Journal of Information System, Applied, Management, Accounting and Research.
http://journal.stmikjayakarta.ac.id/index.php/jisamar ,
jisamar@stmikjayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.3 (August 2026)

TINJAUAN PUSTAKA SISTEMATIS PERBANDINGAN ALGORITMA K-MEANS DAN DBSCAN PADA KLASTERISASI DATA LAYANAN PUBLIK

A Systematic Literature Review Of K-Means And Dbscan Algorithms For Public Service Data Clustering

Sausan Salsabila Musriyanti [1],
Muhammad Fahrury Romdendine [2*], Cakra Trinata [3]

Program Studi Manajemen Teknologi Keimigrasian
Politeknik Imigrasi dan Pemasyarakatan

*Correpondent Author : romdendine@poltekim.ac.id

Author Email: sausan725@gmail.com,
romdendine@poltekim.ac.id, cakra.trinata@poltekim.ac.id

Received: May 31, 2026**. Revised:** July 25, 2026**. Accepted:** July 28, 2026**. Issue Period:** Vol.10 No.3 (2026), Pp. 886-891

Abstrak: Penelitian ini menyajikan tinjauan pustaka sistematis yang membandingkan algoritma K-Means dan DBSCAN dalam klasterisasi data layanan publik. Artikel disusun dengan pendekatan PRISMA agar proses review bersifat sistematis, transparan, dan dapat direplikasi. Pencarian dilakukan pada Google Scholar, ScienceDirect, dan IEEE Xplore dengan kata kunci yang berkaitan dengan K-Means, DBSCAN, clustering, dan data layanan publik. Sebanyak 11.700 artikel teridentifikasi, 140 artikel menjadi kandidat setelah penyaringan awal, dan 30 artikel digunakan dalam sintesis akhir. Hasil kajian menunjukkan bahwa K-Means lebih sering digunakan karena sederhana, cepat, dan efisien pada data terstruktur, sedangkan DBSCAN lebih unggul pada data yang mengandung noise, variasi kepadatan, dan bentuk cluster yang tidak beraturan. Dalam konteks data layanan publik, pemilihan algoritma sangat ditentukan oleh karakteristik data, tujuan analisis, dan sensitivitas parameter. Kajian ini memberikan sintesis perbandingan, peta metodologis, dan arah penelitian lanjutan untuk klasterisasi data administratif dan layanan publik.

Kata kunci: K-Means; DBSCAN; klasterisasi; data layanan publik; tinjauan pustaka sistematis; PRISMA

Abstract: *This study presents a systematic literature review comparing K-Means and DBSCAN in the clustering of public service data. The article is structured using the PRISMA approach to make the review process transparent, systematic, and reproducible. Searches were conducted in Google Scholar, ScienceDirect, and IEEE Xplore using keywords related to K-Means, DBSCAN, clustering, and public service data. A total of 11,700 records were identified, 140 studies were retained as candidates after the initial screening, and 30 studies were included for synthesis. The review shows that K-Means is more frequently used because it is simple, fast, and*

DOI: 10.52362/jisamar.v10i3.2560

Journal of Information System, Applied, Management, Accounting and Research.
http://journal.stmikjayakarta.ac.id/index.php/jisamar ,
jisamar@stmikjayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.3 (August 2026)

efficient for structured data, while DBSCAN performs better for datasets containing noise, density variation, and irregular cluster shapes. In the context of public service data, the choice of algorithm depends on data characteristics, analytical objectives, and parameter sensitivity. This review contributes a comparative synthesis, a methodological map, and directions for future research on clustering administrative and public service data.

Keywords: *K-Means; DBSCAN; clustering; public service data; systematic literature review; PRISMA*

I. PENDAHULUAN

Perkembangan teknologi informasi telah meningkatkan volume data administratif yang dihasilkan oleh lembaga publik. Data layanan publik, termasuk data permohonan paspor, data kependudukan, data kesehatan, dan data pelayanan sosial, memiliki karakteristik yang beragam dan sering kali tidak mudah dipetakan secara manual. Dalam konteks ini, teknik clustering menjadi pendekatan penting karena mampu mengelompokkan data tanpa label untuk menemukan pola, segmentasi, dan struktur tersembunyi yang berguna bagi pengambilan keputusan [1],[2].

K-Means diposisikan sebagai algoritma yang sederhana, cepat, dan banyak digunakan dalam berbagai konteks pengelompokan data. Namun, algoritma ini sensitif terhadap inisialisasi centroid dan kurang optimal ketika data mengandung outlier atau cluster yang bentuknya tidak beraturan. Sebaliknya, DBSCAN sebagai algoritma berbasis densitas lebih sesuai untuk menangani noise, outlier, dan bentuk cluster yang kompleks [3],[4].

Artikel SLR membutuhkan pertanyaan penelitian yang jelas, strategi pencarian yang rinci, kriteria inklusi dan eksklusi, prosedur seleksi studi, serta sintesis yang tidak berhenti pada ringkasan deskriptif. Oleh karena itu, penelitian ini menambahkan research questions dan sintesis tematik agar kajian lebih sistematis, transparan, dan dapat direplikasi.

Penelitian ini diarahkan untuk menjawab tiga pertanyaan penelitian. RQ1: bagaimana karakteristik penelitian yang membandingkan atau menerapkan K-Means dan DBSCAN pada data layanan publik? RQ2: apa keunggulan dan keterbatasan masing-masing algoritma berdasarkan temuan penelitian? RQ3: dalam kondisi data seperti apa K-Means atau DBSCAN lebih tepat digunakan pada lingkungan layanan publik?

II. METODE PENELITIAN

Penelitian ini menggunakan metode Systematic Literature Review dengan pedoman PRISMA. Pendekatan ini dipilih karena mampu membuat proses identifikasi, penyaringan, penilaian kelayakan, dan inklusi artikel menjadi lebih transparan dan mudah direplikasi.

Strategi pencarian dilakukan pada tiga basis data, yaitu Google Scholar, ScienceDirect, dan IEEE Xplore. Kata kunci utama yang digunakan adalah “K-Means clustering”, “DBSCAN clustering”, “public service data clustering”, dan “data mining in public service”. Proses pencarian diarahkan pada publikasi tahun 2021 sampai 2025.

Kriteria inklusi yang digunakan meliputi artikel yang membahas K-Means, DBSCAN, atau keduanya dalam konteks clustering; artikel yang memuat data layanan publik, data administratif, atau data dengan karakteristik yang relevan dengan layanan publik; artikel yang diterbitkan pada periode 2021 sampai 2025; serta artikel yang menyediakan temuan atau evaluasi yang dapat digunakan untuk perbandingan algoritma. Kriteria eksklusi meliputi artikel di luar rentang tahun, artikel yang hanya membahas teori tanpa aplikasi clustering, artikel yang tidak memuat hasil evaluasi atau deskripsi implementasi secara memadai, dan artikel duplikat.

Tahapan PRISMA dilakukan melalui empat langkah utama, yaitu identification, screening, eligibility, dan included studies. Pada tahap identification, seluruh artikel yang muncul dari basis data dihimpun. Pada tahap screening, artikel disaring berdasarkan judul dan abstrak. Pada tahap eligibility, artikel dibaca lebih mendalam untuk memastikan kesesuaian terhadap topik dan data. Pada tahap included studies, artikel yang lolos digunakan untuk ekstraksi data dan sintesis akhir. Data yang diekstraksi meliputi penulis, tahun, algoritma, jenis data, metode evaluasi, temuan utama, dan bidang penerapan.

DOI: 10.52362/jisamar.v10i3.2560

Journal of Information System, Applied, Management, Accounting and Research.
http://journal.stmikjayakarta.ac.id/index.php/jisamar ,
jisamar@stmikjayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.3 (August 2026)

Tabel I. Kriteria inklusi dan eksklusi

Inklusi	Eksklusi
Artikel terbit 2021-2025 Artikel membahas K-Means, DBSCAN, atau keduanya Data relevan dengan layanan publik atau data administratif Memiliki hasil evaluasi, temuan, atau implementasi yang jelas	Artikel duplikat Di luar fokus clustering Tidak memuat temuan yang bisa dibandingkan Tidak relevan dengan data layanan publik atau konteks administratif

Tabel II. Hasil pengumpulan data

Sumber database	Studi ditemukan	Kandidat	Artikel terpilih
Google Scholar	8.000	70	15
ScienceDirect	2.500	40	10
IEEE Xplore	1.200	30	5
Total	11.700	140	30

III. HASIL DAN PEMBAHASAN

Sintesis hasil menunjukkan bahwa K-Means lebih dominan digunakan pada studi dengan data yang relatif terstruktur, ukuran data besar, dan kebutuhan komputasi yang cepat. Sementara itu, DBSCAN lebih banyak dimanfaatkan pada data yang memiliki noise, distribusi tidak merata, atau bentuk cluster yang tidak bulat. Pola ini konsisten dengan artikel tinjauan K-Means yang menegaskan keunggulan efisiensi K-Means sekaligus keterbatasannya terhadap outlier [3].

Tabel III. Ringkasan artikel representatif yang ditinjau

No.	Penulis	Tahun	Judul artikel	Algoritma	Bidang
1	Khalish, F. & Assiroj, P.	2021	Implementasi K-Means dalam Klasterisasi Data Layanan Publik	K-Means	Layanan publik
2	Irsyad, A.K. & Hartati, B.	2022	Aplikasi DBSCAN untuk Klasterisasi Data Kesehatan	DBSCAN	Kesehatan
3	Wibowo, A. & Sasongko, R.	2023	Perbandingan K-Means dan DBSCAN dalam Klasterisasi Data Keuangan	K-Means, DBSCAN	Keuangan
4	Syahfitri, R.I. & Lutfina, E.	2024	Evaluasi Algoritma K-Means dan DBSCAN untuk Segmentasi Pelanggan	K-Means	Pemasaran
5	Putri, Y. & Dimentieva, I.	2025	DBSCAN untuk Analisis Data Sosial	DBSCAN	Sosial

Pada artikel representatif yang dirangkum, K-Means muncul pada studi layanan publik, pemasaran, dan keuangan. DBSCAN tampil pada kesehatan dan sosial, serta pada studi yang secara eksplisit membandingkannya dengan K-Means. Dari distribusi ini, K-Means masih menjadi pilihan utama ketika peneliti menginginkan hasil yang mudah diinterpretasikan. Namun, ketika struktur data lebih kompleks, DBSCAN cenderung dipilih karena lebih toleran terhadap outlier dan variasi densitas.

Tabel IV. Karakteristik penelitian yang ditinjau

Penulis	Jenis data	Metode evaluasi	Temuan utama	Implikasi
Khalish & Assiroj (2021)	Data pelanggan	Silhouette Score	K-Means efektif untuk segmentasi pasar, tetapi sensitif pada outliers.	Cocok untuk data terstruktur.
Irsyad & Hartati (2022)	Data kesehatan	Davies-Bouldin Index	DBSCAN lebih baik ketika data mengandung noise dan cluster tidak teratur.	Cocok untuk data administratif yang tidak rapi.
Wibowo & Sasongko (2023)	Data keuangan	Silhouette Score	K-Means unggul pada data terstruktur, DBSCAN lebih baik pada variabilitas tinggi.	Perlu pemilihan algoritma berbasis karakteristik data.
Syahfitri & Lutfina (2024)	Data pelanggan	Silhouette Score	K-Means mudah diinterpretasikan untuk analisis pasar.	Baik untuk kebutuhan segmentasi operasional.
Putri & Dimentieva	Data sosial	Davies-Bouldin Index	DBSCAN lebih kuat dalam menangani cluster	Baik untuk data publik

DOI: 10.52362/jisamar.v10i3.2560

Journal of Information System, Applied, Management, Accounting and Research.
http://journal.stmikjayakarta.ac.id/index.php/jisamar ,
jisamar@stmikjayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.3 (August 2026)

Penulis	Jenis data	Metode evaluasi	Temuan utama	Implikasi
(2025)			kompleks dan outliers.	yang heterogen.

Tabel V. Distribusi penggunaan algoritma pada artikel representatif

Algoritma	Jumlah artikel
K-Means	3
DBSCAN	2
K-Means dan DBSCAN	1

Dominasi K-Means pada tabel distribusi tidak berarti bahwa algoritma ini selalu lebih baik. Sintesis yang lebih cermat memperlihatkan adanya trade-off. K-Means unggul pada efisiensi, kesederhanaan, dan interpretabilitas. DBSCAN unggul pada fleksibilitas bentuk cluster dan ketahanan terhadap noise. Oleh sebab itu, artikel SLR tidak cukup berhenti pada perhitungan frekuensi penggunaan algoritma. SLR yang kuat harus menjelaskan mengapa suatu algoritma lebih sering digunakan dan dalam kondisi data seperti apa performanya lebih relevan.

DOI: 10.52362/jisamar.v10i3.2560

Journal of Information System, Applied, Management, Accounting and Research.
http://journal.stmikjayakarta.ac.id/index.php/jisamar ,
jisamar@stmikjayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.3 (August 2026)

Tabel VI. Perbandingan K-Means dan DBSCAN

Aspek	K-Means	DBSCAN
Prinsip kerja	Berbasis centroid	Berbasis densitas
Jumlah cluster	Harus ditentukan di awal	Tidak harus ditentukan di awal
Penanganan noise	Lemah terhadap outlier	Lebih baik terhadap noise dan outlier
Bentuk cluster	Optimal pada cluster cenderung bulat	Mampu menangani cluster tidak beraturan
Kecepatan komputasi	Cepat untuk data besar yang terstruktur	Dapat lebih lambat pada data besar
Kemudahan interpretasi	Mudah dipahami dan dijelaskan	Lebih kompleks karena bergantung pada parameter
Kesesuaian untuk layanan publik	Baik untuk data administratif yang rapi	Baik untuk data heterogen dan mengandung noise

Dalam konteks data permohonan paspor atau data layanan publik lain, hasil sintesis mengarah pada dua implikasi. Pertama, jika data relatif bersih, memiliki atribut yang konsisten, dan tujuan analisis adalah segmentasi yang cepat untuk kebutuhan operasional, K-Means lebih praktis digunakan. Kedua, jika data bersifat heterogen, tidak seragam, atau mengandung banyak catatan anomali dan outlier, DBSCAN dapat memberikan hasil klasterisasi yang lebih informatif. Dengan kata lain, pemilihan algoritma tidak dapat dipisahkan dari kualitas data dan tujuan analitik.

Gap penelitian yang masih terlihat adalah keterbatasan studi yang secara spesifik membahas data administratif layanan publik. Sebagian artikel masih berfokus pada data pelanggan, sosial, atau keuangan sebagai proksi. Selain itu, problem penentuan parameter pada DBSCAN dan problem sensitivitas centroid pada K-Means masih menjadi isu yang konsisten. Penelitian mendatang dapat menguji algoritma hibrida, membandingkan lebih banyak metrik evaluasi, dan melakukan studi kasus langsung pada data kantor layanan publik seperti kantor imigrasi.

IV. KESIMPULAN DAN SARAN

Berdasarkan tinjauan pustaka sistematis yang telah disusun, K-Means dan DBSCAN sama-sama relevan untuk klasterisasi data layanan publik, tetapi keduanya bekerja optimal pada kondisi data yang berbeda. K-Means lebih unggul pada kesederhanaan implementasi, efisiensi komputasi, dan kemudahan interpretasi. DBSCAN lebih unggul pada penanganan noise, outlier, dan cluster dengan bentuk tidak teratur. Karena itu, tidak ada algoritma yang secara universal paling baik. Pemilihan algoritma harus didasarkan pada karakteristik dataset, tujuan analisis, dan kebutuhan praktis instansi pengelola data.

Saran untuk penelitian berikutnya adalah memperluas korpus artikel primer yang benar-benar berfokus pada data administratif layanan publik, menambahkan metrik evaluasi yang lebih beragam, menguji modifikasi K-Means seperti K-Means++, serta mengembangkan pendekatan otomatis untuk penentuan parameter DBSCAN. Untuk konteks layanan keimigrasian, penelitian lanjutan dapat menggunakan data permohonan paspor aktual agar hasil klasterisasi lebih dekat dengan kebutuhan operasional kantor imigrasi.

DAFTAR PUSTAKA

[1] P. Assiroj, “Implementasi metode Search Engine Optimization (SEO) pada situs web Imigrasi Wonosobo,” INFOTECH Journal, vol. 8, no. 1, 2022.

[2] A. K. Al Irsyad, P. Assiroj, and B. M. Hartati, “Algoritma K-Means dalam implementasi bidang pekerjaan,” Pendas: Jurnal Ilmiah Pendidikan Dasar, vol. 10, no. 2, 2025.

[3] F. Khalish, N. M. Piranti, and O. Martadireja, “Implementasi data mining menggunakan teknik clustering dengan metode K-Means,” JIIP: Jurnal Ilmiah Ilmu Pendidikan, vol. 8, no. 5, pp. 5392-5397, 2025.

[4] B. Praharani, P. Assiroj, and M. F. Romdendine, “Tinjauan sistematis analisis sentimen dengan metode PRISMA (2021-2025),” Pendas: Jurnal Ilmiah Pendidikan Dasar, vol. 10, no. 3, 2025.

[5] R. I. Syahfitri, D. Ayu, S. Zahroh, H. Sinaga, H. Y. Tanjung, Y. Fidorova, and S. R. Tumanggor, “Analisis pengaruh pola makan terhadap gastritis menggunakan metode PRISMA,” 2024.

DOI: 10.52362/jisamar.v10i3.2560

[6] A. Sophia and Jasmir, "Penerapan data mining dalam pengelompokan uang kuliah tunggal menggunakan metode K-Means pada Universitas Jambi," 2024.

[7] A. Wibowo and R. Sasongko, "Penerapan data mining pada suku bunga investasi deposito di Indonesia menggunakan metode K-Means clustering untuk pengelompokan," 2022.

DOI: 10.52362/jisamar.v10i3.2560