



PENERAPAN METODE ST-DBSCAN MENGGUNAKAN ALGORITMA PARTICLE SWARM OPTIMIZATION PADA KLASTERISASI IPLM PROVINSI DI INDONESIA

ALGHIFARID B. MOHAMAD¹, SALMUN K. NASIB^{2*}, SITI NURMARDIA ABDUSSAMAD³, HASAN S. PANIGORO⁴,
DJIHAD WUNGGULI⁵

^{1,2,3} Program Studi Statistika, Jurusan Matematika, Universitas Negeri Gorontalo

^{4,5} Program Studi Matematika, Jurusan Matematika, Universitas Negeri Gorontalo

*Penulis korespondensi: salmun@ung.ac.id

Abstrak

Indeks Pembangunan Literasi Masyarakat (IPLM) merupakan ukuran terhadap upaya yang dilakukan oleh pemerintah daerah (baik di tingkat provinsi maupun kabupaten/kota) dalam membina dan mengembangkan perpustakaan sebagai sarana pembelajaran sepanjang hayat untuk membangun budaya literasi di masyarakat. Penelitian ini bertujuan untuk mengelompokkan Provinsi di Indonesia berdasarkan elemen-elemen yang mempengaruhi IPLM menggunakan metode Spatio Temporal Density Based Spatial Clustering of Applications with Noise (ST-DBSCAN) yang dioptimalkan dengan Particle Swarm Optimazation (PSO). Data yang digunakan meliputi 7 Variabel dengan 68 Observasi. Berdasarkan evaluasi Silhouette Coefficient (0.104994), penggunaan algoritma PSO menghasilkan nilai parameter ST-DBSCAN antara lain ϵ_1 sebesar 2.73, ϵ_2 sebesar 2, dan MinPts sebesar 5. Parameter-parameter tersebut menghasilkan 4 cluster utama dan satu cluster noise, dimana *cluster 0* sebanyak 13 Provinsi, *cluster 1* sebanyak 3 Provinsi, *cluster 2* sebanyak 4 Provinsi, *cluster 3* sebanyak 3 Provinsi, dan *cluster noise* sebanyak 11 Provinsi.

Kata kunci: PSO, ST-DBSCAN, Silhouette Coefficient, IPLM.

Abstract

The index of Community Literacy Development (IPLM) is a measure of the efforts made by local governments (both at the provincial and district/city levels) in fostering and developing libraries as a means of lifelong learning to build a culture of literacy in the community. This study aims to cluster provinces in Indonesia based on elements that affect IPLM using the Spatio Temporal Density Based Spatial Clustering of Applications with Noise (ST-DBSCAN) method optimized with Particle Swarm Optimazation (PSO). The data used includes 7 variables with 68 observations. Based on the Silhouette Coefficient evaluation (0.104994), the use of the PSO algorithm produces ST-DBSCAN parameter values including ϵ_1 of 2.73, ϵ_2 of 2, and MinPts of 5. These parameters produce 4 main clusters and one noise cluster, where cluster 0 is 13 provinces, cluster 1 is 3 provinces, cluster 2 is 4 provinces, cluster 3 is 3 provinces, and noise cluster is 11 provinces.

Keywords: PSO, ST-DBSCAN, Silhouette Coefficient, IPLM.

2020 Mathematics Subject Classification: 62H30, 90C59, 62P25.

Diterima: 18-02-25, direvisi: 23-03-25, dimuat: 02-05-25.

1 Pendahuluan

Indeks Pembangunan Literasi Masyarakat (IPLM) merupakan ukuran terhadap upaya yang dilakukan oleh pemerintah daerah (baik di tingkat provinsi maupun kabupaten/kota) dalam membina dan mengembangkan perpustakaan sebagai sarana pembelajaran sepanjang hayat untuk membangun budaya literasi di masyarakat. IPLM diukur dengan tujuan untuk memahami kondisi berbagai jenis perpustakaan di seluruh Indonesia, mencakup aspek distribusi perpustakaan, koleksi, tenaga perpustakaan, hingga jumlah pemustaka [1]. Berdasarkan data Badan Pusat Statistik (BPS) Indonesia, skor Indeks Pembangunan Literasi Masyarakat (IPLM) Indonesia pada tahun 2023 adalah 68,19. Predikat IPLM dikelompokkan menjadi lima kategori: "sangat rendah" (0-29,9), "rendah" (30-49,9), "sedang" (50-79,9), "tinggi" (80-89,9), dan "sangat tinggi" (90-100). Meskipun skor IPLM Indonesia saat ini masih masuk dalam kategori "sedang", beberapa provinsi di Indonesia telah mencapai predikat IPLM "tinggi". Provinsi tersebut adalah Sulawesi Selatan dan Daerah Istimewa Yogyakarta, dengan skor masing-masing sebesar 86,74 dan 85,09. Sementara itu, skor IPLM provinsi lainnya yang termasuk dalam 10 besar masih berada dalam klasifikasi "sedang".

Indonesia masih menghadapi tantangan dalam memenuhi kebutuhan koleksi bahan bacaan yang memadai bagi jumlah penduduknya, serta akses yang belum merata pada setiap wilayah untuk memperolehnya. Persoalan utama dalam kecakapan literasi diungkap oleh Plt Kepala Perpustakaan Nasional (Perpusnas), E. Aminudin Aziz. Persoalan tersebut meliputi ketersediaan bahan bacaan yang belum sesuai dengan minat baca calon pembaca, program dan aktivitas yang belum terarah, dan sumber daya manusia yang kurang. Pusat Pengembangan Perpustakaan dan Pemasarakatan Minat Baca Perpusnas RI menyusun elemen-elemen yang mempengaruhi Indeks Pembangunan Literasi Masyarakat di setiap provinsi di Indonesia. Elemen-elemen ini mencakup tujuh indikator/aspek, yaitu pemerataan layanan perpustakaan, kecukupan koleksi, kecukupan tenaga perpustakaan, jumlah kunjungan masyarakat per hari, jumlah perpustakaan yang memenuhi SNP, partisipasi masyarakat dalam kegiatan sosialisasi, dan jumlah anggota perpustakaan. Kemiripan dari Karakteristik unsur Pembangunan Literasi Masyarakat pada setiap wilayah dimanfaatkan untuk mengelompokkan setiap wilayah di Indonesia. Tujuannya adalah agar kebijakan pihak-pihak terkait dapat disesuaikan dengan kondisi indikator pembangunan literasi masyarakat di setiap kelompok yang terbentuk [2].

Penelitian terdahulu mengenai Indeks Pembangunan Literasi Masyarakat oleh Warsito, dkk. (2023) dengan metode analisis deskriptif untuk analisis IPLM dan *Structural Equation Modelling* (SEM) untuk analisis Indeks Literasi masyarakat (ILM). Analisis ILM memberikan hasil bahwa dari estimasi hubungan signifikan pada tingkat signifikansi 5%, terlihat bahwa ketujuh variabel ILM memiliki koefisien jalur positif, seperti kemampuan membaca yang menunjukkan kausalitas antar variabel, sehingga terdapat hubungan positif di antara ketujuh variabel yang digunakan. Berdasarkan uji F statistik, menunjukkan adanya perbedaan yang signifikan terhadap nilai ILM berdasarkan aspek status pekerjaan dan tingkat kemudahan akses informasi. Hasil akhir diperoleh IPLM Kota Salatiga mencapai angka 72,83, yang masuk dalam kategori tingkat sedang (memenuhi standar) [3]. Shoimah (2023) menggunakan metode analisis deskriptif dan pengukuran IPLM dengan hasil Nilai Indeks Pembangunan Literasi Masyarakat (IPLM) Kabupaten Probolinggo pada tahun 2021 sebesar 10,67, sedikit lebih rendah dibandingkan dengan IPLM nasional yang mencapai 12,993 dan IPLM Provinsi Jawa Timur sebesar 13,44 [4].

Penelitian-penelitian tersebut tidak menunjukkan adanya pola pengelompokan karakteristik unsur atau indikator dalam indeks pembangunan literasi masyarakat (IPLM) di setiap wilayah. Analisis pengelompokan ini penting untuk memastikan bahwa kebijakan yang dirancang dapat disesuaikan dengan kondisi spesifik indikator IPLM pada masing-masing kelompok, sehingga diharapkan dapat berkontribusi pada peningkatan signifikan angka IPLM di Indonesia. Berdasarkan penelitian yang dilakukan [2], data IPLM mengandung banyak outlier atau *noise* yang dapat me-

nyebabkan hasil pengelompokkan tidak representatif. Salah satu metode pengelompokkan yang dapat mengatasi masalah tersebut adalah *Spatio Temporal Density Based Spatial Clustering of Application with Noise* (ST-DBSCAN). ST-DBSCAN adalah metode pengelompokkan yang mengidentifikasi cluster berdasarkan jarak kedekatan dan hubungan tetangga. Metode ini merupakan pengembangan dari *Density Based Spatial Clustering of Application with Noise* (DBSCAN) yang dirancang untuk mengatasi keterbatasan DBSCAN dalam pemrosesan data spasial dan temporal. ST-DBSCAN mampu mengelompokkan data berdasarkan dimensi spasial dan temporal serta membedakan titik *noise* pada data dengan kepadatan yang beragam, sehingga lebih adaptif terhadap berbagai jenis data dengan distribusi kepadatan yang berbeda.

Algoritma ST-DBSCAN memerlukan empat parameter, yaitu *Epsilon1* (ϵ_1), *Epsilon2* (ϵ_2), Minimum Points (MinPts), dan $\Delta\epsilon$ [5]. Pemilihan parameter dalam metode ST-DBSCAN sangat penting karena secara langsung mempengaruhi pembentukan *cluster*. Pengoptimalan parameter dapat dilakukan dengan menerapkan metode metaheuristic [6]. Penggunaan algoritma metaheuristic yang belum pernah digunakan untuk mengoptimasi nilai parameter ST-DBSCAN adalah algoritma Particle Swarm Optimization (PSO). Menurut Hartono (2022) penerapan algoritma PSO untuk mengoptimasi K-Means menghasilkan nilai parameter yang optimal untuk pengelompokkan tingkat kesejahteraan di Jawa Timur yaitu 10 partikel untuk nilai parameter tertinggi [7]. Oleh karena itu, berdasarkan uraian diatas penelitian ini akan menggunakan algoritma PSO untuk mengoptimalkan nilai parameter dari ST-DBSCAN pada pengelompokkan provinsi di Indonesia berdasarkan Indeks Pembangunan Literasi Masyarakat tahun 2022 dan 2023.

2 Tinjauan Pustaka

2.1 Analisis Cluster

Analisis *cluster* adalah teknik multivariat yang bertujuan untuk mengelompokkan objek atau kasus berdasarkan karakteristik yang dimilikinya. *Cluster* dapat diartikan sebagai 'kelompok,' dan secara umum, analisis ini akan menghasilkan beberapa *cluster* [8]. Metode ini berfungsi untuk menganalisis kemiripan antara sampel yang berbeda, dengan prinsip bahwa sekelompok objek dikelompokkan ke dalam beberapa *cluster*, di mana objek-objek dalam satu *cluster* memiliki tingkat kemiripan yang tinggi, namun berbeda atau tidak serupa dengan objek-objek di *cluster* lain [9].

2.2 Uji Multikolinearitas

Analisis *cluster* memerlukan asumsi bahwa tidak ada korelasi antara variabel-variabel, sehingga uji multikolinieritas perlu dilakukan untuk mengetahui apakah terdapat korelasi di antara variabel [10]. Salah satu cara uji asumsi multikolinieritas adalah dengan menggunakan *Variance Inflation Factor* (VIF). Nilai VIF itu sendiri dapat diperoleh dengan rumus [11]:

$$\text{VIF} = \frac{1}{1 - R^2} \quad (1)$$

dengan R^2 adalah koefisien determinasi antar variabel independen. Jika nilai $\text{VIF} > 10$ artinya terindikasi masalah multikolinearitas dan jika nilai $\text{VIF} \leq 10$ artinya tidak terindikasi masalah multikolinearitas [11].

2.3 Standarisasi Data

Standarisasi data bertujuan untuk mengurangi perbedaan dalam satuan yang signifikan antara variabel-variabel yang diteliti dan menyetarakan jangkauan setiap variabel sehingga memiliki skala yang sama. Teknik standarisasi yang digunakan dalam penelitian ini adalah *z-score*. *Z-score*, atau yang juga dikenal sebagai skor baku, adalah angka yang menunjukkan seberapa jauh suatu skor mentah berada dari skor rata-rata (μ) dalam satuan simpangan baku (σ) [12]. Proses standarisasi data melibatkan penggunaan persamaan berikut [13] :

$$Z = \frac{x_i - \bar{x}}{\sigma} \quad (2)$$

dengan Z adalah *z-score* dari suatu variabel, x Nilai data dari suatu variabel, \bar{x} adalah rata-rata data (*mean*) dari suatu variabel, dan σ adalah standar deviasi dari suatu variabel.

2.4 Jarak Euclidean

Clustering pada dasarnya adalah metode untuk mengelompokkan data yang memiliki karakteristik serupa antara satu observasi dengan yang lain. Untuk mengukur kesamaan karakteristik dalam suatu cluster, dapat digunakan ukuran jarak (*distance measure*). Salah satu ukuran jarak yang paling umum digunakan adalah jarak Euclidean, yang merupakan metode untuk menghitung jarak antara dua observasi dalam ruang Euclidean [14]. Untuk mengukur kesamaan data dengan rumus Euclidean Space, dapat dihitung rumus berikut [15]:

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^n (\mathbf{x}_i - \mathbf{y}_i)^2} \quad (3)$$

dengan $d(\mathbf{x}, \mathbf{y})$ adalah jarak antara \mathbf{x} dan \mathbf{y} , \mathbf{x} adalah data pusat *cluster*, \mathbf{y} adalah data atribut *cluster* i adalah data setiap observasi, n adalah jumlah data, \mathbf{x}_i adalah data pusat *cluster* ke- i , \mathbf{y}_j adalah data atribut ke- j .

2.5 Parameter Tuning

algoritma ST-DBSCAN membutuhkan beberapa parameter, antara lain epsilon1 (ϵ_1), epsilon2 (ϵ_2), dan MinPts [16]. ϵ_1 adalah jarak untuk atribut spasial, ϵ_2 adalah jarak untuk atribut temporal, sedangkan MinPts Menentukan jumlah minimum titik dalam kombinasi spasial-temporal yang dibutuhkan untuk membentuk sebuah kluster. ST-DBSCAN dirancang untuk menangani data spatio-temporal dengan memperhitungkan kedekatan spasial dan temporal dalam proses clustering. Oleh karena itu, pengaturan parameter yang tepat memungkinkan algoritma menyesuaikan diri dengan karakteristik unik data dan meningkatkan akurasi hasil pengelompokan [17]. Pada penelitian ini, metode K-Nearest Neighbor akan digunakan untuk membangkitkan nilai awal dari MinPts dan metode *Elbow* akan digunakan dalam menentukan nilai awal dari ϵ_1 .

2.5.1 K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* (KNN) merupakan salah satu metode klasifikasi yang mengelompokkan data berdasarkan jarak terdekatnya dengan beberapa data lain atau tetangga (*neighbor*). Jarak antara dua objek, x dan y , dapat dihitung menggunakan berbagai metode, salah satunya adalah Jarak Euclidean [18].

2.5.2 Elbow

Metode *elbow* adalah teknik visual menggunakan grafik untuk menentukan jumlah *cluster* optimal dalam suatu data. Metode ini membandingkan nilai *Sum of Square Error* (SSE) untuk setiap jumlah *cluster* yang berbeda. Dalam analisis *cluster*, metode *elbow* dijabarkan dengan rumus berikut [19]:

$$SSE = \sum_{k=1}^k \sum_{x_i \in S_k} \|X_i - C_k\|^2 \quad (4)$$

dengan k adalah jumlah *cluster*, x_i adalah nilai atribut dari data ke- i , C_k adalah jumlah *cluster* i pada *cluster* ke- k , dan $\|$ adalah menghitung jarak *euclidean*.

2.6 Particle Swarm Optimization

Particle Swarm Optimization (PSO), yang pertama kali dikembangkan oleh Kennedy dan Eberhart pada tahun 1995, adalah algoritma yang terinspirasi dari perilaku kelompok hewan, seperti kawanan burung atau ikan. PSO didasarkan pada dua pendekatan utama, yaitu keterkaitannya dengan artificial life secara umum, serta teori kawanan burung, gerombolan ikan, dan swarm secara khusus [20]. PSO bertujuan untuk mengoptimalkan fungsi objektif dari suatu masalah dengan mengeksplorasi ruang pencarian secara kolektif sambil mempertahankan sejumlah solusi kandidat atau partikel [21]. Berikut ini merupakan persamaan dari algoritma PSO ([20]):

$$X_i(t) = x_{i1}(t), x_{i2}(t), x_{i3}(t), \dots, x_{iN}(t)$$

$$V_i(t) = v_{i1}(t), v_{i2}(t), v_{i3}(t), \dots, v_{iN}(t)$$

Keterangan:

- X : Posisi partikel.
- V : Kecepatan partikel.
- i : Indeks partikel pada dimensi ruang N .
- t : Iterasi ke- t pada dimensi ruang N .
- N : Dimensi ruang.

Berikut ini adalah model yang menggambarkan tahapan perbaikan status partikel ([20]):

$$V_i(t) = \omega V_i(t-1) + c_1 r_1 (X_i^L - X_i(t-1)) + c_2 r_2 (X^G - X_i(t-1)) \quad (5)$$

$$X_i(t) = V_i(t) + X_i(t-1) \quad (6)$$

Keterangan:

- $X_i^L = X_{i1}^L, X_{i2}^L, \dots, X_{iN}^L$
- $X^G = X_1^G, X_2^G, \dots, X_N^G$
- X_i^L : Local best dari partikel ke- i
- X^G : Global best dari seluruh kelompok partikel
- c : Konstanta bernilai positif yang disebut learning factor
- r : Bilangan random positif dengan rentang 0 sampai 1
- ω : Parameter inersia.

Persamaan (5) digunakan untuk mendapatkan posisi partikel terbaik (*local best*), dan jarak saat ini adalah posisi kawanan terbaik (*global best*). Setelah itu, partikel bergerak menuju posisi yang baru sesuai dengan persamaan (6).

2.7 Evaluasi Kluster *Silhouette Coefficient*

Teknik evaluasi kluster yang digunakan pada penelitian ini adalah metode *Silhouette Coefficient* (SC), digunakan untuk menilai kualitas dan kekuatan hasil pengelompokan. Setelah memperoleh nilai SC untuk data spasial, dilanjutkan dengan menghitung nilai SC untuk data temporal. Nilai SC global yang diperoleh akan menjadi acuan dalam memilih kombinasi parameter yang optimal. Semakin mendekati 1, semakin baik kualitas pengelompokan yang dihasilkan dari kombinasi parameter tersebut [22]. Persamaan dalam mengevaluasi kluster menggunakan Silhouette Coefficient diuraikan sebagai berikut [23]:

$$S_i = \frac{(b_i - a_i)}{\max(b_i - a_i)} \quad (7)$$

dengan a_i adalah rata-rata jarak objek i dengan seluruh objek dan b_i adalah nilai terkecil rata-rata objek i dengan objek pada *cluster* yang berbeda.

2.8 ST-DBSCAN (*Spatio Temporal Density Based Spatial Clustering of Application with Noise*)

Spatio-Temporal Density Based Spatial Clustering of Applications with Noise (ST-DBSCAN) adalah algoritma yang dihasilkan dari modifikasi algoritma DBSCAN. ST-DBSCAN memungkinkan pengelompokan data spasial dan temporal berdasarkan atribut non-spasial, spasial, dan temporal. Algoritma DBSCAN tidak dapat mendeteksi titik *noise* saat menangani data dengan kepadatan yang berbeda, namun ST-DBSCAN mampu mengatasi masalah ini dengan menyesuaikan faktor kepadatan untuk setiap kluster. Algoritma ST-DBSCAN memerlukan empat parameter, yaitu *epsilon1* (ϵ_1), *epsilon2* (ϵ_2), MinPts, dan $\Delta\epsilon$. *epsilon1* adalah jarak untuk atribut spasial, sedangkan *epsilon2* adalah jarak untuk atribut temporal. Sedangkan $\Delta\epsilon$ digunakan untuk mencegah penggabungan kluster yang tidak sesuai karena perbedaan signifikan dalam nilai atribut non-spasial dari objek lain [17]. Tahapan dari algoritma ST-DBSCAN [17] yaitu:

1. Menentukan nilai ϵ dan MinPts.
2. Menghitung semua matriks jarak antar objek berdasarkan aspek spasial dan temporal.
3. Membentuk matriks jarak untuk semua pasangan dari n objek berdasarkan spasial dan temporal.
4. Dimulai dari titik awal, kemudian mengambil semua titik berdasarkan aspek spasial dan temporal dengan ketentuan berikut:

$$A = \{x \mid x \leq \epsilon_1, x \in \text{matriks jarak spasial}\}$$

$$B = \{x \mid x \leq \epsilon_2, x \in \text{matriks jarak temporal}\}$$

5. Mengambil semua irisan dari aspek spasial dan aspek temporal dengan ketentuan berikut:

$$A \cap B = \{x \mid x \in A \text{ dan } x \in B\}$$

6. Jika jumlah objek pada irisan lebih kecil dari nilai MinPts, maka titik dinilai sebagai *noise*.
7. Kluster terbentuk jika titik memenuhi nilai parameter ϵ_1 , ϵ_2 , dan MinPts.

2.9 Indeks Pembangunan Literasi Masyarakat

Indeks Pembangunan Literasi Masyarakat merupakan data yang menggambarkan tingkat perkembangan literasi masyarakat, yang diperoleh dari unsur-unsur pengembangan literasi masyarakat (UPLM) berdasarkan data sekunder serta aspek masyarakat (AM). Tujuannya adalah untuk mendukung pembinaan dan pengembangan perpustakaan sebagai sarana pembelajaran seumur hidup demi meningkatkan literasi masyarakat. [24]. Pusat Pengembangan Perpustakaan dan Pemasarakan Minat Baca Perpustakaan RI telah menyusun UPLM yang terdiri atas tujuh komponen, yaitu Pemerataan layanan perpustakaan (UPLM1), ketercukupan koleksi (UPLM2), kecukupan tenaga perpustakaan (UPLM3), tingkat kunjungan masyarakat per hari (UPLM4), jumlah perpustakaan yang dibina sesuai dengan Standar Nasional Perpustakaan (UPLM5), keterlibatan masyarakat dalam kegiatan sosialisasi (UPLM6), dan jumlah anggota perpustakaan (UPLM7) [24].

3 Hasil dan Pembahasan

3.1 Sumber data

Penelitian ini menggunakan data sekunder yang diperoleh dari Indeks Pembangunan Literasi Masyarakat Tahun 2022 dan 2023 dalam publikasi IPLM 2022 dan IPLM 2023 pada laman <https://ckan.perpusnas.go.id>. Unit Penelitian ini adalah Provinsi. Populasi dan sampel yang digunakan dalam penelitian ini adalah Indeks Pembangunan Literasi Masyarakat 34 Provinsi di Indonesia Tahun 2023 dan 2024 sebanyak 68 baris data. Serta Variabel yang digunakan adalah tujuh variabel independen, yaitu Pemerataan Layanan Perpustakaan (X_1), Ketercukupan Koleksi Perpustakaan (X_2), Ketercukupan Tenaga Perpustakaan (X_3), Tingkat Kunjungan Masyarakat per hari (X_4), Jumlah Perpustakaan yang dibina sesuai SNP (X_5), Keterlibatan Masyarakat dalam sosialisasi perpustakaan (X_6), dan Perkembangan jumlah anggota perpustakaan (X_7).

3.2 Uji Multikolinearitas

Berikut adalah hasil uji multikolinearitas yang disajikan pada tabel berikut.

Tabel 1: Hasil Uji Multikolinearitas

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
X_1	-	2,42975	2,37992	1,14915	1,47824	5,17663	2,77699
X_2	2,42975	-	1,95375	1,30484	1,68061	4,23329	1,78730
X_3	2,37992	1,95375	-	1,22629	1,45777	3,16274	2,18575
X_4	1,14915	1,30484	1,22629	-	1,17025	1,29019	1,15754
X_5	1,47824	1,68061	1,45777	1,17025	-	1,62672	1,27602
X_6	5,17663	4,23329	3,16274	1,29019	1,62672	-	4,70498
X_7	2,77699	1,78730	2,18575	1,15754	1,27602	4,70498	-

Berdasarkan hasil uji multikolinearitas di atas, nilai VIF antar variabel adalah < 10 yang artinya tidak terindikasi adanya multikolinearitas sehingga variabel-variabel tersebut dapat digunakan untuk proses analisis kluster pada penelitian ini.

3.3 Jarak Euclidean

Clustering pada dasarnya adalah metode untuk mengelompokkan data yang memiliki karakteristik serupa antara satu observasi dengan yang lain. Untuk mengukur kesamaan karakteristik dalam suatu cluster, dapat digunakan ukuran jarak (*distance measure*). Salah satu ukuran jarak

yang paling umum digunakan adalah jarak Euclidean, yang merupakan metode untuk menghitung jarak antara dua observasi dalam ruang Euclidean [14]. Berikut adalah hasil perhitungan yang disajikan dalam tabel berikut:

Tabel 2: Matriks Jarak Euclidean

N	1	2	3	...	66	67	68
1	0	4.493	1.190	...	4.635	1.089	3.895
2	4.493	0	4.924	...	0.555	4.576	1.765
3	1.190	4.924	0	...	5.056	1.763	4.392
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
66	4.635	0.555	5.056	...	0	4.717	1.797
67	1.089	4.576	1.763	...	4.717	0	3.982
68	3.895	1.765	4.392	...	1.797	3.982	0

Nilai ini digunakan untuk mengidentifikasi titik inti berdasarkan jumlah tetangga yang melebihi ambang batas *MinPts*. Melalui proses penghubungan, titik-titik inti yang berdekatan dalam radius *epsilon* akan saling terhubung membentuk sebuah *cluster*. Sementara itu, titik-titik yang tidak memenuhi kriteria sebagai titik inti dan tidak dapat terhubung ke *cluster* lain dianggap sebagai noise berdasarkan jarak euclidean.

3.4 Parameter Tuning

Terdapat tiga parameter ST-DBSCAN yang digunakan dalam penelitian ini adalah ϵ_1 , ϵ_2 , dan *Minimum Points* (MinPts). ϵ_1 ditentukan dengan menggunakan metode Elbow, ϵ_2 ditentukan dengan menyesuaikan rentan waktu yaitu 2 karena pada penelitian ini menggunakan rentan waktu 2 tahun, serta MinPts ditentukan dengan menggunakan metode *Nearest Neighbor*.

3.4.1 K-Nearest Neighbor

Hasil pembangkitan nilai awal menggunakan *Nearest Neighbors* disajikan sebagai berikut.

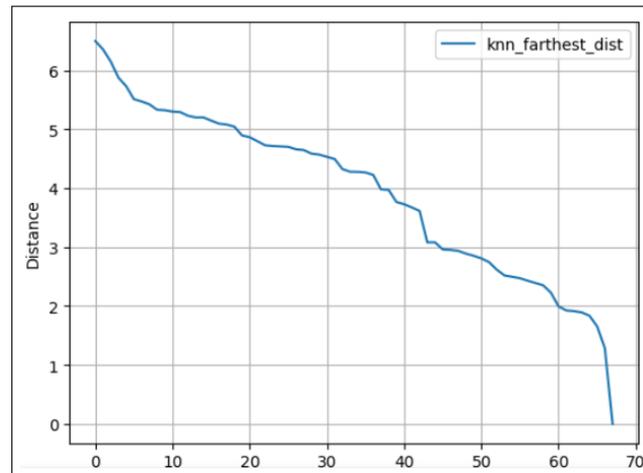
Tabel 3: Hasil *Nearest Neighbors*

Minimum Points	Shape of Distance Matrix
16	26.088

Nilai tersebut adalah pembangkit awal dari MinPts yang akan digunakan sebagai batasan eksplorasi pada algoritma PSO.

3.4.2 Elbow

Pencarian nilai pembangkit awal untuk ϵ_1 dilakukan menggunakan metode elbow dengan mengidentifikasi titik di mana penurunan inertia mulai melambat secara signifikan. berikut adalah visualisasi dari metode *elbow*.



Gambar 1: Grafik *Elbow*

Gambar 1 menunjukkan bahwa *elbow point* terbentuk pada titik 3.1, yang mengindikasikan adanya penurunan yang melambat pada inertia di titik tersebut. Berdasarkan hasil ini, nilai awal yang digunakan oleh peneliti ditetapkan sebesar 3.1.

3.5 Particle Swarm Optimization

Pencarian parameter optimal dilakukan untuk menentukan nilai *epsilon* 1 dan MinPts yang akan digunakan dalam proses pengelompokan data menggunakan metode ST-DBSCAN. Nilai hasil penyetalan parameter ini digunakan sebagai batas dalam pencarian parameter *epsilon* dan minimum point pada algoritma PSO. Algoritma tersebut menjalankan iterasi sesuai dengan batasan melalui fungsi objektif yang bertujuan menemukan *Gbest* sebagai *epsilon* 1 dan *Pbest* sebagai MinPts. Penelitian ini mencatat bahwa algoritma PSO melakukan iterasi maksimal ketika menemukan solusi yang optimal, dengan informasi terkait iterasi telah disajikan dalam tabel berikut.

Tabel 4: Informasi Iterasi

Iterasi Ke-	ϵ_1	MinPts	Fitness Values
1	3,08	16	1,000
2	3,08	14	0,236
3	3,10	12	0,249
⋮	⋮	⋮	⋮
67	2,73	5	0,199

PSO secara intrinsik mengoptimalkan untuk meminimalkan nilai fungsi tujuan, maka nilai ϵ_1 dan MinPts yang optimal dipilih berdasarkan nilai fungsi tujuan terendah. Dari proses iterasi yang disajikan, diperoleh nilai ϵ_1 dan MinPts dengan nilai fungsi tujuan yang optimal berhenti pada iterasi ke-67. Oleh karena itu, berdasarkan **tabel 4** parameter optimal yang dihasilkan adalah nilai ϵ_1 sebesar 2.73 dan MinPts sebesar 5.

3.6 Evaluasi Silhouette Coefficient

Hasil evaluasi nilai parameter menggunakan *Silhouette Coefficient* (SC) telah disajikan dalam tabel berikut.

Tabel 5: Hasil evaluasi *Silhouette Coefficient*

ϵ_1	ϵ_2	MinPts	SC	Jumlah Klaster
2.73	2	5	0.104994	5

Tabel 5 menunjukkan nilai SC dari parameter yang optimal adalah sebesar 0.104994 dengan jumlah klaster yang optimal adalah 5 klaster. Selanjutnya akan dilakukan pembentukan klaster menggunakan algoritma ST-DBSCAN.

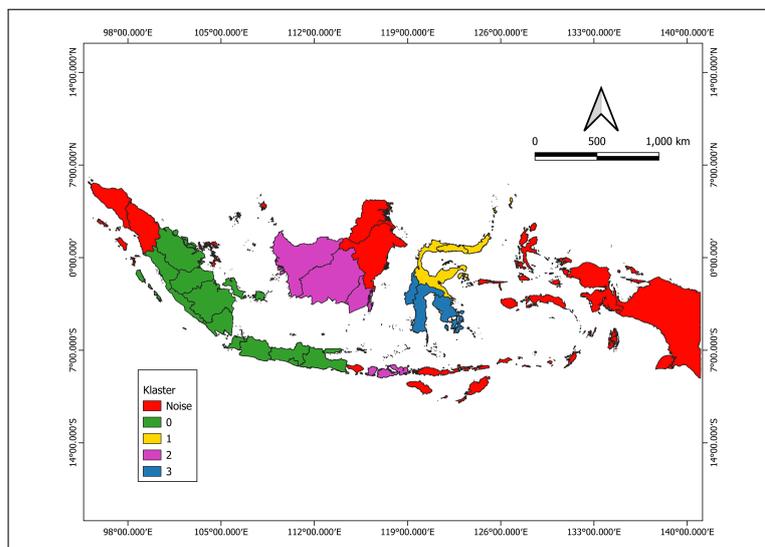
3.7 Pembentukan Cluster dengan ST-DBSCAN

Pembentukan *cluster* menggunakan metode ST-DBSCAN sangat bergantung pada penerapan nilai parameter yang digunakan, yaitu ϵ_1 , ϵ_2 , dan MinPts. Parameter yang akan digunakan adalah nilai parameter optimal yang diperoleh dari tahapan sebelumnya, yaitu ϵ_1 sebesar 2.73, ϵ_2 sebesar 2, dan MinPts sebesar 5. Hasil klasterisasi provinsi di Indonesia berdasarkan faktor yang mempengaruhi IPLM dengan menggunakan parameter tersebut telah disajikan dalam tabel berikut.

Tabel 6: Hasil *Cluster* ST-DBSCAN

<i>Cluster</i>	Provinsi
0	Banten, Bengkulu, DI Yogyakarta, DKI Jakarta, Jambi, Jawa Barat, Jawa Tengah, Jawa Timur, Kepulauan Bangka Belitung, Lampung, Riau, Sumatera Barat, Sumatera Selatan
1	Gorontalo, Sulawesi Tengah, Sulawesi Utara
2	Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, NTB
3	Sulawesi Barat, Sulawesi Selatan, Sulawesi Tenggara
Noise	Aceh, Bali, Kalimantan Timur, Kalimantan Utara, Kepulauan Riau, Maluku, Maluku Utara, NTT, Papua, Papua Barat, Sumatera Utara

Tabel 6 menunjukkan klaster yang terbentuk adalah sebanyak 5 klaster yang terdiri atas klaster 0 sebanyak 13 Provinsi, klaster 1 sebanyak 3 Provinsi, klaster 2 sebanyak 4 Provinsi, klaster 3 sebanyak 3 Provinsi, serta klaster yang terdeteksi sebagai *noise* sebanyak 11 Provinsi. Berikut adalah pemetaan hasil analisis *cluster*.



Gambar 2: Pemetaan Klaster Menggunakan QGIS

3.8 Interpretasi Pembentukan *cluster* ST-DBSCAN

Karakteristik setiap *cluster* pada klasterisasi 34 Provinsi di Indonesia menggunakan ST-DBSCAN berdasarkan faktor-faktor yang memengaruhi nilai IPLM dapat dilihat berdasarkan nilai rata-rata pada variabel untuk setiap *cluster* yang disajikan dalam tabel berikut.

Tabel 7: Nilai Rata-rata Variabel pada Setiap Klaster

Variabel	Klaster 0	Klaster 1	Klaster 2	Klaster 3
X_1	0.25964	0.19937	0.22706	0.22348
X_2	0.98656	1.13937	0.94921	1.07690
X_3	0.29561	0.30618	0.43075	0.40543
X_4	0.34851	0.27905	0.29608	0.41395
X_5	1.52783	1.15630	1.27053	1.17712
X_6	0.65916	0.65228	0.67056	0.64057
X_7	0.80292	0.81127	0.82899	0.80842

Berdasarkan **Tabel 7**, nilai rata-rata dari variabel X_1 (pemerataan layanan perpustakaan) dan X_5 (jumlah perpustakaan yang dibina sesuai Standar Nasional Perpustakaan) tertinggi berada di *cluster* 0 dibandingkan dengan *cluster* 1, *cluster* 2, dan *cluster* 3. Nilai rata-rata yang tinggi bermakna bahwa faktor yang paling berpengaruh terhadap nilai IPLM pada wilayah-wilayah yang berada di *cluster* 0 adalah pemerataan layanan perpustakaan dan jumlah perpustakaan yang dibina sesuai standar nasional perpustakaan.

Nilai rata-rata dari variabel X_2 (ketercukupan koleksi perpustakaan) tertinggi berada di *cluster* 1 dibandingkan dengan *cluster* 0, *cluster* 2 dan *cluster* 3. Hal ini bermakna bahwa faktor yang paling berpengaruh terhadap nilai IPLM pada wilayah-wilayah yang berada di *cluster* 1 adalah ketercukupan koleksi perpustakaan.

Nilai rata-rata dari variabel X_3 (ketercukupan tenaga perpustakaan), X_6 (keterlibatan masyarakat dalam sosialisasi perpustakaan) dan X_7 (perkembangan jumlah anggota perpustakaan) yang tertinggi berada di *cluster* 2 dibandingkan dengan *cluster* 0, *cluster* 1 dan *cluster* 3. Hal ini bermakna bahwa faktor yang paling berpengaruh terhadap nilai IPLM pada wilayah-wilayah yang berada di *cluster* 2 adalah ketercukupan tenaga perpustakaan, keterlibatan masyarakat dalam sosialisasi perpustakaan, dan perkembangan jumlah anggota perpustakaan.

Nilai rata-rata dari variabel X_4 (tingkat kunjungan masyarakat per hari) yang tertinggi berada di *cluster* 3 dibandingkan *cluster* 0, *cluster* 1 dan *cluster* 2. Hal ini bermakna bahwa faktor yang paling berpengaruh terhadap nilai IPLM pada wilayah-wilayah yang berada di *cluster* 3 adalah tingkat Kunjungan masyarakat per hari.

Cluster noise dalam metode ST-DBSCAN merujuk pada titik-titik data yang tidak termasuk dalam *cluster* manapun karena tidak memiliki cukup tetangga dalam radius spasial (ϵ_1) dan temporal (ϵ_2) yang memenuhi ambang batas minimum titik (MinPts). Titik-titik ini dianggap sebagai outlier atau anomali karena berada di area dengan kepadatan rendah secara spasial-temporal.

4 Kesimpulan

Berdasarkan hasil penggunaan algoritma PSO dalam menentukan parameter ST-DBSCAN yang optimal, diperoleh nilai parameter $\epsilon_1 = 2,73$, $\epsilon_2 = 2$ dan MinPts = 5, dengan *cluster* yang dihasilkan berjumlah 4 *cluster* serta terdapat *cluster* noise. Parameter tersebut dianggap optimal dibuktikan dengan hasil evaluasi berdasarkan nilai *Silhouette Coefficient* (SC) sebesar 0,104994.

Hasil *cluster* yang diperoleh berdasarkan nilai parameter yang optimal pada klasterisasi Provinsi di Indonesia berdasarkan faktor yang mempengaruhi IPLM menggunakan metode ST-DBSCAN

adalah sebanyak 5 *cluster* yaitu , *cluster* 0, *cluster* 1, *cluster* 2, *cluster* 3, dan *cluster* noise. *cluster* 0 terdiri dari 13 Provinsi, *cluster* 1 terdiri dari 3 Provinsi, *cluster* 2 terdiri dari 4 Provinsi, *cluster* 3 terdiri dari 3 Provinsi, serta *cluster* noise terdiri dari 11 Provinsi.

Daftar Pustaka

- [1] B. Warsito, A. R. Hakim, and E. Fatmawati, "Pengukuran indeks pembangunan literasi masyarakat (iplm) kota salatiga sebagai dasar penyusunan program pengembangan dan pembinaan perpustakaan," *Jurnal Pustaka Ilmiah*, vol. 9, no. 2, pp. 75–84, 2023.
- [2] W. Anbiya, F. C. Garini, and T. Purwandari, "Optimalisasi pembangunan perpustakaan di indonesia berdasarkan indikator indeks pembangunan literasi masyarakat: Aplikasi metode k-medoids clustering," *BIAStatistics Journal of Statistics Theory and Application*, vol. 2022, no. 1, pp. stat5–stat5, 2022.
- [3] B. Warsito, H. Muharam, A. R. Hakim, E. Fatmawati, Y. Y. Prasetyawan, *et al.*, "Dinamika literasi masyarakat salatiga: Perbandingan indeks pembangunan literasi dan indeks literasi 2022," *Media Pustakawan*, vol. 30, no. 1, pp. 93–105, 2023.
- [4] F. Shoimah, "Mengukur indeks pembangunan literasi masyarakat (iplm) kabupaten probolinggo," *Jurnal Pangripta*, vol. 6, no. 1, pp. 1164–1175, 2023.
- [5] M. I. P. Eriansya and M. Syafrullah, "Implementasi algoritma st-dbscan dan k-means untuk pengelompokan indeks pembangunan manusia kabupaten/kota pulau jawa tahun 2014-2016 berbasis web di badan pusat statistik," *SKANIKA: Sistem Komputer Dan Teknik Informatika*, vol. 1, no. 3, pp. 1026–1032, 2018.
- [6] R. Nooraeni and A. N. Fahira, "Optimasi parameter st-dbscan dengan knn dan algoritma genetika studi kasus: Data bencana alam di pulau jawa 2021," *Jurnal Komputasi*, vol. 11, no. 1, pp. 24–33, 2023.
- [7] A. D. Hartono, D. Arifianto, and I. Saifudin, "Implementasi algoritma k-means dengan particle swarm optimization untuk pengelompokan tingkat kesejahteraan di provinsi jawa timur," *Jurnal Aplikasi Sistem Informasi dan Elektronika*, vol. 4, no. 2, pp. 53–61, 2022.
- [8] S. Santoso, "Statistik multivariat edisi revisi," *Jakarta: PT Elex Media Komputindo*, vol. 224, 2014.
- [9] B. Yan and G. Deng, "Improved clustering algorithm based on density-isoline," *Open Journal of Statistics*, vol. 5, no. 4, pp. 303–310, 2015.
- [10] A. R. Pertiwi, F. Madina, F. Yulistina, A. R. N. Mawaddah, and A. Oktaviarina, "Analisis clustering metode hierarki single linkage terhadap indeks pembangunan manusia (ipm) indonesia tahun 2023," *Jurnal Sains Indonesia*, vol. 5, no. 1, pp. 18–24, 2024.
- [11] Y. Haitovsky, "Multicollinearity in regression analysis: Comment," *The Review of economics and statistics*, pp. 486–489, 1969.
- [12] A. R. Aziz, B. Warsito, and A. Prahutama, "Pengaruh transformasi data pada metode learning vector quantization terhadap akurasi klasifikasi diagnosis penyakit jantung," *Jurnal Gaussian*, vol. 10, no. 1, pp. 21–30, 2021.

- [13] G. W. Milligan and M. C. Cooper, "A study of standardization of variables in cluster analysis," *Journal of classification*, vol. 5, pp. 181–204, 1988.
- [14] M. Faisal, E. Zamzami, *et al.*, "Comparative analysis of inter-centroid k-means performance using euclidean distance, canberra distance and manhattan distance," in *Journal of Physics: Conference Series*, vol. 1566, p. 012112, IOP Publishing, 2020.
- [15] P.-E. Danielsson, "Euclidean distance mapping," *Computer Graphics and image processing*, vol. 14, no. 3, pp. 227–248, 1980.
- [16] A. Johar, A. Vatesia, and I. A. Donny, "Implementasi metode spatio temporal clustering dengan algoritma st-dbscan pada titik api kebakaran hutan indonesia (2015-2020)," *Rekursif: Jurnal Informatika*, vol. 11, no. 1, pp. 1–9, 2023.
- [17] D. Birant and A. Kut, "St-dbscan: An algorithm for clustering spatial–temporal data," *Data & knowledge engineering*, vol. 60, no. 1, pp. 208–221, 2007.
- [18] J. Saputra, Y. Sa'adati, V. Y. P. Ardhana, and M. Afriansyah, "Klasifikasi kematangan buah alpukat mentega menggunakan metode k-nearest neighbor berdasarkan warna kulit buah," *Resolusi: Rekayasa Teknik Informatika dan Informasi*, vol. 3, no. 5, pp. 196–203, 2023.
- [19] V. B. Samosir, A. M. Widodo, N. Anwar, B. A. Sekti, and N. Erzed, "Identifikasi outlier menggunakan teknik data mining clustering untuk analisis data tracer study pada fakultas ilmu komputer universitas esa unggul," *IKRA-ITH Informatika: Jurnal Komputer dan Informatika*, vol. 8, no. 1, pp. 162–174, 2024.
- [20] R. Eberhart and J. Kennedy, "Particle swarm optimization," in *Proceedings of the IEEE international conference on neural networks*, vol. 4, pp. 1942–1948, Citeseer, 1995.
- [21] H. Verma, D. Verma, and P. K. Tiwari, "A population based hybrid fcm-pso algorithm for clustering analysis and segmentation of brain image," *Expert systems with applications*, vol. 167, p. 114121, 2021.
- [22] B. A. Suryani, M. N. Hayati, and S. Prangga, "Pengelompokan kabupaten/kota di pulau Kalimantan pada tahun 2020 dan 2021 berdasarkan indeks pembangunan manusia menggunakan metode algoritma st-dbscan," *JURNAL Riset PEMBANGUNAN*, vol. 6, no. 1, pp. 1–9, 2023.
- [23] J. Kogan, *Introduction to clustering large and high-dimensional data*. Cambridge University Press, 2007.
- [24] E. Fatmawati, "Perencanaan dasar dalam mengukur kajian indeks pembangunan literasi masyarakat (iplm) dan indeks literasi masyarakat (ilm)," *IQRA: Jurnal Perpustakaan dan Informasi*, vol. 17, no. 1, pp. 172–205, 2023.