

## Enhancing the Efficacy of the K-Means Algorithm through Particle Swarm Optimization for Forest Fire Clustering

Putri Yuli Utami\*<sup>1</sup>, Sucipto<sup>2</sup>, Murni Ramadhani<sup>3</sup>, Mega Otafyani<sup>4</sup>

<sup>1,3,4</sup>Information System, Faculty of Engineering and Computer Science, Universitas Muhammadiyah Pontianak, Indonesia

<sup>2</sup>Informatics, Faculty of Engineering and Computer Science, Universitas Muhammadiyah Pontianak, Indonesia

Email: [putriyuli@unmuhpnk.ac.id](mailto:putriyuli@unmuhpnk.ac.id)

Received : Jun 24, 2025; Revised : Aug 8, 2025; Accepted : Sep 29, 2025; Published : Apr 18, 2026

### Abstract

Forest and land fires are a recurring ecological disaster in Indonesia, particularly in West Kalimantan, where peatlands and tropical climates contribute to high vulnerability. Effective identification of fire-prone areas is critical for mitigation efforts, yet conventional clustering methods such as K-Means suffer from limitations, especially in determining optimal cluster numbers and centroid initialization. This study proposes an enhanced clustering approach by integrating the Particle Swarm Optimization (PSO) algorithm with K-Means to improve the accuracy of hotspot clustering in Mempawah Regency. The research utilizes hotspot and weather datasets from January 2023 to March 2024, incorporating variables such as temperature, humidity, rainfall, and wind speed. Data preprocessing includes normalization using Min-Max Scaling. PSO is applied to determine the optimal number of clusters (K) within a range of 2 to 10 by evaluating Davies-Bouldin Index (DBI), Silhouette Coefficient (SC), and inertia. Experimental results show that the optimal configuration—20 iterations and 20 particles—yields a DBI of 0.897 and an SC of 0.464, indicating standard cluster quality. Visual validation using PCA demonstrates clear cluster separation, supporting the evaluation results. Compared to visual methods like Elbow, which suggest K=3-4, the PSO-KMeans approach identifies K=10 as optimal, providing better clustering performance. This research highlights the effectiveness of optimal K values in the K-Means algorithm using group intelligence in improving spacial data modeling and supporting strategic decision making for local forest fire mitigation efforts.

**Keywords :** Clustering, Forest Fire, K-Means, Particle Swarm Optimization, Spatial Analysis, Wildfire Risk

This work is an open access article and licensed under a Creative Commons Attribution-Non Commercial 4.0 International License



## 1. PENDAHULUAN

Kebakaran hutan dan lahan (karhutla) merupakan bencana ekologis yang hampir setiap tahun melanda berbagai wilayah di Indonesia, termasuk Provinsi Kalimantan Barat yang memiliki tingkat kerawanan tinggi akibat dominasi lahan gambut dan iklim tropis yang rentan terhadap kekeringan saat musim kemarau [1]-[2]. Dampak karhutla sangat signifikan terhadap lingkungan dan masyarakat, mencakup hilangnya keanekaragaman hayati, kerusakan ekosistem, peningkatan emisi gas rumah kaca, serta gangguan kesehatan akibat paparan asap [3]-[4]. Secara umum, kebakaran hutan dan lahan di Indonesia sebagian besar disebabkan oleh aktivitas manusia, baik secara sengaja maupun tidak, yang mencakup sekitar 99,9% dari total kejadian, sedangkan hanya 0,01% disebabkan oleh faktor alam [5]. Berdasarkan data Badan Penanggulangan Bencana Daerah (BPBD) Kalimantan Barat, tercatat lebih dari 13.057 hektar lahan terbakar sepanjang Januari hingga Agustus 2024, dengan lonjakan kasus di beberapa kabupaten rawan, termasuk Kabupaten Mempawah yang mencatatkan 271 titik panas, terutama pada Februari (122 titik) dan April (115 titik). Sementara itu, data dari Kementerian Lingkungan Hidup dan Kehutanan (KLHK) melalui sistem pemantauan Sipongi menunjukkan bahwa selama Januari hingga

Mei 2023, luas karhutla di Kabupaten Mempawah mencapai 400,95 hektar, menjadikannya wilayah dengan kebakaran terluas ketiga di Kalimantan Barat setelah Kubu Raya dan Ketapang [6]. Rasio antara jumlah titik panas dan luas kebakaran yang mencapai rata-rata 1,48 hektar per titik menunjukkan bahwa sebaran api sangat dipengaruhi oleh distribusi hotspot, sehingga pemanfaatan data spasial menjadi sangat penting dalam upaya mitigasi dan pengendalian risiko karhutla secara lebih efektif.

Namun demikian, upaya identifikasi dan pencegahan karhutla masih menghadapi berbagai tantangan, salah satunya adalah belum tersedianya pemetaan area rawan karhutla yang akurat dan terperinci di tingkat lokal [7]. Pengelompokan berdasarkan data hotspot menjadi pendekatan potensial untuk menentukan wilayah prioritas yang memerlukan penanganan lebih lanjut [8]. Salah satu algoritma yang sering digunakan untuk proses pengelompokan tersebut adalah K-Means karena kemampuannya dalam membagi data ke dalam beberapa kelompok berdasarkan kemiripan karakteristik [9]-[11]. Meskipun demikian, algoritma ini memiliki sejumlah keterbatasan, seperti keharusan menentukan jumlah kluster sejak awal, sensitivitas terhadap inisialisasi titik pusat (centroid) awal, serta penurunan performa ketika digunakan pada dataset berukuran besar [12]-[13]. Kelemahan-kelemahan tersebut dapat menyebabkan hasil klusterisasi menjadi kurang stabil dan kurang akurat, sehingga diperlukan pendekatan optimasi untuk meningkatkan kualitas pengelompokannya [14]-[15].

Penelitian sebelumnya di Provinsi Riau telah menggunakan algoritma Chebysev Distance K-Means untuk mengelompokkan data titik panas tahun 2016 menjadi tiga kategori tingkat kerawanan. Hasil klusterisasi menghasilkan nilai Davies Bouldin Index (DBI) sebesar 0,361, yang menunjukkan bahwa hasil pengelompokan cukup optimal. Hasil klusterisasi tersebut kemudian divisualisasikan secara spasial menggunakan Google Maps API [16]. Selain itu, penelitian lain mengembangkan model prediksi emisi karbon akibat karhutla di wilayah Sumatera dengan menggunakan algoritma Functional Link Neural Network (FLNN) yang dioptimasi menggunakan Particle Swarm Optimization (PSO). Dataset yang digunakan mencakup data emisi dari Global Fire Emissions Database (GFED) serta data iklim dari ERA5 selama periode 1998–2022. Hasil analisis menunjukkan bahwa faktor cuaca merupakan variabel paling dominan dalam memengaruhi emisi, dan model FLNN-PSO memberikan nilai Mean Absolute Error (MAE) terendah sebesar 254,28 gC/m<sup>2</sup>/bulan.

Salah satu metode optimasi yang digunakan dalam penelitian ini adalah Particle Swarm Optimization (PSO), algoritma metaheuristik yang meniru perilaku sosial kawanan dalam pencarian solusi optimal [17]-[20]. PSO diterapkan untuk mengoptimalkan inisialisasi pusat kluster pada algoritma K-Means, sehingga menghasilkan kelompok yang lebih stabil dan merepresentasikan pola sebaran titik panas karhutla secara lebih akurat [21]-[22]. Evaluasi hasil klusterisasi dilakukan dengan dua metrik evaluasi kuantitatif, yaitu Davies–Bouldin Index (DBI) dan Silhouette Coefficient (SC), yang masing-masing mengukur tingkat pemisahan antar kluster dan kekompakan data dalam satu kluster [23]. Untuk mendukung proses klusterisasi, dilakukan tahap praproses data seperti pembersihan, standarisasi format, dan sinkronisasi waktu antara data cuaca dan data titik panas. Selanjutnya, dilakukan analisis untuk mengkaji hubungan antara parameter cuaca seperti suhu, kelembaban, curah hujan, dan kecepatan angin dengan kemunculan titik panas di wilayah Kabupaten Mempawah. Hasil analisis ini menjadi dasar dalam pemodelan karakteristik daerah rawan kebakaran dan sebagai informasi pendukung dalam upaya mitigasi karhutla di tingkat lokal.

Penelitian ini bertujuan untuk mengembangkan model pemetaan areal rawan karhutla yang unggul secara teknis serta relevan untuk mendukung pengambilan keputusan strategis di tingkat lokal. Sistem yang dikembangkan diarahkan untuk menghasilkan peta kerawanan karhutla dengan akurasi dan keandalan yang lebih tinggi, sekaligus mendukung prioritas tindakan penanggulangan berdasarkan tingkat risiko kebakaran yang teridentifikasi. Selain itu, hasil penelitian ini diharapkan dapat menjadi dasar ilmiah dalam penyusunan kebijakan mitigasi oleh instansi terkait, seperti BPBD, KLHK, dan pemerintah daerah.

## 2. METODE

### 2.1. Data collection

Data *collection* dilakukan menjadi dua bagian yaitu pengambilan foto udara dengan bantuan UAV dan pengambilan data lapangan untuk memvalidasi model dalam pendugaan karakteristik daerah terjadi kebakaran seperti area terjadi kebakaran lahan.

Data yang digunakan dalam penelitian ini diperoleh dari dua sumber utama yang relevan dengan analisis cuaca dan hotspot kebakaran, khususnya di wilayah Kabupaten Mempawah, Kalimantan Barat. Pertama, data cuaca yang mencakup variabel seperti suhu udara (*temperatur*), kelembaban udara (*humidity*), curah hujan (*rainfall*), dan kecepatan angin (*wind speed*) diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Dataset ini mencatat kondisi cuaca harian di wilayah Mempawah selama periode 1 Januari 2023 hingga 22 Maret 2024, dengan total 1080 entry. Data cuaca tersebut dapat diakses melalui portal resmi BMKG di <https://dataonline.bmkg.go.id/dataonline-home>.

Kedua, data mengenai hotspot kebakaran hutan dan lahan diperoleh dari Badan Riset dan Inovasi Nasional (BRIN) melalui platform <https://data.brin.go.id/>. Data ini mencakup informasi terkait lokasi dan intensitas hotspot yang terdeteksi di wilayah Mempawah selama periode yang sama.

Tabel 1. Metadata Dataset

No	Date	Latitude	Longitude	Confidence	Temperature	Humidity	Rainfall	Wind Speed
0	01/01/2023	0.450680	109.0634613	8	32.5	84	0.0	5
1	01/01/2023	0.453569	109.0591271	8	32.5	84	0.0	5
2	02/01/2023	0.451459	109.06387329	8	30.8	86	0.0	4
3	02/01/2023	0.452000	109.06044006	8	30.8	86	0.0	4
4	02/01/2023	0.447479	109.06324768	8	30.8	86	0.0	4
...	...	...	...	...	...	...	...	...
1075	15/10/2023	0.389263	109.06603241	8	32.4	85	0.0	2
1076	15/10/2023	0.385301	109.07318115	8	32.4	85	0.0	2
1077	16/10/2023	0.386018	109.07308197	8	33.0	92	2.6	4
1078	06/02/2024	0.368347	109.03370667	8	32.2	88	33.6	5
1079	22/03/2024	0.343276	109.03041077	8	34.8	69	0.0	4

### 2.2. Data Preparation

Data *Preparation* dibagi dalam tiga tahapan yaitu *data selection*, *data cleaning*, dan *data transformatio*. *data selection*, *data cleaning*, dan *data transformation*. *Data selection* ialah melakukan seleksi parameter data *hotspot* yaitu parameter id, tanggal (WIB), waktu (WIB), lintang, bujur, tingkat kepercayaan, satelit, radius kemungkinan, kecamatan, kabupaten, jarak sungai. Seleksi parameter yang digunakan dalam dataset iklim yaitu curah hujan, kelembaban, dan suhu. Tahap *data cleansing* dilakukan proses pembersihan data untuk menghindari *missing value*. Selanjutnya *data transformation* merubah bentuk data kedalam proses algoritma yang digunakan dengan melakukan normalisasi data.

Proses *data preparation* dimulai dengan memuat data dari file *Excel*, kemudian mengonversi kolom tanggal (*Date*) ke format *datetime* untuk memastikan kesesuaian tipe data. Selanjutnya, kolom *longitude* dan *latitude* yang awalnya menggunakan koma sebagai pemisah desimal dikonversi menjadi titik dan diubah ke dalam tipe *float*. Kolom-kolom numerik lainnya seperti *confidence*, *temperature*, *humidity*, *rainfall*, dan *wind\_speed* juga dikonversi ke tipe numerik, dan nilai yang hilang pada kolom-kolom tersebut diisi menggunakan nilai rata-rata masing-masing kolom untuk menjaga konsistensi data. Setelah proses pembersihan selesai, data disimpan kembali dalam format CSV agar dapat digunakan

kembali. Terakhir, dipilih lima fitur numerik utama yang relevan untuk keperluan analisis clustering, yaitu *confidence*, *temperature*, *humidity*, *rainfall*, dan *wind\_speed*.

Tabel 2. Data Preparation

No	Date	Latitude	Longitude	Confidence	Temperature	Humidity	Rainfall	Wind Speed
0	01/01/2023	0.450680	109.063461	8	32.5	84	0.0	5
1	01/01/2023	0.453569	109.059128	8	32.5	84	0.0	5
2	02/01/2023	0.451459	109.063873	8	30.8	86	0.0	4
3	02/01/2023	0.452000	109.060440	8	30.8	86	0.0	4
4	02/01/2023	0.447479	109.063248	8	30.8	86	0.0	4
...	...	...	...	...	...	...	...	...
1075	15/10/2023	0.389263	109.066032	8	32.4	85	0.0	2
1076	15/10/2023	0.385301	109.073181	8	32.4	85	0.0	2
1077	16/10/2023	0.386018	109.073082	8	33.0	92	2.6	4
1078	06/02/2024	0.368347	109.033707	8	32.2	88	33.6	5
1079	22/03/2024	0.343276	109.030411	8	34.8	69	0.0	4

### 2.2.1. Normalisasi Data

Normalisasi bertujuan mengubah nilai fitur data ke dalam rentang yang sama, biasanya antara 0 dan 1, agar setiap fitur memiliki skala yang setara dan tidak ada fitur yang mendominasi karena nilai numerik yang lebih besar. Proses normalisasi dilakukan menggunakan metode Min-Max Scaling untuk mengubah skala data numerik ke rentang 0 hingga 1.

Rumus Min-Max Scaling

Untuk setiap nilai data  $x$  pada fitur tertentu, nilai tersebut diubah menjadi  $x'$  dengan rumus:

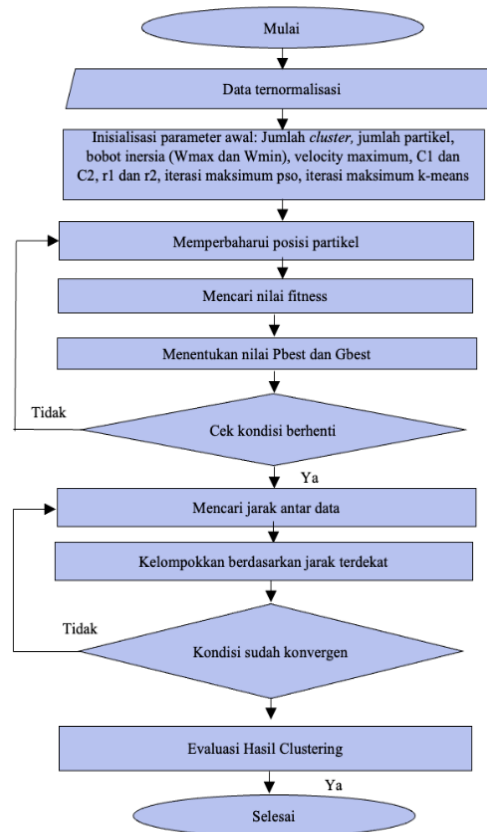
$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

Dimana:

- $x$  = nilai asli pada fitur tersebut
- $x_{\min}$  = nilai minimum pada fitur tersebut di seluruh dataset
- $x_{\max}$  = nilai maksimum pada fitur tersebut di seluruh dataset
- $x'$  = nilai hasil normalisasi yang berada di rentang [0, 1]

### 2.3. Modelling PSO-KMeans

Algoritma *Particle Swarm Optimization* (PSO) pencarian solusi dilakukan secara acak dari suatu populasi dimana setiap partikel memiliki nilai *fitness* yang harus dievaluasi untuk setiap generasi berdasarkan *local best* (pbest) dan *global best* (gbest) yang merupakan pengalaman dari setiap partikel dalam menghasilkan solusi terbaik [24]-[25]. PSO digunakan untuk menentukan jumlah kluster optimal serta menginisialisasi titik pusat (centroid) awal pada algoritma K-Means. Tahapan penelitian terlihat pada Gambar 1.



Gambar 1. Langkah Penelitian Optimasi Algoritma K-Means menggunakan Particle Swarm Optimization( PSO)

Particle Swarm Optimization (PSO) adalah algoritma optimasi yang terinspirasi dari perilaku sosial kawanan burung atau ikan. Algoritma ini bekerja dengan menggerakkan sekelompok partikel (solusi kandidat) dalam ruang pencarian untuk menemukan solusi optimal. Gerakan partikel dipengaruhi oleh pengalaman terbaik masing-masing partikel dan seluruh populasi [26]-[27]. Berikut merupakan tahapan proses optimasi PSO yang dilakukan:

1. Inisialisasi Populasi Partikel

Pada tahap ini, dilakukan inisialisasi posisi dan kecepatan awal sejumlah partikel  $N$

secara acak dalam ruang pencarian jumlah kluster. Setiap partikel mewakili solusi kandidat berupa nilai jumlah kluster  $k$ , dengan rentang [28]-[29]:

$$k \in [k_{min}, k_{max}] \tag{2}$$

Setiap partikel memiliki atribut berikut:

- Posisi  $x_i$  (nilai  $k$  kandidat),
- Kecepatan  $v_i$
- Posisi terbaik pribadi  $pbest_i$
- Posisi terbaik global  $gbest$

2. Evaluasi Nilai Fitness Awal

Setiap partikel dievaluasi berdasarkan hasil klusterisasi K-Means menggunakan nilai  $k = x_i$  untuk menghasilkan pusat kluster dan distribusi data. Evaluasi dilakukan menggunakan tiga metrik:

- Davies-Bouldin Index (DBI):

$$DBI = \frac{1}{k} \sum_{i=1}^k \text{Max}_{j \neq i} \left( \frac{S_i + S_j}{M_{ij}} \right) \quad (3)$$

Dimana  $S_i$  adalah deviasi rata-rata dalam kluster  $i$   $M_{ij}$  adalah jarak antara centroid kluster  $i$  dan  $j$

- Silhouette Coefficient (SC):

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (4)$$

Dimana  $a(i)$  adalah rata-rata jarak dari titik ke titik lain dalam kluster yang sama, dan  $b(i)$  adalah jarak minimum ke titik dari kluster terdekat lainnya.

- Inertia (Within-Cluster Sum of Squares / WCSS):

$$\text{Inertia} = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (5)$$

Di mana  $C_i$  adalah himpunan data pada kluster ke- $i$ , dan  $\mu_i$  adalah centroid dari kluster ke- $i$ .

Fungsi fitness disusun untuk meminimalkan DBI dan inertia, serta memaksimalkan SC. Contoh bentuk fungsi fitness gabungan:

$$\text{Fitness} = \alpha \cdot \left( \frac{1}{SC + \varepsilon} \right) + \beta \cdot DBI + \gamma \cdot \text{Inertia} \quad (6)$$

Dengan  $\alpha, \beta, \gamma$  adalah bobot yang ditentukan sesuai prioritas (misal: sama rata = 1), dan  $\varepsilon$  adalah nilai kecil untuk menghindari pembagian nol.

### 3. Pembaruan Kecepatan dan Posisi Partikel

Setiap partikel memperbarui kecepatannya  $v_i$  dan posisinya  $x_i$  menggunakan persamaan berikut [30]-[31]:

- Update kecepatan:

$$v_i(t + 1) = w \cdot v_i(t) + c_1 \cdot r_1 \cdot (pbest_i - x_i(t)) + c_2 \cdot r_2 \cdot (gbest - x_i(t)) \quad (7)$$

- Update posisi:

$$x_i(t + 1) = x_i(t) + v_i(t + 1) \quad (8)$$

Keterangan:

- $w$  = bobot inersia (0.5 – 1),
- $c_1, c_2$  = koefisien percepatan (biasanya 1.5 – 2),
- $r_1, r_2$  = angka acak [0,1],
- $pbest_i$  = posisi terbaik partikel  $i$
- $gbest$  = posisi terbaik dari seluruh partikel.

### 4. Update Nilai pbest dan gbest

Setelah posisi diperbarui, partikel dievaluasi kembali. Jika nilai fitness partikel lebih baik dari sebelumnya (pbest), maka diperbarui. Jika fitness lebih baik dari nilai global terbaik saat ini (gbest), maka gbest juga diperbarui.

5. Iterasi Hingga Konvergen

Langkah 2–4 diulang hingga jumlah iterasi maksimum tercapai atau nilai *gbest* tidak mengalami perubahan signifikan dalam beberapa iterasi berturut-turut . Solusi akhir berupa:

- Jumlah kluster optimal  $k^* = gbest$
- Inisialisasi titik pusat terbaik dari partikel terbaik,
- Nilai DBI, SC, dan inerti yang optimal.

6. Finalisasi Klasterisasi dengan K-Means

Hasil PSO berupa nilai *k* dan centroid digunakan untuk inisialisasi akhir pada algoritma K-Means. Hal ini memperbaiki kelemahan K-Means dalam hal sensitivitas terhadap pemilihan centroid awal.

Setelah proses evaluasi selesai, kualitas kluster yang terbentuk diklasifikasikan berdasarkan nilai Silhouette Score (SC) dan Davies-Bouldin Index (DBI). Interpretasi nilai-nilai ini penting untuk menentukan seberapa baik struktur kluster yang dihasilkan oleh proses optimasi. Tabel berikut menunjukkan kategori kualitas kluster berdasarkan nilai SC dan DBI:

Tabel 1. Kategori Kualitas Kluster Berdasarkan SC dan DBI

<i>Silhouette Score (SC)</i>	Kategori Struktur Cluster	<i>Davies-Bouldin Index (DBI)</i>	Kategori DBI
$>0.7$	Struktur cluster sangat kuat	Sangat rendah	Cluster sangat baik
$0.6 - 0.7$	Struktur cluster kuat	Rendah	Cluster baik
$0.4 - 0.6$	Struktur cluster standar	Sedang	Cluster standar
$0.2 - 0.4$	Struktur cluster lemah	Tinggi	Cluster kurang baik
$\leq 0.2$	Struktur cluster sangat lemah	Sangat tinggi	Cluster sangat buruk

Dengan acuan klasifikasi ini, interpretasi hasil evaluasi kuantitatif menjadi lebih sistematis, sehingga kualitas kluster yang dihasilkan oleh algoritma PSO-KMeans dapat dijelaskan secara lebih objektif dan konsisten.

3. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil eksperimen dari proses klasterisasi titik panas kebakaran hutan dan lahan (karhutla) menggunakan algoritma K-Means yang dioptimasi dengan Particle Swarm Optimization (PSO). Evaluasi dilakukan berdasarkan variasi parameter iterasi dan populasi partikel untuk menentukan jumlah kluster optimal. Hasil pengelompokan dianalisis menggunakan metrik Davies-Bouldin Index (DBI) dan Silhouette Coefficient (SC), serta divisualisasikan menggunakan metode PCA guna melihat pola sebaran kluster yang terbentuk.

3.1. Clustering dengan Particle Swarm Optimization (PSO)

Algoritma Particle Swarm Optimization (PSO) digunakan untuk mengoptimalkan jumlah kluster pada metode KMeans. PSO mengevaluasi kinerja kluster berdasarkan dua metrik utama, yaitu Davies-Bouldin Index (DBI) dan Silhouette Score (SC). PSO mencari jumlah kluster yang optimal dengan mempertimbangkan batasan jumlah kluster antara 2 dan 10. Setelah menemukan jumlah kluster yang optimal, kualitas kluster dievaluasi menggunakan nilai DBI dan SC, yang kemudian dikategorikan untuk menggambarkan kekuatan struktur kluster. Hasil eksperimen menunjukkan bahwa kombinasi parameter iterasi dan ukuran populasi berpengaruh pada kualitas kluster yang dihasilkan, dengan nilai DBI dan SC yang dihitung dan dikategorikan untuk memberikan gambaran tentang seberapa baik atau buruk kluster yang ditemukan.

Tabel 2. Parameter Particle Swarm Oprimization

Parameter PSO	Deskripsi	Nilai yang Digunakan	
maxiter (iterasi)	Jumlah iterasi maksimum	20, 50, 75	[31]
swarmsize (popsize)	Ukuran populasi partikel dalam PSO, menentukan jumlah partikel yang dievaluasi pada setiap iterasi.	5, 10, 20	
lb (lower bound)	Batas bawah untuk jumlah kluster yang diuji	2	[32]
ub (upper bound)	Batas atas untuk jumlah kluster yang diuji	10	
evaluate_k_pso	Fungsi evaluasi untuk menghitung kualitas kluster berdasarkan DBI dan SC.	Fungsi evaluate_k_pso	

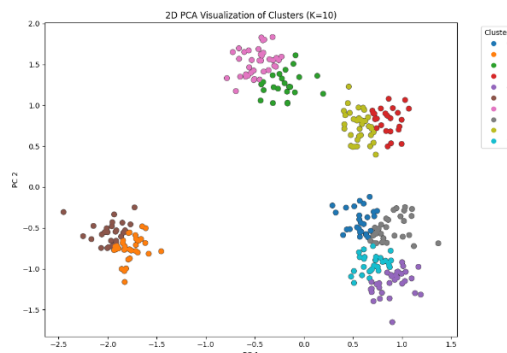
### 3.2. Evaluasi Kluster PSO

Tabel 3. Evaluasi Kualitas Kluster PSO

Iterasi	Populasi	PSO - Optimal K	PSO - DBI	PSO - Silhouette Score	PSO - SC Category	PSO - DBI Category
20	5	10	1.104	0.328	kluster lemah	tinggi (buruk)
20	10	10	0.983	0.357	kluster lemah	tinggi (buruk)
20	20	10	0.897	0.464	kluster standar	sedang
50	5	10	1.044	0.344	kluster lemah	tinggi (buruk)
50	10	10	0.974	0.360	kluster lemah	tinggi (buruk)
50	20	10	0.981	0.356	kluster lemah	tinggi (buruk)
75	5	10	1.041	0.346	kluster lemah	tinggi (buruk)
75	10	10	0.914	0.412	kluster standar	sedang
75	20	10	0.905	0.455	kluster standar	sedang

Berdasarkan hasil evaluasi dengan variasi iterasi dan jumlah populasi pada algoritma PSO, diperoleh bahwa nilai *Optimal K* cenderung stabil pada angka 10 di seluruh kombinasi parameter. Namun, kualitas kluster yang terbentuk menunjukkan variasi yang signifikan. Nilai DBI (Davies-Bouldin Index) tertinggi ditemukan pada iterasi 20 dengan populasi 5 (DBI = 1.104), menunjukkan kualitas kluster yang buruk, sementara nilai DBI terbaik (terendah) sebesar 0.897 tercapai pada iterasi 20 dan populasi 20, mengindikasikan kluster yang lebih baik. Hal ini sejalan dengan nilai *Silhouette Score* yang tertinggi sebesar 0.464 pada konfigurasi tersebut, menunjukkan struktur kluster yang lebih jelas. Secara umum, peningkatan jumlah populasi cenderung memperbaiki kualitas kluster, ditunjukkan dengan pergeseran dari kategori "Struktur kluster lemah" dan "DBI tinggi (buruk)" ke "Struktur kluster standar" dan "DBI sedang". Ini menandakan bahwa konfigurasi PSO yang lebih kompleks dapat meningkatkan efektivitas dalam pengelompokan data.

### 3.3. Visualisasi Kluster PSO



Gambar 2. Visualisasi Kluster

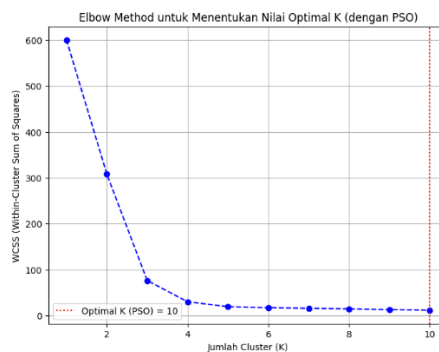
Visualisasi hasil PCA dengan jumlah kluster 10 menunjukkan bahwa data yang telah direduksi ke dalam dua dimensi utama mampu merepresentasikan variasi antar data secara cukup baik. Titik-titik data dalam plot PCA tersebar membentuk kelompok-kelompok yang merepresentasikan masing-masing kluster hasil optimasi PSO. Meskipun terdapat tumpang tindih antar beberapa kluster, sebagian besar kluster terlihat memiliki pemisahan yang cukup jelas, terutama pada konfigurasi PSO dengan iterasi dan populasi tinggi. Hal ini memperkuat hasil evaluasi kuantitatif, di mana konfigurasi dengan iterasi 20 dan populasi 20 menghasilkan DBI terendah dan *silhouette score* tertinggi, yang artinya kluster memiliki batas yang lebih tegas dan struktur yang lebih kuat. PCA berperan penting dalam membantu memvisualisasikan struktur internal data dan mendukung interpretasi bahwa jumlah kluster 10 adalah solusi optimal berdasarkan hasil evaluasi dan visualisasi.

Hasil eksperimen yang dilakukan menunjukkan bahwa algoritma Particle Swarm Optimization (PSO) berhasil meningkatkan performansi algoritma K-Means dalam menentukan jumlah kluster optimal untuk identifikasi daerah rawan kebakaran hutan dan lahan di Kabupaten Mempawah. Evaluasi dilakukan dengan menguji beberapa konfigurasi parameter jumlah iterasi (20, 50, 75) dan ukuran populasi (5, 10, 20). Secara keseluruhan, PSO secara konsisten memilih jumlah kluster optimal sebanyak 10. Konfigurasi terbaik ditemukan pada iterasi 20 dengan populasi 20, menghasilkan nilai Davies-Bouldin Index (DBI) terendah sebesar 0.897 dan Silhouette Score (SC) tertinggi sebesar 0.464, yang dikategorikan sebagai struktur kluster standar dan DBI sedang. Sebaliknya, konfigurasi dengan iterasi 20 dan populasi 5 menghasilkan nilai DBI tertinggi (1.104) dan SC terendah (0.328), yang menunjukkan struktur kluster lemah. Hasil ini mengindikasikan bahwa peningkatan populasi partikel pada PSO dapat meningkatkan kualitas kluster secara signifikan. Visualisasi hasil kluster menggunakan PCA menunjukkan bahwa penyebaran kluster pada jumlah K=10 cukup terdefinisi dengan baik. Hasil ini menguatkan metrik kuantitatif bahwa PSO mampu memberikan solusi optimal dalam pengelompokan titik panas karhutla berbasis data cuaca.

#### 4. DISKUSI

Hasil penelitian menunjukkan bahwa integrasi PSO dengan K-Means secara signifikan meningkatkan kualitas klusterisasi data hotspot kebakaran hutan, terutama dalam menentukan jumlah kluster optimal dan mengatasi kelemahan inisialisasi centroid pada K-Means. Dibandingkan dengan metode konvensional seperti Chebysev Distance K-Means, pendekatan ini lebih adaptif dan mampu menangani data spasial kompleks. Pemanfaatan PSO-K-K-Means dalam mengelompokkan data spasial untuk menentukan kluster optimal secara otomatis lebih komprehensif dan hasil klusterisasi lebih stabil.

##### 4.1. Visualisasi Kluster K- PSO



Gambar 3. Visualisasi Kluster K-PSO

Gambar 4 menunjukkan hasil Elbow Method yang digunakan untuk menentukan jumlah cluster optimal (K) dengan bantuan algoritma Particle Swarm Optimization (PSO). Grafik menggambarkan hubungan antara jumlah cluster (K) dengan nilai WCSS (Within-Cluster Sum of Squares), yang

mencerminkan seberapa baik data dikelompokkan. Terlihat bahwa nilai WCSS mengalami penurunan signifikan dari  $K=1$  hingga  $K=3$ , kemudian penurunannya mulai melambat hingga  $K=10$ . Secara visual, titik siku (elbow) terlihat pada  $K=3$  atau  $K=4$ , yang umumnya dianggap sebagai indikasi jumlah cluster optimal karena setelah titik tersebut, penambahan jumlah cluster tidak lagi memberikan pengurangan WCSS yang signifikan. Namun, garis vertikal merah yang ditampilkan pada  $K=10$  menunjukkan bahwa hasil optimasi menggunakan PSO menemukan nilai  $K$  optimal berada pada angka 10. Hal ini menunjukkan bahwa PSO mempertimbangkan faktor-faktor lain dalam pencarian solusi global yang mungkin tidak terlihat melalui analisis visual semata. Oleh karena itu, meskipun Elbow Method secara konvensional mungkin menyarankan  $K=3$  atau 4, hasil PSO menunjukkan bahwa  $K=10$  memberikan hasil clustering yang dianggap paling optimal. Untuk memastikan validitas hasil ini, sebaiknya dilakukan evaluasi tambahan menggunakan metrik seperti Silhouette Coefficient atau Davies-Bouldin Index.

Berdasarkan hasil yang ditunjukkan pada Gambar 6, metode Elbow digunakan untuk menentukan jumlah cluster optimal ( $K$ ) dengan mengamati nilai Within-Cluster Sum of Squares (WCSS) pada setiap nilai  $K$ . Hasil grafik menunjukkan bahwa nilai WCSS mengalami penurunan tajam hingga  $K=3$ , kemudian mulai melandai pada nilai  $K$  selanjutnya. Secara visual, titik elbow yang mengindikasikan jumlah cluster optimal berada pada kisaran  $K=3$  hingga  $K=4$ . Namun, hasil optimasi menggunakan Particle Swarm Optimization (PSO) menunjukkan bahwa jumlah cluster optimal adalah  $K=10$ , sebagaimana ditandai dengan garis vertikal merah pada grafik.

Secara visual nilai  $K=10$  tidak menunjukkan titik elbow yang jelas, PSO mampu mengevaluasi solusi secara lebih menyeluruh dan menemukan bahwa pada  $K=10$ , hasil klusterisasi menghasilkan pemisahan data yang lebih baik berdasarkan kriteria optimasi tertentu. Oleh karena itu, nilai  $K=10$  ditetapkan sebagai jumlah cluster optimal dalam proses klusterisasi ini. Hasil ini mengindikasikan bahwa PSO dapat memberikan solusi yang lebih akurat dibandingkan hanya mengandalkan metode visual tradisional. Validasi lebih lanjut terhadap pemilihan  $K=10$  dapat dilakukan melalui pengukuran metrik lain seperti Silhouette Score atau Davies-Bouldin Index (DBI) untuk memperkuat kesimpulan.

## 5. KESIMPULAN

Penelitian ini menunjukkan bahwa penggunaan algoritma Particle Swarm Optimization (PSO) dalam mengoptimalkan metode K-Means mampu meningkatkan performa klusterisasi pada data hotspot kebakaran hutan dan lahan di Kabupaten Mempawah, Kalimantan Barat. PSO secara efektif menentukan jumlah kluster optimal dan menghasilkan pembagian data yang lebih terstruktur serta memiliki kualitas kluster yang lebih baik. Untuk pengembangan kedepan, penelitian ini dapat ditingkatkan dengan integrasi data lingkungan multisumber, validasi temporal pada data multi tahun untuk menghasilkan informasi lebih komprehensif dan adaptif.

## UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada Majelis Pendidikan Tinggi Penelitian dan Pengembangan Pimpinan Pusat Muhammadiyah atas dukungan dana Penelitian dengan Nomor Kontrak 0258.35/I.3/D/2025 serta LPPM Universitas Muhammadiyah Pontianak yang telah memberikan support untuk penelitian ini.

## DAFTAR PUSTAKA

- [1] A. Rachman, B. H. Saharjo, and E. I. K. Putri, "Forest and Land Fire Prevention Strategies in the Forest Management Unit Kubu Raya, South Ketapang, and North Ketapang in West Kalimantan Province," *Jurnal Ilmu Pertanian Indonesia*, vol. 25, no. 2, pp. 213–223, Apr. 2020, doi: 10.18343/jipi.25.2.213.

- 
- [2] A. Amri *et al.*, “Dampak bencana kebakaran hutan terhadap lingkungan dan upaya penanggulangan di Indonesia,” *Jurnal Riset Tindakan Indonesia*, vol. 9, no. 2, pp. 159–166, 2024, doi: 10.29210/30035130000.
- [3] P. N. Utami and Y. Primawardani, “Upaya Pemenuhan Hak Atas Lingkungan Hidup Terhadap Kebakaran Hutan bagi Masyarakat Riau,” *Jurnal HAM*, vol. 12, no. 3, p. 367, Dec. 2021, doi: 10.30641/ham.2021.12.367-384.
- [4] I. P. Anhar, Rina Mardiana, and Rai Sita, “Dampak Kebakaran Hutan dan Lahan Gambut terhadap Manusia dan Lingkungan Hidup (Studi Kasus: Desa Bunsur, Kecamatan Sungai Apit, Kabupaten Siak, Provinsi Riau),” *Jurnal Sains Komunikasi dan Pengembangan Masyarakat [JSKPM]*, vol. 6, no. 1, pp. 75–85, Mar. 2022, doi: 10.29244/jskpm.v6i1.967.
- [5] N. Suryadiputra, “Manual For The Control Of Fire In Peatlands And Peatland Forest,” Wetlands International, Indonesia Programme and Wildlife Habitat Canada, 2005.
- [6] I. Kejadian, B. Karhutla, and D. Kalimantan Barat, “Satgas Informasi Bencana Bpbd Provinsi Kalimantan Barat Reaktif, Cepat dan Tangguh,” 2023.
- [7] A. Baroroh, “Studi Kasus: Pulau Rupa, Bengkalis,” 2018.
- [8] S. Primantoro, R. Goejantoro, and D. S. Prangga, “Clustering Titik Panas Bumi Pada Potensi Kebakaran Hutan Menggunakan K-Affinity Propagation Geothermal Hotspot Clustering on Forest Fire Potential Using K-Affinity Propagation,” vol. 15, no. 2, 2024, doi: 10.30872/eksponensial.v15i2.1299.
- [9] H. Maulana Ilmi, M. Kurniawan, U. Al Faruq, R. Rotul Muhima, K. kunci, and T. Panas Bumi, “Perbandingan K-Means Dan K-Medoids Untuk Pengelompokkan Data Titik Panas Bumi Di Pulau Kalimantan Comparison Of K-Means And K-Medoids For Hotspot Data Clustering On The Island Of Kalimantan,” vol. 13, no. 1, 2024.
- [10] F. N. Dhewayani *et al.*, “Implementasi K-Means Clustering untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM,” *Jurnal Teknologi dan Informasi*, doi: 10.34010/jati.v12i1.
- [11] R. Kusumastuti, E. Bayunanda, A. Muhammad Rifa, M. Ryandy Ghonim Asgar, and F. Inti Ilmawati, “Clustering Titik Panas Menggunakan Algoritma Agglomerative Hierarchical Clustering (AHC) Hot Spot Clustering Using Agglomerative Hierarchical Clustering (AHC) Algorithm,” *Cogito Smart Journal* |, vol. 8, no. 2.
- [12] M. Gul and M. A. Rehman, “Big data: an optimized approach for cluster initialization,” *\*J Big Data\**, vol. 10, no. 1, Dec. 2023, doi: 10.1186/s40537-023-00798-1.
- [13] A. A. Khan, M. S. Bashir, A. Batool, M. S. Raza, and M. A. Bashir, “K-Means Centroids Initialization Based on Differentiation Between Instances Attributes,” *International Journal of Intelligent Systems*, vol. 2024, no. 1, 2024, doi: 10.1155/2024/7086878.
- [14] M. E. Celebi, H. A. Kingravi, and P. A. Vela, “A Comparative Study of Efficient Initialization Methods for the K-Means Clustering Algorithm,” 2012.
- [15] A. N. Ikhsan, D. S. Hadmoko, and P. Widayani, “Spatial Modeling of Forest and Land Fire Susceptibility Using the Information Value Method in Kotawaringin Barat Regency, Indonesia,” *Fire*, vol. 6, no. 4, Apr. 2023, doi: 10.3390/fire6040170.
- [16] I. Daqiqil Id and Tr. Angraini, “Penentuan Daerah Rawan Titik Api di Provinsi Riau Menggunakan Clustering Algoritma K-Means (Determination of Fire Point Prone Areas in Riau Province Using Clustering K-Means Algorithm),” 2018.
- [17] A. Manik, E. B. Nababan, and T. Tulus, “Improved Support Vector Machine Performance Using Particle Swarm Optimization In Credit Risk Classification,” *Jurnal Teknik Informatika (Jutif)*, vol. 3, no. 6, pp. 1739–1746, Dec. 2022, doi: 10.20884/1.jutif.2022.3.6.615.
- [18] J. Ghorpade-Aher and V. A. Metre, “Clustering Multidimensional Data with PSO based Algorithm”, arXiv preprint arXiv:1402.6428. 2014 Feb 26.
-

- 
- [19] C. Gong, H. Chen, W. He, and Z. Zhang, "Improved multi-objective clustering algorithm using particle swarm optimization," *PLoS One*, vol. 12, no. 12, Dec. 2017, doi: 10.1371/journal.pone.0188815.
- [20] L. Imanuela Mustamu and Y. Sibaroni, "FUEL INCREASE SENTIMENT ANALYSIS USING SUPPORT VECTOR MACHINE WITH PARTICLE SWARM OPTIMIZATION AND GENETIC ALGORITHM AS FEATURE SELECTION," *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 3, pp. 521–528, Jun. 2023, doi: 10.52436/1.jutif.2023.4.3.881.
- [21] L. Dey and S. Chakraborty, "Canonical PSO Based K -Means Clustering Approach for Real Datasets," *Int Sch Res Notices*, vol. 2014, pp. 1–11, Nov. 2014, doi: 10.1155/2014/414013.
- [22] G. K. Patel, V. K. Dabhi, and H. B. Prajapati, "Clustering Using a Combination of Particle Swarm Optimization and K-means," *Journal of Intelligent Systems*, vol. 26, no. 3, pp. 457–469, Jul. 2017, doi: 10.1515/jisys-2015-0099.
- [23] S. Hasana and D. Fitriana, "A Study on Enhanced Spatial Clustering Using Ensemble DBscan and UMAP to Map Fire Zone in Greater Jakarta, Indonesia," *Jurnal Riset Informatika*, vol. 5, no. 3, pp. 409–418, Jun. 2023, doi: 10.34288/jri.v5i3.557.
- [24] E. Lu, L. Xu, Y. Li, Z. Ma, Z. Tang, and C. Luo, "A Novel Particle Swarm Optimization with Improved Learning Strategies and Its Application to Vehicle Path Planning," *Math Probl Eng*, vol. 2019, 2019, doi: 10.1155/2019/9367093.
- [25] L. Zajmi, F. Y. H. Ahmed, and A. A. Jaharadak, "Concepts, Methods, and Performances of Particle Swarm Optimization, Backpropagation, and Neural Networks," *Applied Computational Intelligence and Soft Computing*, vol. 2018, 2018, doi: 10.1155/2018/9547212.
- [26] D. Freitas, L. G. Lopes, and F. Morgado-Dias, "Particle Swarm Optimisation: A historical review up to the current developments," Mar. 01, 2020, *MDPI AG*. doi: 10.3390/E22030362.
- [27] A. G. Gad, "Particle Swarm Optimization Algorithm and Its Applications: A Systematic Review," *Archives of Computational Methods in Engineering*, vol. 29, no. 5, pp. 2531–2561, Aug. 2022, doi: 10.1007/s11831-021-09694-4.
- [28] A. M. Ikotun, F. Habyarimana, and A. E. Ezugwu, "Cluster validity indices for automatic clustering: A comprehensive review," Jan. 30, 2025, *Elsevier Ltd*. doi: 10.1016/j.heliyon.2025.e41953.
- [29] I. W. Septiani, Abd. C. Fauzan, and M. M. Huda, "Implementasi Algoritma K-Medoids Dengan Evaluasi Davies-Bouldin-Index Untuk Klasterisasi Harapan Hidup Pasca Operasi Pada Pasien Penderita Kanker Paru-Paru," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 3, no. 4, p. 556, Jul. 2022, doi: 10.30865/json.v3i4.4055.
- [30] Q. Li, S. Y. Liu, and X. S. Yang, "Influence of initialization on the performance of metaheuristic optimizers," *Applied Soft Computing Journal*, vol. 91, Jun. 2020, doi: 10.1016/j.asoc.2020.106193.
- [31] Y. Ruan *et al.*, "Application of K-PSO Clustering Algorithm and Game Theory in Rock Mass Quality Evaluation of Maji Hydropower Station," *Applied Sciences (Switzerland)*, vol. 13, no. 14, Jul. 2023, doi: 10.3390/app13148467.
-