Jurnal Surya Informatika
Journal homepage: https://jurnal.umpp.ac.id/index.php/surya_informatika
P-ISSN: 2477-3042; E-ISSN: 3026-3034
Vol. 16, No. 2, November 2026, Pp. 217-227

Perbandingan Kinerja K-Means dan PSO K-Means Untuk Klasterisasi Dokumen Kependudukan Berbasis SIG

Bintang Aprilia*[1)], Muhammad Hatta[2)], Mesi Febima[3)]
1. Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Catur Insan Cendekia, Indonesia
2. Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Catur Insan Cendekia, Indonesia
3. Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Catur Insan Cendekia, Indonesia

Article Info

Kata Kunci: *Davies-Bouldin Index*; dokumen kependudukan*; K-means clustering; Particle swarm optimization*; sistem informasi geografis

Keywords: *civil registration documents; Davies-Bouldin Index; geographic information system; K-means clustering; Particle swarm optimization*

Article history:
Received: 31 Juli 2026
Revised: 15 Agustus 2026
Accepted: 17 Agustus 2026
Available online: 1 November 2026

DOI:
10.48144/suryainformatika.v16i2.2520

* Corresponding author.
Bintang Aprilia
bintangaprilia046@gmail.com

ABSTRAK

Dinas Kependudukan dan Pencatatan Sipil Kabupaten Cirebon belum memiliki analisis berbasis data yang dapat mengelompokkan tingkat kebutuhan layanan dokumen kependudukan sekaligus menampilkan persebarannya secara spasial. Penelitian ini bertujuan mengelompokkan persebaran dokumen kependudukan pada 40 kecamatan dan membangun sistem informasi geografis untuk memvisualisasikan hasil pengelompokan. Data periode 2024–2025 terdiri atas Kartu Keluarga, KTP, Kartu Identitas Anak, Akta Kelahiran, dan Akta Kematian. Tahapan analisis mengikuti *CRISP-DM*, meliputi pemahaman masalah, pemahaman data, pembersihan, seleksi atribut, normalisasi *Min-Max*, pemodelan *K-Means*, optimasi *centroid* menggunakan *Particle Swarm Optimization*, evaluasi, dan penerapan hasil ke sistem. Tiga cluster digunakan untuk merepresentasikan kategori tinggi, sedang, dan rendah. *K-Means* menghasilkan distribusi cluster 13, 17, dan 10 kecamatan pada 2024 serta 15, 15, dan 10 kecamatan pada 2025. *K-Means* + PSO menghasilkan distribusi 14, 10, dan 16 kecamatan pada kedua tahun. Optimasi menurunkan SSE dari 4,314 menjadi 3,779 atau sebesar (12,97%) dan DBI dari 1,230 menjadi 1,050 atau sebesar (14,62%) pada 2024, serta menurunkan SSE dari 4,053 menjadi 3,783 atau sebesar (6,67%) dan DBI dari 1,170 menjadi 1,073 atau sebesar (5,11%). pada 2025. Seluruh fitur sistem untuk admin dan kepala dinas dinyatakan berhasil melalui pengujian *black box*. Hasil ini menunjukkan bahwa PSO memperbaiki kualitas cluster dan integrasi SIG memudahkan interpretasi persebaran wilayah.

ABSTRACT

The Cirebon Regency Population and Civil Registration Office does not yet have a data-driven analysis capable of categorizing the demand for population documents while simultaneously displaying their spatial distribution. This study aims to categorize the distribution of population documents across 40 subdistricts and develop a geographic information system to visualize the results of this categorization. Data for the 2024–2025 period consists of Family Cards, ID Cards, Child Identity Cards, Birth Certificates, and Death Certificates. The analysis stages follow the CRISP-DM framework, including problem understanding, data understanding, data cleaning, attribute selection, Min-Max normalization, K-Means modeling, centroid optimization using Particle Swarm Optimization, evaluation, and implementation of the results into the system. Three clusters were used to represent the high, medium, and low categories. K-Means produced a cluster distribution of 13, 17, and 10 subdistricts in 2024 and 15, 15, and 10 subdistricts in 2025. K-Means + PSO produced a distribution of 14, 10, and 16 subdistricts in both years. The optimization reduced the SSE from 4.314 to 3.779 (a decrease of 12.97%) and the DBI from 1.230 to 1.050 (a decrease of 14.62%) in 2024, and reduced the SSE from 4.053 to 3.783 (a 6.67% decrease) and the DBI from 1.170 to 1.073 (a 5.11% decrease) in 2025. All system features for administrators and department heads were confirmed to function successfully through black-box testing. These results indicate that the PSO improves cluster quality and that GIS integration facilitates the interpretation of spatial distribution.

Jurnal Surya Informatika
Journal homepage: https://jurnal.umpp.ac.id/index.php/surya_informatika
P-ISSN: 2477-3042; E-ISSN: 3026-3034
Vol. 16, No. 2, November 2026, Pp. 217-227

1. PENDAHULUAN

Pemanfaatan *machine learning* dalam pelayanan publik memungkinkan data historis tidak hanya disimpan sebagai arsip, tetapi diolah menjadi informasi yang mendukung perencanaan dan pengambilan keputusan. Pada administrasi kependudukan, data penerbitan dan kepemilikan dokumen dapat digunakan untuk mengenali pola kebutuhan layanan pada setiap wilayah. Pendekatan *clustering* relevan untuk kondisi tersebut karena mampu membentuk kelompok berdasarkan kemiripan karakteristik tanpa memerlukan label kelas. Penerapan *K-means* pada data kemiskinan dan berbagai indikator kewilayahan menunjukkan bahwa algoritma ini dapat menyederhanakan data menjadi kelompok yang lebih mudah diinterpretasikan [1], [2].

Dinas Kependudukan dan Pencatatan Sipil Kabupaten Cirebon mengelola data Kartu Keluarga, KTP, Kartu Identitas Anak, Akta Kelahiran, dan Akta Kematian dari seluruh kecamatan. Data tersebut sebelumnya lebih banyak digunakan sebagai laporan administratif dan belum dianalisis secara terstruktur untuk membedakan wilayah dengan tingkat kepemilikan dokumen tinggi, sedang, dan rendah. Ketiadaan pengelompokan menyebabkan kebutuhan layanan antarwilayah sulit dibandingkan, sedangkan penyajian dalam bentuk tabel belum memberikan gambaran spasial yang mudah dipahami. Kondisi ini membatasi pemanfaatan data historis sebagai dasar penentuan prioritas pelayanan dan pengembangan administrasi kependudukan berbasis data [3]. Variabel yang digunakan dalam proses pengelompokan berupa jumlah Kartu Keluarga, KTP, Kartu Identitas Anak, Akta Kelahiran, dan Akta Kematian pada masing-masing kecamatan.

K-means dipilih karena prosesnya sederhana, efisien, dan sesuai untuk melakukan pengelompokan data numerik. Namun, hasil *K-means* sensitif terhadap pemilihan *centroid* awal. *Centroid* yang kurang representatif dapat menyebabkan proses iterasi menghasilkan solusi lokal sehingga kualitas cluster yang terbentuk belum optimal. Kondisi tersebut menjadi lebih penting pada data kewilayahan karena setiap wilayah memiliki karakteristik jumlah dokumen yang berbeda dan tingkat variasi antarkecamatan yang tidak seragam. *Particle Swarm Optimization* (PSO) sesuai digunakan untuk mengatasi permasalahan tersebut karena memiliki kemampuan pencarian global dalam menentukan posisi *centroid* yang lebih optimal sebelum proses *K-means* dilakukan. Penerapan kombinasi PSO dan *K-means* pada pengelompokan indikator pembangunan manusia dan data kewilayahan menunjukkan bahwa optimasi *centroid* dapat meningkatkan kualitas hasil *clustering* dibandingkan K-means konvensional [3], [4]. Penggunaan PSO dalam optimasi proses clustering juga telah diterapkan pada pendekatan *hybrid* untuk meningkatkan kualitas pembentukan cluster melalui pencarian solusi yang lebih optimal [5]. Penelitian lain juga menunjukkan bahwa PSO dapat digunakan untuk menentukan *centroid* awal yang lebih optimal sehingga kualitas cluster mengalami peningkatan berdasarkan metrik evaluasi yang digunakan [3].

Penelitian yang menggabungkan clustering dengan Sistem Informasi Geografis (SIG) telah digunakan untuk memetakan berbagai fenomena kewilayahan, termasuk distribusi tenaga pendidik dan objek berbasis geospasial [6], [7], [8]. Penerapan K-means dalam pemetaan berbasis SIG memungkinkan hasil pengelompokan tidak hanya ditampilkan dalam bentuk tabel, tetapi juga divisualisasikan berdasarkan lokasi wilayah sehingga pola persebarannya lebih mudah dipahami. Meskipun demikian, penelitian terdahulu masih banyak yang berfokus pada salah satu sisi, yaitu peningkatan kualitas cluster melalui optimasi tanpa visualisasi spasial atau visualisasi SIG menggunakan *K-means* konvensional tanpa optimasi *centroid*. Kesenjangan tersebut menjadi dasar penelitian ini untuk mengintegrasikan *K-means* yang dioptimasi PSO dengan Sistem Informasi Geografis dalam menganalisis persebaran dokumen kependudukan Kabupaten Cirebon. Hasil pengelompokan divisualisasikan menggunakan *Leaflet* karena merupakan pustaka *JavaScript open-source* yang ringan dan mendukung pembuatan peta interaktif berbasis web, termasuk penggunaan *marker*, *polygon*, *popup*, dan *GeoJSON* [8].

Penelitian ini memiliki dua tujuan. Pertama, mengelompokkan data persebaran dokumen kependudukan periode 2024–2025 pada 40 kecamatan ke dalam tiga kategori kebutuhan layanan berdasarkan jumlah Kartu Keluarga, KTP, Kartu Identitas Anak, Akta Kelahiran, dan Akta Kematian. Kedua, membangun sistem berbasis SIG yang menampilkan hasil pengelompokan secara spasial untuk mendukung analisis oleh admin dan kepala dinas. Kualitas hasil *K-means* dan *K-means* + PSO dibandingkan menggunakan *Sum of Squared Error* (SSE) dan *Davies-Bouldin Index* (DBI). Nilai kedua metrik yang lebih kecil digunakan sebagai indikator bahwa cluster memiliki tingkat kekompakan yang lebih baik dan pemisahan antarkelompok yang lebih optimal.

2. METODE PENELITIAN

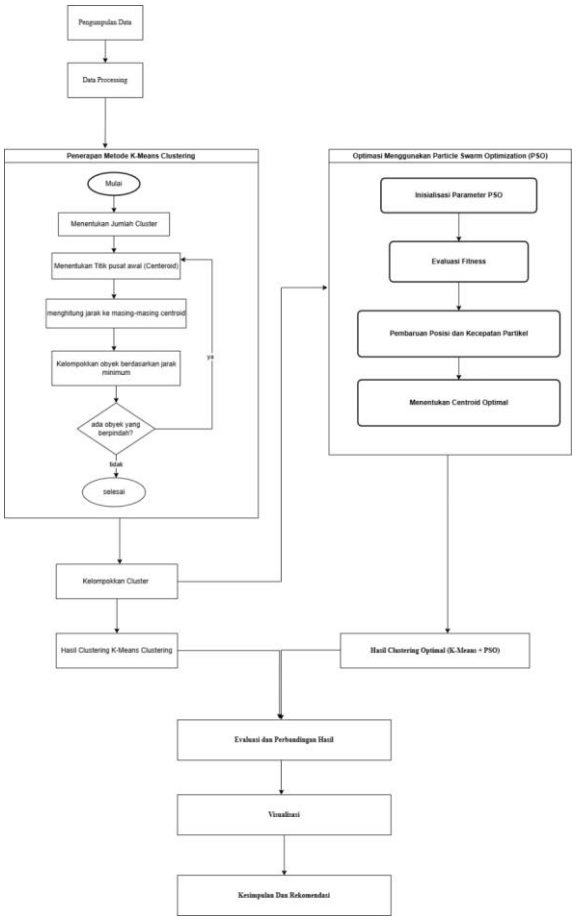
2.1 Data dan Kerangka Penelitian

Penelitian menggunakan data primer yang diperoleh dari Dinas Kependudukan dan Pencatatan Sipil Kabupaten Cirebon melalui observasi, wawancara, dan pengumpulan data historis. Data awal terdiri atas 465 baris dan 126 kolom yang mencakup data pada tingkat kabupaten, kecamatan, dan desa. Unit analisis dalam penelitian ini dibatasi pada 40 kecamatan di Kabupaten Cirebon. Data penelitian terdiri atas dua periode pengamatan, yaitu tahun 2024 dan 2025, yang dipisahkan menjadi dua dataset. Pada masing-masing tahun, digunakan lima atribut numerik

Jurnal Surya Informatika
Journal homepage: https://jurnal.umpp.ac.id/index.php/surya_informatika
P-ISSN: 2477-3042; E-ISSN: 3026-3034
Vol. 16, No. 2, November 2026, Pp. 217-227

dokumen kependudukan, yaitu Kartu Keluarga (KK), KTP, Kartu Identitas Anak (KIA), Akta Kelahiran, dan Akta Kematian. Dengan demikian, setiap dataset tahunan memiliki lima atribut yang digunakan sebagai variabel dalam proses klasterisasi. Proses klasterisasi dilakukan secara terpisah untuk tahun 2024 dan 2025, sehingga hasil pengelompokan masing-masing tahun dapat dianalisis dan dibandingkan tanpa mencampurkan karakteristik data antarperiode. Dengan pendekatan tersebut, terdapat dua proses klasterisasi, yaitu klasterisasi dataset tahun 2024 menggunakan lima atribut dan klasterisasi dataset tahun 2025 menggunakan lima atribut yang sama.

Proses analisis mengikuti *CRISP-DM* yang terdiri atas *business understanding*, *data understanding*, data *preparation*, *modeling*, *evaluation*, dan *deployment* [9]. *Business understanding* mendefinisikan kebutuhan pengelompokan wilayah dan visualisasi spasial. Data understanding mengidentifikasi struktur data, level wilayah, dan atribut yang relevan. Data preparation menghasilkan dataset numerik bersih. *Modeling* membandingkan *K-means* dengan *K-means* + PSO, *evaluation* menggunakan SSE dan DBI, sedangkan *deployment* menerapkan hasil ke dalam sistem web SIG. Alur penelitian yang digunakan ditunjukkan pada Gambar 1.



Gambar 1. Alur penelitian

Tabel 1. Karakteristik dataset penelitian

Komponen	Keterangan
Data awal	465 baris; 126 kolom
Unit analisis	40 kecamatan
Periode	2024 dan 2025
Atribut per tahun	KK, KTP, KIA, Akta Kelahiran, Akta Kematian
Kondisi akhir	Tidak ada nilai kosong dan duplikasi

2.2 Persiapan dan Normalisasi Data

Persiapan data dilakukan dengan menyeleksi 40 record kecamatan, menghapus kolom tahun dan atribut yang tidak digunakan, memeriksa nilai kosong, memeriksa duplikasi nama kecamatan, serta memastikan seluruh atribut *clustering* bertipe numerik. Validasi menunjukkan tidak terdapat nilai kosong dan tidak terdapat duplikasi kecamatan pada dataset akhir. Pemisahan data per tahun dilakukan agar perbedaan skala dan karakteristik tahun 2024 dan 2025 tidak saling memengaruhi selama proses *clustering*.

Rentang setiap atribut berbeda sehingga perhitungan jarak dapat didominasi oleh atribut dengan nilai terbesar. Oleh karena itu, data dinormalisasi ke rentang 0 sampai 1 menggunakan *Min-Max Normalization*. Normalisasi diterapkan secara terpisah pada setiap atribut dan tahun dengan persamaan (1). Nilai x menyatakan data awal, *x_min* menyatakan nilai minimum atribut, dan *x_max* menyatakan nilai maksimum atribut.

$$x' = (x - x\_min) / (x\_max - x\_min) \quad (1)$$

2.3 Pemodelan K-Means

Jumlah cluster ditetapkan sebanyak tiga berdasarkan titik tekukan pada *Elbow Method.* Pada data tahun 2024, grafik menunjukkan perubahan WCSS mulai melandai pada K = 3 dengan nilai 137,4. Ketiga cluster diberi makna Cluster 1 sebagai kategori tinggi, Cluster 2 sebagai kategori sedang, dan Cluster 3 sebagai kategori rendah. Penamaan kategori didasarkan pada nilai *centroid* masing-masing cluster setelah klasterisasi selesai, bukan urutan indeks dari algoritma. Cluster dengan *centroid* tertinggi pada seluruh atribut dokumen kependudukan disebut kategori tinggi, yang terendah disebut rendah, dan sisanya sedang.

K-means dimulai dengan penentuan *centroid* awal, kemudian setiap data dihitung jaraknya terhadap seluruh *centroid* menggunakan *Euclidean Distance* pada persamaan (2). Data ditempatkan pada cluster dengan jarak minimum. *Centroid* diperbarui menggunakan rata-rata seluruh anggota cluster sebagaimana persamaan (3). Perhitungan jarak, penempatan anggota, dan pembaruan *centroid* diulang hingga tidak terjadi perpindahan anggota atau perubahan *centroid* yang berarti.

$$d(x,c)=\sqrt{\sum_{j=1}^{m}(x_j\text{-}c_j)^2} \quad (2)$$

Jurnal Surya Informatika
Journal homepage: https://jurnal.umpp.ac.id/index.php/surya_informatika
P-ISSN: 2477-3042; E-ISSN: 3026-3034
Vol. 16, No. 2, November 2026, Pp. 217-227

$$C_j^{(t+1)} = \frac{1}{n_j} \sum_{i=1}^{n_j} X_i \quad (3)$$

2.4 Optimasi Centroid Menggunakan PSO

Pada model hibrid, satu partikel merepresentasikan seluruh *centroid*. Tiga cluster dengan lima atribut menghasilkan ruang pencarian berdimensi 15. Posisi awal partikel dibentuk dari data normalisasi yang diberi variasi terbatas pada rentang 0 sampai 1. Penelitian menggunakan 12 partikel, maksimum 8 iterasi, bobot inersia 0,7, konstanta kognitif c1 = 2, dan konstanta sosial c2 = 2. Jumlah partikel dan iterasi yang digunakan lebih rendah dibandingkan beberapa penelitian PSO yang umumnya menggunakan 30–50 partikel dan 100–500 iterasi. Pemilihan parameter tersebut disesuaikan dengan kebutuhan sistem yang digunakan untuk operasional harian, dengan mempertimbangkan efisiensi waktu komputasi. Konsekuensi dari penggunaan parameter tersebut dibahas lebih lanjut pada subbab Keterbatasan Penelitian.

Kecepatan awal setiap partikel diinisialisasi secara acak dengan batas yang terkontrol. Inisialisasi ini bertujuan agar partikel dapat mengeksplorasi ruang solusi tanpa mengalami perpindahan posisi yang terlalu jauh pada iterasi awal. Karena personal best pada awal proses sama dengan posisi awal partikel, arah pergerakan pada tahap awal lebih banyak dipengaruhi oleh komponen sosial, yaitu menuju posisi global best yang telah ditemukan.

Setiap partikel dievaluasi menggunakan SSE sebagai fungsi *fitness*. Posisi terbaik individu disimpan sebagai *personal best* dan posisi terbaik seluruh populasi disimpan sebagai *global best*. Kecepatan diperbarui dengan persamaan (4), kemudian posisi diperbarui dengan persamaan (5). Variabel w adalah bobot inersia, r1 dan r2 adalah bilangan acak, *pbest*_i adalah posisi terbaik partikel ke-i, dan *gbest* adalah posisi terbaik global. Setelah iterasi PSO selesai, *centroid global best* digunakan sebagai *centroid* awal *K-means* hingga proses *clustering* mencapai konvergensi. PSO digunakan untuk memperluas pencarian solusi dan mengurangi ketergantungan *K-means* pada pemilihan *centroid* awal [3]-[5].

$$v_i(t+1) = w v_i(t) + c_1 r_1 \left(pbest_i - x_i(t)\right) + c_2 r_2 \left(gbest - x_i(t)\right) \quad (4)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (5)$$

Tabel 2. Parameter *Particle swarm optimization*

Parameter	Nilai
Jumlah cluster	3
Jumlah partikel	12
Iterasi maksimum	8
Bobot inersia (w)	0,7
c1 dan c2	2 dan 2
Velocity awal	Acak

2.5 Evaluasi dan Implementasi Sistem

SSE mengukur jumlah kuadrat jarak setiap data terhadap *centroid* cluster-nya seperti pada persamaan (6). Nilai yang lebih kecil menunjukkan anggota cluster lebih dekat dengan pusat kelompok. DBI mengevaluasi rasio kedekatan *internal* dan jarak antarsentroid. *Scatter* S_i dihitung dari rata-rata jarak anggota terhadap *centroid*, M_ij merupakan jarak *centroid* i dan j, sedangkan R_ij = (S_i + S_j)/M_ij. DBI dihitung dengan persamaan (7); semakin kecil nilainya, semakin baik keseimbangan antara *cohesion* dan *separation* [10].

$$SSE = \sum_{k=1}^{K} \sum_{x_i \in C_k} \|x_i - c_k\|^2 \quad (6)$$

$$DBI = \frac{1}{K} \sum_{i=1}^{K} \max_{j \neq i} \left(\frac{S_i + S_j}{M_{ij}} \right) \quad (7)$$

Hasil analisis diterapkan dalam aplikasi web berbasis SIG. Model perhitungan *K-Means* dan *K-Means* + PSO diimplementasikan menggunakan *Python*, sedangkan aplikasi web dikembangkan menggunakan *phyton*, basis data *MySQL*, dan API *Leaflet* untuk visualisasi peta. Proses pengolahan dan analisis data dalam penelitian ini menggunakan kerangka *CRISP-DM*, yang meliputi pemahaman data, persiapan data, pemodelan menggunakan *K-Means* dan *K-Means* + PSO, serta evaluasi hasil *clustering*. Hasil dari proses tersebut kemudian diimplementasikan ke dalam aplikasi untuk mendukung penyajian dan pemanfaatan hasil analisis. Admin dapat mengelola data kependudukan, data training, menjalankan proses *K-Means* dan *K-Means* + PSO, melihat riwayat perhitungan, membandingkan hasil evaluasi, melihat peta SIG, dan menghasilkan laporan. Sementara itu, kepala dinas memperoleh akses untuk melihat dashboard, data, peta persebaran cluster, serta laporan. Dengan demikian, *CRISP-DM* menjadi kerangka utama yang digunakan dalam penelitian untuk mengolah data hingga menghasilkan informasi clustering yang kemudian diterapkan pada aplikasi web berbasis SIG.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Persiapan Data

Seleksi data menghasilkan 40 kecamatan untuk masing-masing tahun dan lima atribut dokumen pada setiap periode. Pemeriksaan kualitas menunjukkan seluruh *record* lengkap, unik, dan numerik. Nilai minimum dan maksimum berbeda pada setiap atribut. Sebagai contoh, tahun 2024 memiliki rentang Kartu Keluarga 10.736-33.541, KTP 22.128-74.787, KIA 2.954-14.786, Akta Kelahiran 7.642-30.824, dan Akta Kematian 367-2.880. Tahun 2025 memiliki rentang Kartu Keluarga 10.920-34.448, KTP 22.421-76.620, KIA 3.462-17.022, Akta Kelahiran 7.717-31.237, dan Akta Kematian 367-2.880. Perbedaan rentang tersebut menegaskan perlunya normalisasi sebelum jarak *Euclidean* dihitung.

Jurnal Surya Informatika
Journal homepage: https://jurnal.umpp.ac.id/index.php/surya_informatika
P-ISSN: 2477-3042; E-ISSN: 3026-3034
Vol. 16, No. 2, November 2026, Pp. 217-227

Hasil normalisasi mempertahankan pola relatif data tetapi menyetarakan skala seluruh atribut. Kecamatan dengan nilai minimum memperoleh nilai 0 dan kecamatan dengan nilai maksimum memperoleh nilai 1 pada atribut terkait. Dataset normalisasi kemudian digunakan secara identik pada *K-means* dan *K-means* + PSO sehingga perbandingan kedua metode dilakukan pada data dan jumlah cluster yang sama.

Tabel 3. Rentang atribut sebelum normalisasi

Tahun	Atribut	Minimum	Maksimum
2024	KK	10.736	33.541
2024	KTP	22.128	74.787
2024	KIA	2.954	14.786
2024	Akta lahir	7.642	30.824
2024	Akta mati	367	2.880
2025	KK	10.920	34.448
2025	KTP	22.421	76.620
2025	KIA	3.462	17.022
2025	Akta lahir	7.717	31.237
2025	Akta mati	367	2.880

3.2 Proses Perhitungan K-Means

Sebelum memperoleh hasil evaluasi SSE dan DBI, proses clustering dilakukan melalui beberapa tahapan perhitungan *K-Means*. Data yang digunakan merupakan data hasil normalisasi lima atribut dokumen kependudukan, yaitu KK, KTP, KIA, Akta Kelahiran, dan Akta Kematian.

1. Penentuan Jumlah Cluster

Jumlah cluster yang digunakan adalah K=3 yang ditetapkan dari hasil pengujian *elbow method*, yang merepresentasikan tiga kategori kebutuhan layanan dokumen kependudukan yaitu:

Tabel 4. Kategori Cluster

Cluster	Keterangan
C1	Cluster dengan rata-rata *centroid* tertinggi setelah proses clustering selesai.
C2	Cluster dengan rata-rata *centroid* di antara kategori tinggi dan rendah.
C3	Cluster dengan rata-rata *centroid* terendah setelah proses clustering selesai.

Berdasarkan analisis nilai *centroid*, Penamaan kategori tidak mengikuti nomor cluster karena indeks C1, C2, dan C3 bersifat arbitrer. Setelah *centroid* akhir diperoleh, rata-rata lima komponennya diurutkan. Pada hasil *K-Means* + PSO tahun 2024, rata-rata *centroid* C2 = 0,646 merupakan yang tertinggi, C1 = 0,415 berada di tengah, dan C3 = 0,184 merupakan yang terendah; sehingga C2 diberi label tinggi, C1 sedang, dan C3 rendah. Pola yang sama muncul pada 2025, dengan rata-rata *centroid* C2 = 0,636, C1 = 0,421, dan C3 = 0,187. Penetapan label untuk K-Means standar dilakukan dengan prosedur pemeringkatan *centroid* yang sama.

2. Inisialisasi *Centroid* Awal

Sebelum proses perhitungan jarak dilakukan, algoritma *K-Means* memerlukan titik pusat awal (*initial centroid*) sebagai acuan pengelompokan data. Pada penelitian ini, jumlah cluster ditentukan sebanyak tiga (K=3), sehingga diperlukan tiga *centroid* awal yang masing-masing merepresentasikan pusat Cluster 1, Cluster 2, dan Cluster 3.

Centroid awal diperoleh dari pemilihan data hasil normalisasi pada lima atribut dokumen kependudukan, yaitu Kartu Keluarga (KK), KTP, Kartu Identitas Anak (KIA), Akta Kelahiran, dan Akta Kematian. Data yang telah dinormalisasi memiliki rentang nilai 0 sampai 1 sehingga setiap atribut memiliki skala yang sama sebelum dilakukan perhitungan jarak *Euclidean*.

$$C_1 = (0{,}566640;\ 0{,}531819;\ 0{,}223788;\ 0{,}472315;\ 0{,}279035)$$
$$C_2 = (0{,}716593;\ 0{,}690632;\ 0{,}568382;\ 0{,}603408;\ 0{,}652328)$$
$$C_3 = (0{,}202524;\ 0{,}194492;\ 0{,}128016;\ 0{,}182450;\ 0{,}214460)$$

3. Perhitungan Jarak Euclidean K-Means

Setelah *centroid* ditentukan, setiap data kecamatan dihitung jaraknya terhadap masing-masing *centroid* menggunakan metode *Euclidean Distance*. Persamaan yang digunakan adalah:

Untuk menghitung jarak ke *centroid* pertama

$$d(x_i, c_k) = \sqrt{\sum_{j=1}^{m} \left(x_{ij} - c_{kj}\right)^2}$$

$$d(X, C_1) = \sqrt{(0{,}432-0{,}566640)^2 + (0{,}381-0{,}531819)^2 + (0{,}276-0{,}223788)^2 + (0{,}419-0{,}472315)^2 + (0{,}215-0{,}279035)^2}$$

$$d(X, C1) = 0{,}2245$$

4. Perbarui *Centroid*

Setelah seluruh data dikelompokkan berdasarkan jarak terdekat, tahap berikutnya adalah memperbarui posisi *centroid*. Pembaruan *centroid* dilakukan dengan menghitung nilai rata-rata dari seluruh data yang berada pada masing-masing cluster. Persamaan yang digunakan yaitu:

$$C_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_i$$

$$X\_1 = (0{,}432;\ 0{,}381;\ 0{,}276;\ 0{,}419;\ 0{,}215)$$
$$X\_2 = (0{,}512;\ 0{,}467;\ 0{,}312;\ 0{,}451;\ 0{,}264)$$
$$X\_3 = (0{,}586;\ 0{,}521;\ 0{,}356;\ 0{,}489;\ 0{,}298)$$

$$C_1 = \left(\frac{0{,}432+0{,}512+0{,}586}{3};\ \frac{0{,}381+0{,}467+0{,}521}{3};\ \frac{0{,}276+0{,}312+0{,}356}{3};\ \frac{0{,}419+0{,}451+0{,}489}{3};\ \frac{0{,}215+0{,}264+0{,}298}{3}\right)$$

$$C1 = (0{,}510;\ 0{,}456;\ 0{,}315;\ 0{,}453;\ 0{,}259)$$

Centroid baru tersebut kemudian digunakan sebagai pusat cluster pada iterasi berikutnya untuk menghitung kembali jarak setiap data. Proses pembaruan *centroid* dilakukan secara berulang hingga

Jurnal Surya Informatika
Journal homepage: https://jurnal.umpp.ac.id/index.php/surya_informatika
P-ISSN: 2477-3042; E-ISSN: 3026-3034
Vol. 16, No. 2, November 2026, Pp. 217-227

tidak terjadi perubahan anggota cluster atau nilai *centroid*.

3.3 Pehitungan K-Means + PSO

1. Inisialisasi Partikel

Posisi awal partikel diperoleh dari *centroid* hasil normalisasi. Contoh satu partikel direpresentasikan sebagai:

$$Xi=(C1,C2,C3)$$
$$X_i=(0{,}566640;\ 0{,}531819;\ 0{,}223788;\ 0{,}472315;\ 0{,}279035;\ 0{,}716593;\ 0{,}690632;\ 0{,}568382;\ 0{,}603408;\ 0{,}652328;\ 0{,}202524;\ 0{,}194492;\ 0{,}128016;\ 0{,}182450;\ 0{,}214460)$$

Pada tahap awal, nilai kecepatan (*velocity*) setiap partikel diinisialisasi dengan nilai acak.

$$Vi=(-0{,}019730,-0{,}019730,\ldots\ldots -0{,}019730)$$

2. Menghitung Nilai *Fitness*

Setiap partikel dievaluasi menggunakan fungsi *fitness* berupa nilai *Sum of Squared Error* (SSE). Nilai *fitness* menunjukkan kualitas *centroid* yang dihasilkan. Semakin kecil nilai SSE, semakin baik posisi partikel tersebut.

$$Fitness=SSE=\sum_{j=1}^{k}\sum_{i=1}^{n}(x_i-c_j)^2\ Fitness(X_1)=4{,}120542$$

Kemudian nilai tersebut dibandingkan dengan partikel lainnya untuk menentukan posisi terbaik.

3. Penentuan *Personal Best* (PBest) dan *Global Best* (GBest)

Setiap partikel menyimpan posisi terbaik berdasarkan nilai fitness terkecil yang pernah diperoleh.

$$PBesti=Xi$$

Apabila nilai *fitness* partikel saat ini lebih kecil dibandingkan nilai sebelumnya, maka posisi tersebut menjadi PBest baru.
Selanjutnya, seluruh PBest dibandingkan untuk memperoleh posisi terbaik global:

$$GBest=min(PBest1,PBest2,\ldots,PBestn)$$

Tabel 5. Penentuan Nilai Gbest

Partikel	Nilai SSE	Keterangan
P1	*4,120542*	-
P2	*3,950221*	Gbest
P3	*4,210331*	-

Berdasarkan tabel tersebut, partikel kedua memiliki nilai SSE terkecil sehingga digunakan sebagai GBest.

4. Pembaruan *Velocity*

Setelah nilai PBest dan GBest diperoleh, kecepatan partikel diperbarui menggunakan persamaan:

$$V_i(t+1)=wV_i(t)+c_1r_1(PBest_i-X_i(t))+c_2r_2(GBest-X_i(t))$$

Contoh pembaruan satu dimensi *centroid*:

$$V_1(1)=\&(0{,}7)(-0{,}019730)+(2)(0{,}5)(PBest\text{-}X)\ \&+(2)(0{,}5)(GBest\text{-}X)$$

$$V_1(1)=0{,}5(PBest\text{-}X)+0{,}5(GBest\text{-}X)$$

Hasil tersebut menjadi kecepatan baru partikel.

5. Perbarui Posisi Partikel

Setelah *velocity* diperoleh, posisi partikel diperbarui menggunakan:

$$X_i(t+1)=X_i(t)+V_i(t+1)$$

$$X\_1(t+1)=0{,}566640+0{,}012500$$
$$X\_1(t+1)=0{,}579140$$

Nilai posisi baru tersebut digunakan kembali untuk menghitung *fitness* pada iterasi berikutnya.

6. *Centroid* Optimal Hasil PSO

Proses pembaruan *velocity* dan posisi dilakukan hingga jumlah iterasi maksimum tercapai. Posisi partikel dengan nilai fitness terbaik (GBest) digunakan sebagai *centroid* optimal untuk proses *K-Means* berikutnya.

$$C_{optimal}=GBest$$

$$C_{optimal}=argmin(Fitness(X_i)))$$

Hasil *centroid* optimal tersebut kemudian digunakan kembali pada tahap clustering *K-Means* sehingga menghasilkan pembagian cluster akhir dan selanjutnya dievaluasi menggunakan SSE dan DBI.

3.4 Hasil Clustering K-Means

K-means standar mencapai konvergensi setelah proses pembaruan *centroid* dan perpindahan anggota berhenti. Pada tahun 2024, metode ini menghasilkan 13 kecamatan pada Cluster 1, 17 kecamatan pada Cluster 2, dan 10 kecamatan pada Cluster 3. Pada tahun 2025, komposisinya berubah menjadi 15 kecamatan pada Cluster 1, 16 kecamatan pada Cluster 2, dan 9 kecamatan pada Cluster 3. Perubahan distribusi tersebut tidak secara langsung menunjukkan adanya kebijakan atau peristiwa tertentu, karena penelitian ini tidak menganalisis faktor eksternal yang menyebabkan perubahan tersebut. Perubahan lebih tepat dipahami sebagai akibat dari perubahan nilai lima atribut dokumen kependudukan pada masing-masing kecamatan antara tahun 2024 dan 2025. Perubahan nilai atribut menyebabkan posisi relatif kecamatan terhadap *centroid* berubah, sehingga beberapa kecamatan dapat berpindah ke cluster yang memiliki karakteristik lebih dekat pada periode berikutnya.

Distribusi anggota pada *K-means* standar juga dipengaruhi oleh posisi *centroid* awal yang digunakan dalam proses klasterisasi. Hasil evaluasi tahun 2024 menunjukkan kontribusi SSE masing-masing cluster

Jurnal Surya Informatika
Journal homepage: https://jurnal.umpp.ac.id/index.php/surya_informatika
P-ISSN: 2477-3042; E-ISSN: 3026-3034
Vol. 16, No. 2, November 2026, Pp. 217-227

sebesar 1,311, 1,080, dan 1,950, dengan total SSE sebesar 4,341. Pada tahun 2025, kontribusi SSE masing-masing cluster sebesar 1,570, 0,853, dan 1,631, dengan total SSE sebesar 4,053. Perbedaan kontribusi SSE antarcluster menunjukkan bahwa tingkat kekompakan anggota dalam setiap cluster tidak sama. Nilai tersebut selanjutnya digunakan sebagai *baseline* untuk membandingkan perubahan kualitas cluster setelah dilakukan optimasi *centroid* menggunakan PSO.

Tabel 6. Distribusi anggota cluster

Tahun	Metode	C1	C2	C3
2024	*K-Means*	13	17	10
2024	*K-means* + PSO	14	10	16
2025	*K-Means*	15	16	9
2025	*K-means* + PSO	14	10	16

3.5 Hasil Clustering K-means + PSO

PSO menelusuri kandidat *centroid* melalui pembaruan posisi berdasarkan *personal best* dan *global best*. Pada kedua *dataset*, posisi yang menghasilkan *fitness* terbaik dipakai sebagai *centroid* awal K-Means. Proses *K-means* lanjutan berhenti ketika seluruh *centroid* memiliki pergeseran nol dan tidak terjadi perpindahan anggota. Hasil akhir menghasilkan distribusi yang konsisten, yaitu 14 kecamatan pada Cluster 1, 10 kecamatan pada Cluster 2, dan 16 kecamatan pada Cluster 3 untuk tahun 2024 maupun 2025.

Centroid hasil PSO tahun 2024 untuk Cluster 1 adalah 0,566640; 0,531819; 0,223788; 0,472315; dan 0,279035. Cluster 2 memiliki *centroid* 0,716593; 0,690632; 0,568382; 0,603408; dan 0,652328, sedangkan Cluster 3 memiliki *centroid* 0,202524; 0,194492; 0,128016; 0,182450; dan 0,214460. Pada tahun 2025, *centroid* Cluster 1 adalah 0,560839; 0,528192; 0,257380; 0,477311; dan 0,279035; Cluster 2 adalah 0,707502; 0,684005; 0,530715; 0,606322; dan 0,652328; sedangkan Cluster 3 adalah 0,200086; 0,192401; 0,143902; 0,184309; dan 0,214460. Nilai *centroid* disimpan bersama hasil cluster untuk menjamin keterlacakan proses perhitungan pada sistem.

Perbedaan jumlah iterasi memperlihatkan karakteristik proses kedua metode. *K-means* standar tahun 2024 mencapai kondisi akhir pada iterasi keempat, sedangkan model dengan *centroid* hasil PSO mencapai konvergensi *K-means* pada iterasi keenam. Pada tahun 2025, *K-means* standar berhenti pada iterasi ketiga dan *K-means* + PSO berhenti pada iterasi kelima. Jumlah iterasi *K-means* lanjutan tidak digunakan sebagai satu-satunya ukuran performa karena PSO menambahkan tahap pencarian kandidat *centroid*. Kualitas akhir tetap ditentukan melalui SSE dan DBI.

Tabel 7. Iterasi konvergensi K-Means

Tahun	Metode	Iterasi akhir
2024	*K-Means*	4
2024	*K-means* + PSO	6
2025	*K-Means*	3
2025	*K-means* + PSO	5

3.6 Perbandingan SSE dan DBI

Perbandingan pada Tabel 8 menunjukkan penurunan kedua metrik setelah *centroid* dioptimasi. Pada tahun 2024, SSE turun dari 4,341413 menjadi 3,778542 dan DBI turun dari 1,230109 menjadi 1,050285. Pada tahun 2025, SSE turun dari 4,053480 menjadi 3,783298 dan DBI turun dari 1,166977 menjadi 1,107336. Grafik perbandingan keseluruhan ditampilkan pada Gambar 2. Karena SSE dan DBI sama-sama menggunakan orientasi minimisasi, hasil tersebut menempatkan *K-means* + PSO sebagai metode terbaik pada kedua periode.

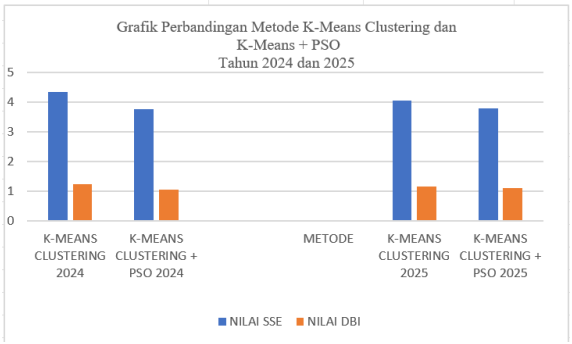
Penurunan SSE menunjukkan bahwa *centroid* hasil optimasi menghasilkan jarak kuadrat total yang lebih kecil antara anggota dan pusat cluster. Penurunan DBI menunjukkan rasio *scatter internal* terhadap jarak antarsentroid juga membaik. Pada tahun 2024, DBI *K-means* dipengaruhi rasio maksimum 1,447594 pada Cluster 1 dan Cluster 3, sedangkan *K-means* + PSO menurunkannya menjadi 1,139290 dan 0,872276. Pada tahun 2025, rasio maksimum *K-means* sebesar 1,387139, sedangkan model hibrid menghasilkan 1,227266 dan 0,867475. Perubahan ini menjelaskan bahwa PSO tidak hanya menurunkan *error internal,* tetapi juga memperoleh konfigurasi pusat cluster dengan pemisahan yang lebih jelas.

Kinerja yang lebih baik berasal dari mekanisme pencarian populasi. *K-means* standar bergerak dari satu kombinasi *centroid* awal, sedangkan PSO mengevaluasi 12 kandidat yang masing-masing merepresentasikan 15 dimensi *centroid*. Informasi personal best mempertahankan pengalaman terbaik tiap partikel dan global best mengarahkan populasi ke solusi dengan SSE terendah. Setelah *centroid* optimal ditemukan, *K-means* tetap digunakan untuk memperbarui pusat berdasarkan anggota aktual. Kombinasi eksplorasi PSO dan pemutakhiran lokal *K-means* mengurangi kemungkinan hasil berhenti pada konfigurasi *centroid* yang kurang baik.

Tabel 8. Perbandingan hasil evaluasi

Tahun	Metode	SSE	DBI
2024	*K-Means*	4,341413	1,230109
2024	*K-means* + PSO	3,778542	1,050285
2025	*K-Means*	4,053480	1,166977
2025	*K-means* + PSO	3,783298	1,107336

Jurnal Surya Informatika
Journal homepage: https://jurnal.umpp.ac.id/index.php/surya_informatika
P-ISSN: 2477-3042; E-ISSN: 3026-3034
Vol. 16, No. 2, November 2026, Pp. 217-227



Gambar 2. Perbandingan SSE dan DBI tahun 2024-2025

Tabel 9. Kontribusi SSE setiap cluster

Tahun	Metode	C1	C2	C3
2024	*K-Means*	1,310668	1,080352	1,950393
2024	*K-means* + PSO	1,352066	1,576565	0,849911
2025	*K-Means*	1,569794	0,853104	1,630583
2025	*K-means* + PSO	1,357669	1,572524	0,853104

Tabel 10. Nilai maksimum komponen DBI

Tahun	Metode	D1	D2	D3
2024	*K-Means*	1,447594	0,795140	1,447594
2024	*K-means* + PSO	1,139290	1,139290	0,872276
2025	*K-Means*	1,387139	0,726655	1,387139
2025	*K-means* + PSO	1,227266	1,227266	0,867475

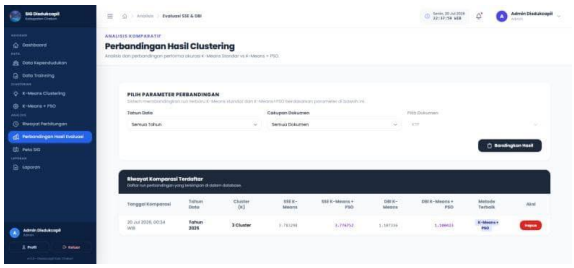
3.7 Implementasi SIG dan Pengujian Sistem

Sistem yang dibangun menyediakan dua jenis pengguna. Admin mengelola data melalui impor *Excel* atau input manual, menjalankan *K-means* dan *K-means* + PSO, melihat riwayat perhitungan, membandingkan SSE dan DBI, menampilkan peta, serta mencetak laporan. Kepala dinas dapat melihat ringkasan dashboard, data kependudukan, peta SIG, dan laporan tanpa mengubah data. Penyimpanan hasil perhitungan memungkinkan pengguna menelusuri tahun, metode, parameter, distribusi cluster, serta nilai evaluasi yang telah dihasilkan.

Visualisasi SIG menggunakan peta Kabupaten Cirebon dengan warna berbeda untuk setiap cluster. Pengguna dapat melihat lokasi kecamatan dan kategori hasil *clustering* secara langsung sehingga pola wilayah tidak hanya dibaca sebagai angka. Tampilan peta pada Gambar 3 mengintegrasikan hasil analitik dengan batas wilayah dan informasi geografis. Penyajian ini menjawab kebutuhan penelitian untuk memberikan gambaran spasial yang lebih informatif daripada tabel administratif biasa.

Halaman perbandingan hasil menyatukan tahun data, jumlah cluster, nilai SSE, nilai DBI, dan rekomendasi metode terbaik dalam satu tampilan. Pengguna dapat memilih tahun serta cakupan dokumen, kemudian sistem mengambil riwayat perhitungan yang sesuai. Desain ini menjaga pemisahan antara proses analitik dan interpretasi hasil: perhitungan disimpan sebagai riwayat, sedangkan antarmuka perbandingan menyajikan ringkasan yang diperlukan untuk keputusan.

Pengujian *black box* dilakukan pada seluruh fungsi utama. Pada akun admin, sepuluh skenario mencakup login, pengelolaan data, pengelolaan data *training*, K-Means, *K-means* + PSO, riwayat perhitungan, perbandingan evaluasi, peta SIG, laporan, dan *logout*. Pada akun kepala dinas, enam skenario mencakup login, *dashboard*, data kependudukan, peta, laporan, dan *logout.* Seluruh skenario memberikan keluaran sesuai harapan dan dinyatakan sukses. Dengan demikian, sistem tidak hanya menghasilkan model *clustering* yang lebih baik, tetapi juga menyediakan fungsi operasional untuk mengelola, menampilkan, dan melaporkan hasil analisis.



Gambar 3. Tampilan perbandingan hasil *clustering*

Gambar 4. Visualisasi peta hasil *clustering*

Peta hasil clustering pada Gambar 4 menunjukkan bahwa persebaran cluster di Kabupaten Cirebon tidak membentuk satu zona geografis yang sepenuhnya terkonsentrasi pada wilayah tertentu. Cluster tinggi, sedang, dan rendah tersebar pada beberapa kecamatan dengan posisi geografis yang berbeda. Berdasarkan visualisasi peta, cluster tinggi tidak hanya berada pada satu kawasan tertentu, seperti pusat kota atau wilayah pesisir, tetapi dapat ditemukan pada beberapa lokasi yang tidak selalu berdekatan. Hal ini menunjukkan bahwa tingkat kebutuhan dokumen kependudukan tidak secara langsung mengikuti kedekatan geografis suatu kecamatan.

Jurnal Surya Informatika
Journal homepage: https://jurnal.umpp.ac.id/index.php/surya_informatika
P-ISSN: 2477-3042; E-ISSN: 3026-3034
Vol. 16, No. 2, November 2026, Pp. 217-227

Secara spasial, terdapat kecamatan yang berdekatan tetapi memiliki kategori cluster yang berbeda. Sebaliknya, beberapa kecamatan yang letaknya relatif berjauhan dapat berada dalam cluster yang sama. Pola tersebut terjadi karena proses clustering didasarkan pada kemiripan lima atribut dokumen kependudukan, yaitu KK, KTP, KIA, Akta Kelahiran, dan Akta Kematian, bukan berdasarkan koordinat atau jarak antarwilayah. Dengan demikian, pola geografis pada peta lebih tepat dipahami sebagai persebaran lokasi dari kelompok yang memiliki karakteristik dokumen serupa, bukan sebagai pembentukan zona spasial berdasarkan kedekatan wilayah.

Dari sisi karakteristik wilayah, hasil visualisasi juga tidak menunjukkan dasar yang cukup untuk menyimpulkan bahwa cluster tertentu secara khusus terkonsentrasi pada wilayah pesisir, pusat pemerintahan, atau wilayah dengan karakteristik geografis tertentu. Hal ini disebabkan variabel geografis seperti jarak dari pusat kota, ketinggian wilayah, kepadatan penduduk, maupun klasifikasi pesisir dan nonpesisir tidak digunakan sebagai atribut dalam proses *clustering*. Oleh karena itu, peta SIG dalam penelitian ini berfungsi untuk menunjukkan lokasi geografis dari hasil pengelompokan, sedangkan pembentukan cluster ditentukan oleh kemiripan data dokumen kependudukan.

Implikasi spasial dari hasil tersebut adalah bahwa Disdukcapil Kabupaten Cirebon tidak dapat menentukan prioritas pelayanan hanya berdasarkan kedekatan geografis antar kecamatan. Kecamatan yang berada dalam cluster tinggi dapat menjadi prioritas berdasarkan karakteristik dokumen yang dimilikinya meskipun lokasinya tidak berdekatan dengan kecamatan tinggi lainnya. Informasi spasial kemudian dapat digunakan sebagai pendukung dalam menentukan jangkauan pelayanan, rute pelayanan, maupun pengelompokan wilayah operasional yang berdekatan. Dengan demikian, integrasi *clustering* dan SIG memberikan informasi mengenai tingkat kebutuhan relatif sekaligus lokasi wilayah yang memiliki karakteristik tersebut.

Tabel 11. Rincian pengujian black box

Aktor	Fungsi	Hasil
Admin	Login administrator	Sukses
Admin	Kelola data kependudukan	Sukses
Admin	Kelola data training	Sukses
Admin	Perhitungan K-Means	Sukses
Admin	Perhitungan *K-means* + PSO	Sukses
Admin	Riwayat perhitungan	Sukses
Admin	Perbandingan evaluasi	Sukses
Admin	Peta SIG	Sukses
Admin	Cetak laporan	Sukses
Admin	Logout	Sukses
Kepala Dinas	Login	Sukses
Kepala Dinas	Dashboard	Sukses
Kepala Dinas	Data kependudukan	Sukses
Kepala Dinas	Peta SIG	Sukses
Kepala Dinas	Laporan	Sukses
Kepala Dinas	Logout	Sukses

3.8 Interpretasi dan Implikasi Hasil

Konsistensi distribusi *K-means* + PSO sebesar 14, 10, dan 16 kecamatan pada kedua tahun menunjukkan bahwa pencarian *centroid* menghasilkan struktur kelompok yang lebih stabil pada data penelitian. Stabilitas tersebut tidak berarti nilai dokumen tahun 2024 dan 2025 sama, karena *centroid* hasil optimasi tetap mengalami perubahan pada setiap atribut. Temuan ini memperlihatkan bahwa PSO mempertahankan pola pengelompokan umum sambil menyesuaikan posisi pusat cluster terhadap perubahan data tahunan.

Kategori tinggi, sedang, dan rendah pada sistem merupakan hasil perbandingan relatif antarkecamatan berdasarkan lima atribut yang telah dinormalisasi. Oleh karena itu, hasil cluster perlu dibaca bersama data asli, *centroid*, dan tahun analisis yang tersimpan dalam riwayat perhitungan. Peta SIG kemudian menambahkan konteks lokasi agar pengguna dapat mengenali keterkaitan antara kategori cluster dan persebaran wilayah, sedangkan tabel evaluasi menunjukkan dasar pemilihan metode secara kuantitatif.

Integrasi proses tersebut membentuk alur analisis yang utuh: data dibersihkan dan dinormalisasi, dua metode *clustering* dijalankan, kualitasnya dibandingkan menggunakan SSE dan DBI, lalu hasil terbaik disajikan pada peta dan laporan. Sistem tidak melakukan prediksi jumlah dokumen, tetapi menyediakan pengelompokan deskriptif yang dapat digunakan sebagai bahan analisis berkala. Dengan penyimpanan riwayat, hasil antarperiode dapat ditelusuri tanpa mengulang interpretasi dari awal.

Tabel 12. Keluaran analisis pada sistem

Keluaran	Informasi yang disajikan
Hasil *clustering*	Tahun, metode, anggota, dan kategori cluster
Evaluasi	Nilai SSE, DBI, dan metode terbaik
Peta SIG	Lokasi kecamatan dan warna kategori cluster
Riwayat	Parameter, *centroid*, distribusi, dan hasil evaluasi
Laporan	Ringkasan hasil yang dapat dicetak

3.9 Perbandingan dengan Penelitian Terdahulu

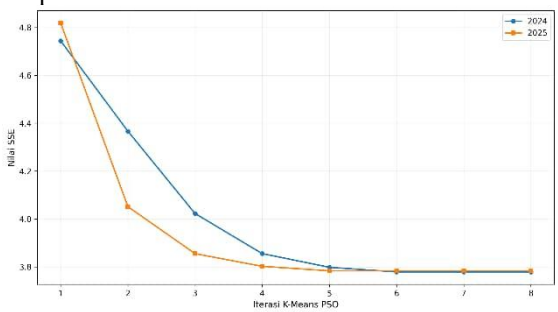
Hasil penelitian menunjukkan bahwa penerapan PSO pada *K-Means* mampu meningkatkan kualitas hasil *clustering* melalui proses optimasi *centroid*. Hal ini ditunjukkan oleh penurunan nilai SSE setelah proses optimasi. Pada penelitian terdahulu,

Jurnal Surya Informatika
Journal homepage: https://jurnal.umpp.ac.id/index.php/surya_informatika
P-ISSN: 2477-3042; E-ISSN: 3026-3034
Vol. 16, No. 2, November 2026, Pp. 217-227

Alam dan Everhard [3] melaporkan penurunan SSE dari 17,288 menjadi 17,255 atau sekitar 0,19%. Sementara itu, Firmansyah, Nugroho, dan Arif [11]. memperoleh penurunan SSE dari 3,1146 menjadi 2,5882 atau sekitar 16,91%, serta penurunan DBI dari 0,7063 menjadi 0,5185 atau sekitar 26,84%.

Pada penelitian ini, berdasarkan hasil pengujian, nilai SSE juga mengalami penurunan setelah dilakukan optimasi menggunakan PSO. Berdasarkan grafik konvergensi, nilai SSE pada tahun 2024 mengalami penurunan dari sekitar 4,74 pada iterasi pertama menjadi 3,78 pada iterasi kedelapan, sedangkan pada tahun 2025 menurun dari sekitar 4,81 menjadi 3,78. Dengan demikian, penurunan tersebut menunjukkan bahwa PSO mampu memperbaiki posisi *centroid* sehingga menghasilkan nilai SSE yang lebih rendah.

Perbandingan tersebut menunjukkan bahwa arah peningkatan kinerja penelitian ini konsisten dengan penelitian terdahulu, yaitu PSO mampu menghasilkan solusi *clustering* dengan nilai SSE yang lebih rendah dibandingkan kondisi awal. Namun, nilai absolut maupun persentase penurunan SSE tidak dapat digunakan untuk menyatakan bahwa metode pada penelitian ini lebih baik atau lebih signifikan dibandingkan penelitian terdahulu secara langsung, karena masing-masing penelitian menggunakan dataset, jumlah objek, jumlah atribut, skala data, jumlah cluster, dan parameter algoritma yang berbeda. Oleh karena itu, perbandingan dengan penelitian terdahulu digunakan untuk menunjukkan konsistensi efektivitas PSO dalam mengoptimalkan *centroid*, bukan untuk menentukan peringkat kinerja antarpenelitian.



Gambar 5. Grafik Konvergensi PSO berdasarkan Nilai SSE

Pada Gambar 5, nilai SSE menunjukkan kecenderungan menurun pada beberapa iterasi awal dan kemudian menjadi relatif stabil pada iterasi berikutnya. Pada tahun 2024, nilai SSE turun dari sekitar 4,74 pada iterasi pertama hingga mendekati 3,78 pada iterasi kedelapan. Sementara itu, pada tahun 2025 nilai SSE turun dari sekitar 4,81 hingga mendekati 3,78. Penurunan tersebut menunjukkan bahwa proses pencarian solusi oleh PSO berhasil menemukan posisi *centroid* yang semakin baik selama iterasi berlangsung.

Penurunan yang semakin kecil pada iterasi akhir menunjukkan bahwa solusi mulai mengalami konvergensi, yaitu perbaikan nilai SSE menjadi relatif kecil dibandingkan iterasi sebelumnya. Kondisi tersebut menunjukkan bahwa PSO tidak sekadar menjalankan iterasi, tetapi melakukan proses pencarian solusi secara bertahap hingga memperoleh posisi *centroid* yang relatif stabil.

Selain nilai terbaik (*Global Best*/GBest), visualisasi konvergensi idealnya menggunakan nilai SSE terbaik dari populasi pada setiap iterasi atau rata-rata SSE seluruh partikel. Penggunaan nilai tersebut dapat memperlihatkan proses optimasi secara lebih jelas, mulai dari kondisi awal populasi hingga solusi terbaik pada iterasi terakhir. Dengan demikian, grafik konvergensi menjadi bukti visual bahwa proses PSO benar-benar melakukan optimasi terhadap *centroid* K-Means.

Penelitian ini memiliki beberapa keterbatasan. Pertama, parameter PSO yang digunakan terdiri atas 12 partikel dan 8 iterasi. Pemilihan 12 partikel dilakukan dengan mempertimbangkan jumlah data yang relatif terbatas, yaitu 40 kecamatan, sehingga jumlah partikel tersebut dinilai cukup untuk menyediakan beberapa kandidat solusi *centroid* tanpa meningkatkan beban komputasi secara berlebihan. Sementara itu, 8 iterasi dipilih karena proses optimasi pada pengujian penelitian telah menunjukkan kecenderungan menuju solusi yang stabil, sehingga penambahan iterasi tidak menjadi fokus utama penelitian. Meskipun demikian, penggunaan jumlah partikel dan iterasi yang terbatas menyebabkan ruang pencarian PSO belum dieksplorasi secara maksimal, sehingga masih terdapat kemungkinan diperolehnya solusi *centroid* dengan nilai SSE dan DBI yang lebih rendah apabila digunakan parameter yang lebih besar atau dilakukan optimasi parameter PSO. Kedua, data penelitian hanya mencakup Kabupaten Cirebon dengan dua periode pengamatan, yaitu 2024 dan 2025, sehingga hasil penelitian belum dapat digeneralisasikan secara langsung ke wilayah lain yang memiliki karakteristik kependudukan berbeda. Penelitian selanjutnya dapat menggunakan periode pengamatan yang lebih panjang, wilayah yang lebih beragam, serta melakukan pengujian terhadap variasi parameter PSO untuk memperoleh hasil optimasi yang lebih komprehensif.

4. KESIMPULAN

Penelitian ini berhasil mengelompokkan persebaran lima jenis dokumen kependudukan pada 40 kecamatan di Kabupaten Cirebon ke dalam tiga cluster dan mengintegrasikan hasilnya dengan Sistem Informasi Geografis (SIG). Metode K-means menghasilkan distribusi cluster yang berbeda antara tahun 2024 dan 2025, sedangkan K-means + PSO menghasilkan distribusi yang lebih konsisten, yaitu 14, 10, dan 16 kecamatan pada masing-masing cluster. Hasil evaluasi menunjukkan bahwa penerapan PSO mampu meningkatkan kualitas clustering dengan

menurunkan SSE sebesar 12,97% dan DBI sebesar 14,62% pada tahun 2024, serta SSE sebesar 6,67% dan DBI sebesar 5,11% pada tahun 2025. Penurunan SSE dan DBI menunjukkan bahwa PSO mampu menghasilkan cluster yang lebih kompak dan memiliki kemiripan karakteristik yang lebih baik. Bagi Dinas Kependudukan dan Pencatatan Sipil Kabupaten Cirebon, hasil ini dapat digunakan sebagai informasi pendukung untuk mengenali karakteristik kebutuhan dokumen pada setiap kelompok kecamatan, sehingga perencanaan pelayanan tidak perlu dilakukan secara seragam, tetapi dapat disesuaikan berdasarkan kondisi masing-masing cluster. Dengan demikian, hasil *clustering* dapat membantu menentukan prioritas wilayah, menyesuaikan distribusi petugas, mengatur penjadwalan pelayanan, serta mendukung perencanaan pelayanan administrasi kependudukan yang lebih terarah.

Kontribusi baru penelitian ini dibandingkan penelitian sebelumnya terletak pada integrasi tiga aspek dalam satu sistem, yaitu optimasi *centroid* menggunakan PSO, clustering persebaran dokumen kependudukan, dan visualisasi hasil secara spasial melalui SIG. Penelitian sebelumnya umumnya menerapkan metode clustering untuk membentuk kelompok berdasarkan karakteristik data atau menggunakan SIG untuk menampilkan persebaran wilayah, tetapi belum mengintegrasikan K-means yang dioptimasi PSO dengan visualisasi SIG untuk menganalisis persebaran dokumen kependudukan pada tingkat kecamatan di Kabupaten Cirebon dalam satu sistem pendukung keputusan. Penelitian ini tidak hanya menghasilkan nilai cluster secara numerik, tetapi juga menerjemahkan hasil tersebut menjadi informasi spasial yang dapat langsung dipahami oleh pihak Disdukcapil. Sistem yang dikembangkan menyediakan pengelolaan data, proses clustering K-means dan K-means + PSO, evaluasi SSE dan DBI, perbandingan hasil, peta persebaran cluster, riwayat perhitungan, serta laporan. Oleh karena itu, penelitian ini memberikan kontribusi berupa pendekatan analisis yang menghubungkan peningkatan kualitas clustering dengan kebutuhan pemetaan dan pengambilan keputusan pelayanan administrasi kependudukan, sehingga hasil penelitian lebih aplikatif bagi Disdukcapil Kabupaten Cirebon.

5. UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Dinas Kependudukan dan Pencatatan Sipil Kabupaten Cirebon atas izin, data, dan dukungan selama pelaksanaan penelitian, serta kepada Program Studi Sistem Informasi Universitas Catur Insan Cendekia atas arahan akademik yang diberikan.

REFERENSI

[1] F. Febriansyah and S. Muntari, "Penerapan Algoritma K-Means untuk Klasterisasi Penduduk Miskin pada Kota Pagar Alam," 2023. https://doi.org/10.14421/jiska.2023.8.1.66-77

[2] S. N. Mayasari and J. Nugraha, "Implementasi K-Means Cluster Analysis untuk Mengelompokkan Kabupaten/Kota Berdasarkan Data Kemiskinan di Provinsi Jawa Tengah Tahun 2022," 2023. [Online]. https://doi.org/10.24002/konstelasi.v3i2.7200

[3] R. Gesit Prasasti Alam and Y. Everhard, "Optimasi K-Means Dengan *Particle Swarm Optimization* Dalam Penentuan Titik Awal Pusat Klaster Data Telekomunikasi K-Means Optimization with *Particle Swarm Optimization* In Determining The Starting Point Of Cluster Centers of Telecommunication Data." https://doi.org/10.62411/tc.v23i1.9743

[4] F. Charisma Hayatina, S. H. Wijaya, M. Kusuma, and D. Hardhienata, "Pengelompokan Publikasi Ilmiah Berdasarkan Bidang Kepakaran Menggunakan Latent Dirichlet Allocation dan Normalized PSO-K-means Clustering of Scientific Publications Based on Field of Expertise Using Latent Dirichlet Allocation and Normalized PSO-K-means." [Online]. Available: http://journal.ipb.ac.id/index.php/jika https://doi.org/10.29244/jika.10.2.121-132

[5] M. W. Ouertani, G. Manita, and O. Korbaa, "Automatic Data Clustering Using Hybrid Chaos Game Optimization with *Particle Swarm Optimization* Algorithm," in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 2677–2687. doi: 10.1016/j.procs.2022.09.326.

[6] F. Valerian and S. Yulianto, "IDENTIFICATION OF THE COVID-19 DISTRIBUTION AREA ON THE ISLAND OF KALIMANTAN USING THE K-MEANS SPATIAL CLUSTERING METHOD," *Jurnal Teknik Informatika (Jutif)*, vol. 3, no. 4, pp. 839–346, Aug. 2022, doi: 10.20884/1.jutif.2022.3.4.314.

[7] R. Ghimire, S. Dhakal, M. Sapkota, Y. Kc, and R. Shrestha, "Web Mapping on Land Cover Change of Kathmandu, Lalitpur and Bhaktapur District," 2022. https://doi.org/10.3126/njg.v21i1.50887

[8] O. Arifin and A. R. Supriyatna, "Sistem Informasi Geografis untuk Pemetaan Lahan Kakao Menggunakan Leaflet JS dan GeoJSON," Jurnal Teknoinfo, vol. 17, no. 1, p. 364, 2023, doi: 10.33365/jti.v17i1.2397.

[9] C. E. D. Vanegas, J. C. G. Mejía, F. A. V. Agudelo, and D. E. S. Duran, "A Representation Based on Essence for the *CRISP-DM* Methodology," *Computacion y Sistemas*, vol. 27, no. 3, pp. 675–689, 2023, doi: 10.13053/CyS-27-3-3446. https://doi.org/10.1080/09537287.2023.2234882

[10] N. Shalsadilla *et al.*, "Penentuan Jumlah Cluster Optimum Menggunakan Davies Bouldin Index dalam Pengelompokan Wilayah Kemiskinan di Indonesia," 2023. https://doi.org/10.29313/statistika.v23i1.1743

[11] A. L. Firmansyah, B. I. Nugroho, and Z. Arif, "Optimasi K-Means Clustering Pada Data Harga Mangga Menggunakan Particle Swarm Optimization," *Jurnal Teknologi Sistem Informasi*, vol. 6, no. 2, pp. 245–259, Sep. 2025, doi: 10.35957/jtsi.v6i2.13158.