wahyudi, Sumanto, and A. S. Budiman, "Analisa Pola Penyebaran Pengguna Layanan Transjakarta dengan Metode *K-Means Clustering*," 2025.
- [19] "Jumlah Angkutan Umum (Unit) AKDP dan AKDK Menurut Kecamatan Di Kabupaten Bogor," Nov. 2024. Accessed: Sep. 18, 2025. [Online]. Available: <https://opendata.bogorkab.go.id/dataset/jumlah-angkutan-umum--unit-akdp-dan-akdkmenurut-kecamatan-di-kabupaten-bogor>
- [20] "Jumlah Angkutan Umum Berizin (Unit) AKDP dan AKDK Menurut Kecamatan Di Kabupaten Bogor," 2024. Accessed: Sep. 18, 2025. [Online]. Available: <https://opendata.bogorkab.go.id/dataset/jumlah-angkutan-umum-berizin-unit-akdp-dan-akdk-menurut-kecamatan-di-kabupaten-bogor>