



Analisis Komparatif *Firefly Algorithm* dan *Particle Swarm Optimization* dalam Optimasi *K-Means* untuk Pengelompokan Ketimpangan Pendapatan Antarprovinsi di Indonesia

Adinda^{1*}, Fahrezal Zubedi¹, Nisky Imansyah Yahya²

¹ Program Studi Statistika, Universitas Negeri Gorontalo, Indonesia.

² Program Studi Matematika, Universitas Negeri Gorontalo, Indonesia.

* Korespondensi: adindahamid12@gmail.com

Sitasi: A. Adinda, F. Zubedi, and N. I. Yahya, "Analisis Komparatif *Firefly Algorithm* dan *Particle Swarm Optimization* dalam Optimasi *K-Means* untuk Pengelompokan Ketimpangan Pendapatan Antarprovinsi di Indonesia", *Jurnal Teknologi Informasi Dan Multimedia*, vol. 8, no. 3, hlm. 535-550, 2026.

<https://doi.org/10.35746/jtim.v8i3.1026>

Diterima: 05-05-2026

Direvisi: 02-06-2026

Disetujui: 30-06-2026



Copyright: © 2026 oleh para penulis. Karya ini dilisensikan di bawah Creative Commons Attribution-ShareAlike 4.0 International License. (<https://creativecommons.org/licenses/by-sa/4.0/>).

Abstract: *Income inequality among provinces in Indonesia reflects differences in welfare levels influenced by the social and economic characteristics of each region. This study aims to compare the performance of Particle Swarm Optimization (PSO) and Firefly Algorithm (FA) in optimizing the number of clusters in K-Means clustering to group Indonesian provinces based on seven factors related to income inequality, namely Human Development Index (HDI), Number of Poor Population, Open Unemployment Rate, Inflation, Provincial Minimum Wage, GDP per Capita, and Labor Force Participation Rate. The data used are secondary data from 2024 covering 38 provinces. The analytical methods include data standardization using Z-Score, K-Means clustering, and cluster number optimization using PSO and FA evaluated by Silhouette Coefficient (SC). The results show that both optimization methods improved clustering quality compared to K-Means without optimization, which yielded an SC of 0.23. PSO produced an SC of 0.28 with an optimal cluster number of 5, while FA produced an SC of 0.39 with an optimal cluster number of 3. FA proved superior in generating a more optimal and representative clustering structure. The clustering results reveal distinct characteristics among clusters that can serve as a basis for formulating more targeted income inequality reduction policies in accordance with the characteristics of each regional group.*

Keywords: *K-Means, PSO, FA, Silhouette Coefficient, Income Inequality*

Abstrak: Ketimpangan pendapatan antarprovinsi di Indonesia mencerminkan perbedaan tingkat kesejahteraan yang dipengaruhi oleh karakteristik sosial dan ekonomi wilayah. Penelitian ini bertujuan untuk membandingkan kinerja *Particle Swarm Optimization* (PSO) dan *Firefly Algorithm* (FA) dalam mengoptimalkan Penentuan jumlah kluster pada metode *K-Means* dilakukan untuk mengelompokkan provinsi di Indonesia berdasarkan tujuh faktor yang memengaruhi ketimpangan pendapatan, yaitu IPM, Jumlah Penduduk Miskin, TPT, Inflasi, UMP, PDRB per Kapita, dan TPAK. Data yang digunakan merupakan data sekunder tahun 2024 yang mencakup 38 provinsi. Metode analisis meliputi standarisasi data menggunakan *Z-Score*, klusterisasi *K-Means*, serta optimasi jumlah cluster menggunakan PSO dan FA dengan evaluasi *Silhouette Coefficient* (SC). Hasil penelitian menunjukkan bahwa kedua metode optimasi mampu meningkatkan kualitas clustering dibandingkan *K-Means* tanpa optimasi yang menghasilkan SC sebesar 0,23. PSO menghasilkan nilai SC sebesar 0,28 dengan lima kluster optimal, sedangkan FA memperoleh nilai SC sebesar 0,39 dengan tiga kluster optimal. FA terbukti lebih unggul dalam menghasilkan struktur pengelompokan yang lebih optimal dan representatif. Hasil pengelompokan menunjukkan perbedaan karakteristik antarcluster yang dapat dijadikan dasar perumusan kebijakan penanggulangan ketimpangan pendapatan yang lebih tepat sasaran sesuai karakteristik masing-masing wilayah.

Kata kunci: *K-Means, PSO, FA, Silhouette Coefficient, Ketimpangan Pendapatan*

1. Pendahuluan

Ketimpangan pendapatan antar wilayah provinsi di Indonesia menjadi tantangan struktural yang belum terselesaikan dalam proses pembangunan nasional [1]. Pesatnya laju pertumbuhan ekonomi tidak secara otomatis diikuti oleh distribusi kesejahteraan yang merata di seluruh lapisan masyarakat, sehingga ketidakseimbangan dalam sebaran pendapatan terus melebar. Kondisi ini mencerminkan perbedaan taraf hidup antar-masyarakat, di mana semakin tingginya ketimpangan menandakan distribusi pendapatan yang kian tidak proporsional. Sebagai negara dengan karakteristik geografis kepulauan yang sangat beragam, Indonesia menghadapi kesenjangan pembangunan antardae-rah yang pada akhirnya memicu terjadinya disparitas pendapatan [2]. Rasio Gini merupakan salah satu indikator yang umum digunakan untuk mengukur tingkat ketimpangan pendapatan [3].

Berdasarkan data Badan Pusat Statistik [4], nilai Rasio Gini di Indonesia memperlihatkan kecenderungan yang tidak stabil dari tahun ke tahun, namun secara keseluruhan masih berada dalam kisaran ketimpangan tingkat menengah. Hal ini mengindikasikan bahwa isu kesenjangan pendapatan masih memerlukan perhatian serius dari berbagai pihak. Ketimpangan tersebut dipengaruhi oleh berbagai faktor, baik yang berasal dari aspek internal maupun eksternal, seperti Indeks Pembangunan Manusia (IPM), Tingkat Pengangguran Terbuka (TPT), inflasi, Upah Minimum Provinsi (UMP), dan Tingkat Partisipasi Angkatan Kerja (TPAK).

Kajian-kajian terdahulu pada umumnya bertumpu pada pendekatan analisis regresi dalam menelaah keterkaitan antarvariabel terhadap ketimpangan pendapatan [6]. Namun demikian, pendekatan tersebut memiliki keterbatasan dalam pengelompokan wilayah berdasarkan kemiripan karakteristiknya. Oleh sebab itu, metode *clustering* sebagai alternatif analisis untuk mengklasifikasikan provinsi berdasarkan kesamaan ciri yang dimilikinya [7]. Salah satu teknik pengelompokan yang banyak digunakan adalah algoritma *K-Means*, mengingat kesederhanaan dan efisiensinya dalam implementasi [8]. Meskipun demikian, *K-Means* memiliki keterbatasan dalam penetapan jumlah *cluster* optimal serta rentan terhadap sensitivitas inisialisasi centroid [9].

Untuk mengatasi kelemahan tersebut, digunakan algoritma optimasi berbasis metaheuristik, yaitu Particle Swarm Optimization (PSO) dan Firefly Algorithm (FA). PSO terinspirasi dari perilaku kawanan hewan yang bergerak mencari solusi optimal berdasarkan *personal best* dan *global best*, sedangkan FA terinspirasi dari perilaku *firefly* yang melakukan pergerakan menuju individu dengan intensitas cahaya yang lebih tinggi, yang mencerminkan solusi dengan nilai objektif yang lebih baik. Harliana dkk. (2021) membuktikan PSO menghasilkan cluster lebih representatif, sementara Wariantia dkk. (2024) membuktikan FA menghasilkan pengelompokan lebih stabil. Namun, studi yang membandingkan keduanya secara langsung dalam konteks ketimpangan pendapatan antarprovinsi di Indonesia belum ada. Atas dasar tersebut, penelitian ini membandingkan performa PSO dan FA dalam mengoptimalkan *K-Means* dalam mengelompokkan 38 provinsi di Indonesia berdasarkan faktor penyebab ketimpangan pendapatan tahun 2024.

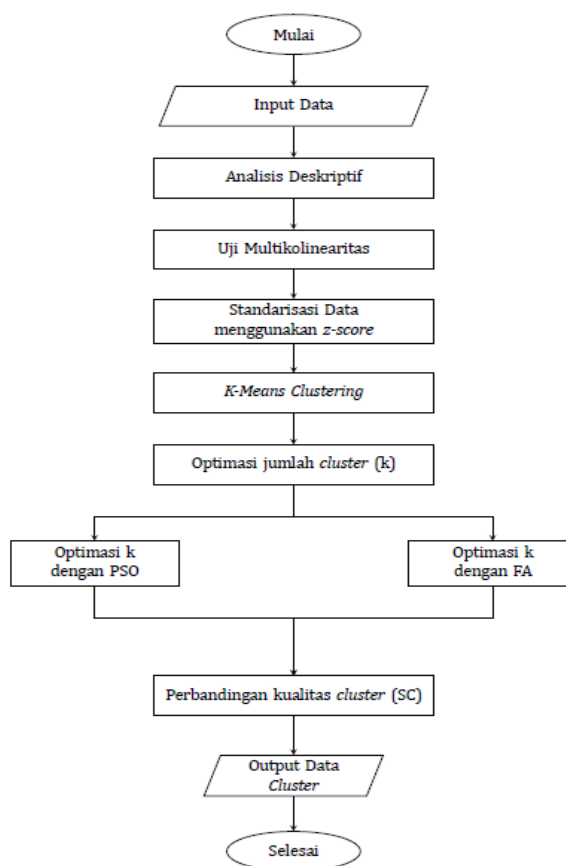
2. Bahan dan Metode

Data sekunder yang digunakan dalam penelitian ini berasal dari publikasi resmi Badan Pusat Statistik (BPS) Tahun 2024., yang tersedia pada laman <https://www.bps.go.id/id>. Selain itu, data mengenai Upah Minimum Provinsi (UMP) diperoleh melalui situs <https://ikpi.or.id/>. Ruang lingkup analisis mencakup seluruh 38 provinsi yang ada di Indonesia. Variabel-variabel yang ditetapkan sebagai indikator penelitian dapat dilihat pada tabel berikut:

Tabel 1. Variabel Penelitian

Simbol	Variabel	Satuan
X1	Indeks Pembangunan Manusia (IPM)	Indeks
X2	Jumlah Penduduk Miskin (JPM)	Ribu
X3	Tingkat Pengangguran Terbuka (TPT)	Persen (%)
X4	Inflasi	Persen (%)
X5	Upah Minimum Provinsi (UMP)	Ribu
X6	PDRB Per Kapita	Ribu
X7	Tingkat Partisipasi Angkatan Kerja (TPAK)	Persen (%)

Studi ini menerapkan pendekatan kuantitatif komparatif. Maksudnya adalah membandingkan performa *Particle Swarm Optimization* (PSO) dan *Firefly Algorithm* (FA) dalam menyempurnakan output pengelompokan *K-Means*. Hal ini bertujuan menentukan jumlah kluster optimal dengan kualitas pengelompokan terbaik, yang diukur menggunakan *Silhouette Coefficient* (SC). Gambar 1 dibawah menyajikan tahapan penelitian yang dilakukan.



Gambar 1. Tahapan Penelitian

2.1. Analisis Cluster

Analisis kluster adalah teknik analisis multivariat yang bertujuan mengelompokkan objek ke dalam beberapa kluster berdasarkan kesamaan karakteristik [10]. Objek dengan tingkat kemiripan yang tinggi akan tergabung dalam kluster yang sama, sehingga setiap kluster memiliki tingkat homogenitas yang tinggi di dalam kelompoknya, namun menunjukkan perbedaan yang jelas dibandingkan dengan kluster lainnya [11]. Teknik ini berguna untuk menyederhanakan data rumit dengan memasukkan objek mirip ke kelompok yang seragam. Dalam penggunaannya, analisis pengelompokan dibagi menjadi dua jenis yaitu *hierarchical* dan *non-hierarchical*. Salah satu teknik *non-hierarchical* yang paling banyak digunakan adalah *K-Means*, yang masuk kategori pengelompokan partisi.

Metode ini membagi data ke beberapa kelompok dengan jumlah yang sudah ditentukan sebelumnya.

2.2. Uji Multikolinearitas

Analisis *cluster* mengasumsikan tidak adanya korelasi antar variabel, sehingga pengujian multikolinearitas perlu dilakukan untuk mengidentifikasi hubungan antar variabel [12]. Salah satu pendekatan yang digunakan adalah *Variance Inflation Factor* (VIF), yang diperoleh melalui rumus di bawah ini:

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

dengan R^2 merupakan koefisien determinasi antar variabel independen. Nilai VIF di atas 10 menunjukkan adanya multikolinearitas, sedangkan VIF kurang dari atau sama dengan 10 berarti tidak ada multikolinearitas.

2.3. Standarisasi Data

Standarisasi data dilakukan untuk mengatasi perbedaan satuan antar variabel agar tidak memengaruhi hasil analisis [13]. Salah satu cara yang sering dipakai adalah transformasi *Z-score*, dengan rumus sebagai berikut:

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

di mana Z merujuk pada nilai standar, x_i merupakan nilai data, \bar{x} adalah nilai rata-rata, serta σ menunjukkan standar deviasi. Cara ini bertujuan menyamakan ukuran data agar semua variabel punya pengaruh yang seimbang saat dianalisis.

2.4. K-Means

K-Means metode pengelompokan nonhierarkis yang membagi data menjadi beberapa kluster berdasarkan tingkat kemiripan karakteristik yang dimiliki setiap objek. Metode ini bertujuan yang meminimalkan variasi dalam *cluster* serta memaksimalkan perbedaan antar-*cluster* melalui proses berulang. Berikut adalah tahapan algoritma *K-Means*:

1. Menetapkan jumlah *cluster* (k)
2. Menginisialisasi pusat *cluster* (centroid) secara *random*.
3. Menghitung jarak setiap data ke centroid terdekat menggunakan jarak Euclidean:

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

dengan $d(x, y)$ adalah jarak antara data dan centroid, x_i merupakan nilai data ke- i , y_i menunjukkan nilai centroid ke- i , dan n sebagai jumlah variabel.

4. Mengelompokkan data ke dalam *cluster* dengan jarak terdekat.
5. Memperbarui pusat *cluster* dengan cara menghitung rata-rata dari anggota-anggota *cluster*:

$$C_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i \quad (4)$$

dengan C_k merujuk pada centroid *cluster* ke- k , n_k adalah jumlah observasi dalam *cluster* ke- k , serta x_i merupakan data ke- i dalam *cluster*.

6. Mengulangi langkah 3–5 hingga tidak terjadi perubahan keanggotaan *cluster*.

2.5. Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) yaitu salah satu algoritma optimasi berbasis *swarm intelligence* yang mencontoh perilaku kolektif anggota kawanan dalam menjelajahi ruang pencarian guna memperoleh solusi yang paling optimal. Dalam PSO, setiap kandidat solusi dimodelkan sebagai partikel yang bergerak di ruang pencarian, dengan mempertimbangkan pengalaman terbaik masing-masing individu (*personal best* atau *pbest*) serta pengalaman terbaik keseluruhan populasi (*global best* atau *gbest*). Pembaruan posisi partikel dilakukan melalui persamaan di bawah ini:

$$v_i^{t+1} = \omega v_i^t + a_1 r_1 (pbest_i - x_i^t) + a_2 r_2 (gbest - x_i^t) \tag{5}$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \tag{6}$$

Keterangan :

- v_i^t : kecepatan partikel ke-*i* pada iterasi ke-*t*
- x_i^t : posisi partikel ke-*i*
- ω : bobot inersia
- a_1 dan a_2 : koefisien pembelajaran
- r_1 dan r_2 : bilangan acak pada interval [0, 1].

Prosedur algoritma PSO diawali dengan inisialisasi posisi dan kecepatan partikel secara acak, dilanjutkan dengan evaluasi fungsi objektif pada setiap partikel. Berikutnya, nilai *pbest* ditetapkan sebagai solusi terbaik individu dan *gbest* sebagai solusi terbaik secara global. Posisi dan kecepatan setiap partikel kemudian diperbarui sesuai persamaan yang berlaku. Proses iterasi terus berjalan hingga memenuhi kriteria penghentian, baik berupa konvergensi maupun tercapainya batas iterasi maksimum. Tabel berikut menyajikan seluruh parameter yang ditetapkan dalam penerapan algoritma PSO pada penelitian ini:

Tabel 2. Parameter PSO

Parameter	Nilai
Jumlah Partikel	15
Maksimum Iterasi	100
Bobot Inersia (ω)	0,9
Koefisien individu (a_1)	1,5
Koefisien kelompok (a_2)	1,0
Batas minimum k	3
Batas maksimum k	10

2.6. Firefly Algorithm (FA)

Firefly Algorithm (FA) adalah algoritma optimasi *swarm intelligence* yang diinspirasi dari pola perilaku *firefly* saat berinteraksi dan berkelompok berdasarkan intensitas cahayanya. Dalam algoritma ini, setiap *firefly* mewakili sebuah solusi, di mana tingkat kecerahannya ditentukan oleh nilai fungsi tujuan. Intensitas cahaya dinyatakan sebagai:

$$I(x) = f(x) \tag{7}$$

dimana $I(x)$ adalah kecerahan cahaya di Lokasi x , dan $f(x)$ merupakan nilai fungsi objektif.

Tingkat ketertarikan antar *firefly* dipengaruhi oleh jarak dan dinyatakan sebagai:

$$\beta(r) = \beta_0 e^{-\gamma r^2} \tag{8}$$

Keterangan :

- $\beta(r)$: tingkat ketertarikan
- β_0 : ketertarikan awal
- γ : koefisien penyerapan cahaya
- r : jarak antar *firefly*

Besaran jarak antara dua *firefly* ditentukan berdasarkan hasil perhitungan yang menggunakan:

$$r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{9}$$

Keterangan :

- r_{ij} : jarak antara *firefly* ke- i dan ke- j
 - x_i dan x_j : posisi setiap *firefly*
- Pergerakan *firefly* mengikuti aturan:

$$x_i = x_i + \beta_0 e^{-\gamma r_{ij}^2} (x_j - x_i) + \alpha(rand - 0.5) \tag{10}$$

dengan x_i menunjukkan posisi *firefly* ke- i , x_j adalah posisi *firefly* yang lebih terang, α merupakan parameter randomisasi, serta *rand* merujuk pada bilangan acak pada interval [0, 1].

Prosedur algoritma FA diawali dengan inisialisasi populasi *firefly* beserta seluruh parameter yang diperlukan, dilanjutkan dengan evaluasi nilai fungsi objektif sebagai representasi intensitas cahaya masing-masing individu. Selanjutnya, tingkat ketertarikan antarindividu ditetapkan berdasarkan jarak yang memisahkan mereka, diikuti dengan pembaruan posisi menuju solusi yang lebih unggul. Siklus ini terus berulang hingga kondisi terminasi terpenuhi, yakni tercapainya iterasi maksimum atau konvergensi. Tabel berikut merangkum parameter yang diterapkan dalam algoritma FA pada penelitian ini.

Tabel 3. Parameter FA

Parameter	Nilai
Jumlah <i>firefly</i>	15
Maksimum iterasi	100
Parameter randomisasi (α)	0,8
Ketertarikan awal (β_0)	1,0
Koefisien penyerapan cahaya (γ)	0,1
Batas minimum k	3
Batas maksimum k	10

2.7. Silhouette Coefficient (SC)

Evaluasi kualitas *cluster* pada penelitian ini menggunakan metode *Silhouette Coefficient* (SC), yang dimanfaatkan untuk mengukur sejauh mana kualitas kesesuaian hasil pengelompokan yang dihasilkan [14]. Rentang nilai SC berada di antara -1 hingga 1, di mana semakin tinggi nilainya mendekati angka 1 maka semakin baik kualitas pengelompokan yang terbentuk. Perhitungan *Silhouette Coefficient* dirumuskan sebagai berikut:

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)} \tag{11}$$

dengan a_i adalah rata-rata jarak objek ke- i terhadap seluruh objek pada *cluster* yang sama, sedangkan b_i adalah jarak rata-rata minimum objek ke- i terhadap *cluster* lain. Nilai SC global dijadikan tolok ukur dalam menetapkan kombinasi parameter terbaik yang menghasilkan pengelompokan paling optimal.

2.8. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) merupakan metode statistik multivariat yang mengubah sekumpulan variabel asli menjadi beberapa komponen utama untuk mengurangi dimensi data, sehingga struktur dan pola data dapat divisualisasikan secara lebih efektif data [15]. Metode ini mereduksi dimensionalitas data tanpa mengorbankan informasi yang esensial. Dalam penelitian ini, PCA digunakan untuk memvisualisasikan hasil pengelompokan K-Means, baik tanpa optimasi maupun dengan PSO dan FA, dengan mengurangi tujuh variabel menjadi dua komponen utama (Dim1 dan Dim2) agar pola pemisahan kelompok dapat dilihat secara visual.

2.9. Ketimpangan Pendapatan

Ketimpangan pendapatan merupakan perbedaan distribusi kesejahteraan antara kelompok berpenghasilan tinggi dan rendah dalam suatu perekonomian. Ketimpangan ini dapat memberikan dampak positif berupa dorongan persaingan antarwilayah, namun juga menimbulkan dampak negatif seperti menurunnya efisiensi ekonomi dan meningkatnya ketidakstabilan sosial. Berbagai faktor memengaruhi ketimpangan pendapatan, seperti Indeks Pembangunan Manusia (IPM), jumlah penduduk miskin, tingkat pengangguran terbuka, inflasi, Upah Minimum Provinsi (UMP), PDRB per kapita, dan Tingkat Partisipasi Angkatan Kerja (TPAK). IPM mencerminkan kualitas kehidupan masyarakat berdasarkan dimensi kesehatan, pendidikan, serta standar hidup layak. Jumlah penduduk miskin menggambarkan ketidakmampuan dalam memenuhi kebutuhan dasar. Tingkat pengangguran terbuka menunjukkan proporsi penduduk yang tidak bekerja, sedangkan inflasi berkaitan dengan kenaikan harga yang memengaruhi daya beli masyarakat. Selain itu, UMP berperan sebagai standar minimum upah pekerja, PDRB per kapita menggambarkan tingkat kemakmuran wilayah, dan TPAK menunjukkan tingkat partisipasi penduduk dalam kegiatan ekonomi. Variabel-variabel tersebut digunakan untuk menggambarkan karakteristik wilayah dalam analisis pengelompokan ketimpangan pendapatan.

3. Hasil dan Pembahasan

3.1. Uji Multikolinearitas

Untuk menjamin tidak terjadinya hubungan linear yang kuat antarvariabel prediktor, dilakukan pengujian multikolinearitas sebagai salah satu tahap awal dalam analisis data. Indikator yang digunakan adalah nilai VIF, di mana nilai kurang dari 10 mengindikasikan bahwa variabel tersebut bebas dari multikolinearitas. Tabel berikut menyajikan hasil pengujian secara lengkap:

Tabel 4. Hasil Uji Multikolinearitas

	X1	X2	X3	X4	X5	X6	X7
X1	-	1,011929	1,255480	1,471360	1,022551	1,212059	1,472159
X2	1,011929	-	1,031703	1,000273	1,399492	1,026742	1,007567
X3	1,255480	1,031703	-	1,034042	1,015019	1,055982	2,118866
X4	1,471360	1,000273	1,034042	-	1,080329	1,000304	1,394210
X5	1,022551	1,399492	1,015019	1,080329	-	1,412778	1,044943
X6	1,212059	1,026742	1,055982	1,000304	1,412778	-	1,174225
X7	1,472159	1,007567	2,118866	1,394210	1,044943	1,174225	-

Berdasarkan hasil uji multikolinearitas di atas, nilai VIF antar variabel adalah kurang dari 10, yang artinya tidak adanya multikolinearitas. Karena telah memenuhi asumsi dari analisis cluster, maka variabel-variabel tersebut dapat digunakan untuk proses analisis cluster pada penelitian ini.

3.2. Standarisasi Data

Perbedaan skala antarvariabel berpotensi memengaruhi hasil analisis yang didasarkan pada pengukuran jarak. Untuk menyamakan skala antarvariabel, dilakukan standarisasi data menggunakan metode Z-Score. Langkah ini bertujuan agar tidak ada variabel yang mendominasi proses pengelompokan. Hasil standarisasi disajikan pada tabel berikut.

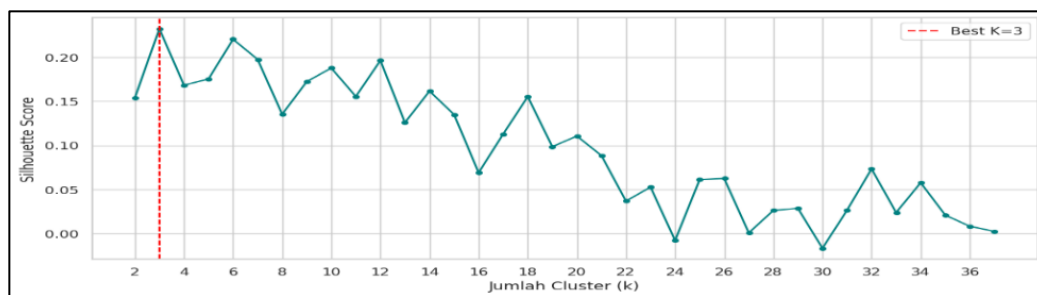
Tabel 5. Hasil Standarisasi Data

No	X1	X2	X3	X4	X5	X6	X7
1	0,32	0,09	0,98	0,66	0,55	-0,66	-1,32
2	0,32	0,51	0,87	0,60	-0,48	-0,18	0,15
3	0,41	-0,34	0,98	-0,78	-0,48	-0,44	-0,10
4	0,47	-0,17	-0,49	-0,37	0,29	1,32	-1,03
5	0,20	-0,39	0,07	-0,17	-0,12	0,04	-0,43
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
35	0,29	2,97	0,29	0,10	-1,70	-0,59	0,71
36	-0,88	-0,57	-0,24	0,22	1,44	-0,37	0,22
37	-2,49	-0,37	-1,17	1,90	1,44	0,56	0,57
38	-3,73	-0,32	-2,19	4,24	1,44	-1,08	4,11

3.3. Pembentukan Cluster dengan K-Means

1. Menentukan Jumlah Cluster (k)

Penentuan jumlah kelompok optimal pada K-Means dilakukan menggunakan metode Silhouette Coefficient (SC) dengan menguji nilai k dari 2 hingga 37. Hasil pengujian yang tersaji pada Gambar 2 menampilkan nilai SC tertinggi diperoleh pada k = 3 sebesar 0,23, sehingga jumlah cluster tersebut dianggap paling optimal.



Gambar 2. Grafik Silhouette Coefficient (SC)

2. Hasil Clustering K-Means

Berdasarkan nilai k optimal (k = 3) dengan SC = 0,23, penerapan K-Means pada data terstandarisasi menghasilkan pengelompokan yang jelas. Setiap objek tergabung dalam cluster dengan centroid terdekat, sehingga terbentuk kelompok yang homogen dan berbeda antar cluster. Tabel berikut menampilkan hasil pengelompokan provinsi berdasarkan faktor ketimpangan pendapatan:

Tabel 6. Hasil Cluster K-Means

Cluster	Provinsi
1	DKI Jakarta, Riau, Aceh, Kepulauan Riau, Sulawesi Utara, Kalimantan Utara, Papua Barat Daya, Banten, Maluku, Kalimantan Timur, Kepulauan Bangka Belitung, Sulawesi Selatan, dan Papua.

Cluster	Provinsi
2	Papua Tengah dan Papua Pegunungan Bali, Bengkulu, DI Yogyakarta, Gorontalo, Jambi, Jawa Barat, Jawa Tengah, Jawa Timur, Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Lampung, Maluku Utara, Nusa Tenggara Barat, Nusa Tenggara Timur, Papua Barat, Papua Selatan, Sulawesi Barat, Sulawesi Tengah, Sulawesi Tenggara, Sumatera Barat, Sumatera Selatan, Sumatera Utara

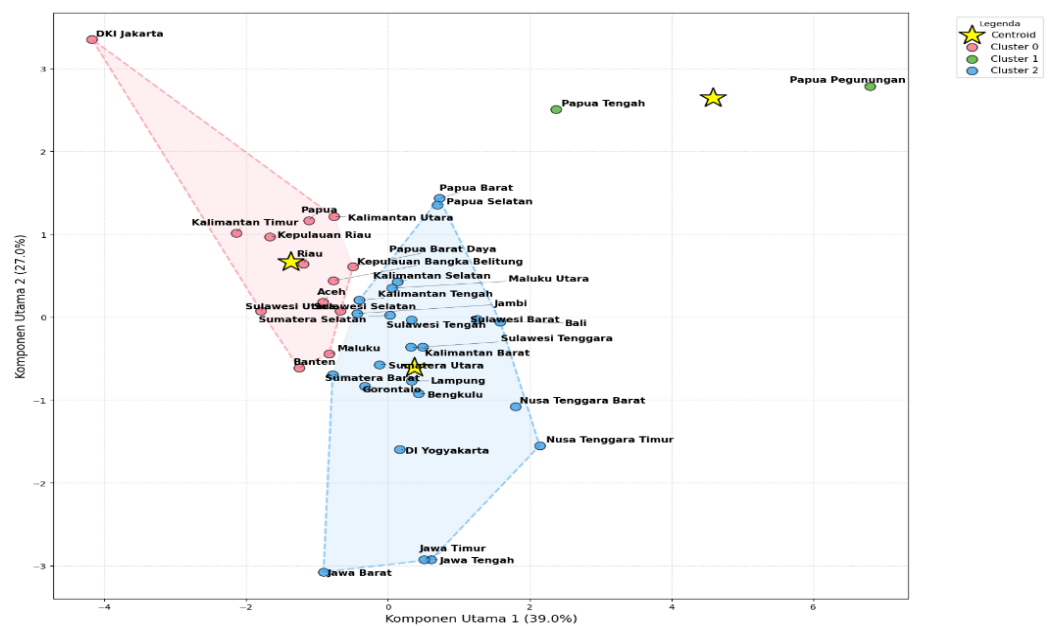
Berdasarkan **Tabel 6**, menunjukkan hasil pengelompokan provinsi ke dalam tiga cluster dengan jumlah anggota yang berbeda pada setiap kelompok. Perbedaan ini mengindikasikan adanya variasi karakteristik antar cluster, sehingga diperlukan analisis lebih lanjut untuk mengidentifikasi ciri masing-masing cluster. Profil karakteristik setiap cluster diungkap melalui perhitungan nilai rerata tiap variabel pada masing-masing kelompok, yang disajikan pada tabel berikut:

Tabel 7. Nilai Rata-rata Variabel pada Setiap Cluster

Cluster	X1	X2	X3	X4	X5	X6	X7
1	0,466	-0,323	0,803	-0,138	0,631	0,608	-0,813
2	-3,110	-0,347	-1,680	3,068	1,438	-0,259	2,338
3	0,007	0,213	-0,308	-0,189	0,482	-0,321	0,256

Berdasarkan **Tabel 7**, cluster 1 memiliki nilai relatif tinggi pada X1 (IPM), X3 (TPT), X5 (UMP), dan X6 (PDRB per kapita), yang menunjukkan pembangunan manusia, tingkat upah, dan aktivitas ekonomi yang lebih baik. Cluster 2 memiliki nilai tertinggi pada X4 (inflasi), X5 (UMP), dan X7 (TPAK), yang mencerminkan tingginya inflasi, upah, dan partisipasi angkatan kerja. Sementara itu, cluster 3 memiliki nilai relatif tinggi pada X2 (jumlah penduduk miskin), X7 (TPAK), dan X1 (IPM), yang menunjukkan tingkat kemiskinan dan partisipasi angkatan kerja yang lebih tinggi dibanding cluster lainnya.

3. Visualisasi Hasil Clustering



Gambar 3. Visualisasi Cluster K-Means

Berdasarkan gambar diatas ditampilkan visualisasi hasil klasterisasi menggunakan PCA mereduksi tujuh variabel menjadi dua komponen utama. Komponen Utama 1 menjelaskan 39,0% variasi data dan Komponen Utama 2 sebesar 27,0%, sehingga total variasi yang menjelaskan mencapai 66,0%. Hasilnya menunjukkan tiga cluster yang terpisah jelas:

Cluster 2 (hijau) sebagai kelompok terbesar mencakup sebagian besar provinsi di Indonesia, Cluster 1 (merah muda) didominasi provinsi di Jawa, Bali, dan Nusa Tenggara, sedangkan Cluster 3 (biru) hanya terdiri dari Papua Tengah dan Papua Pegunungan yang memiliki karakteristik paling berbeda. Temuan ini menunjukkan adanya pengelompokan wilayah yang jelas berdasarkan indikator penelitian.

3.4. Pembentukan Cluster K-Means dengan Optimasi Jumlah Cluster (k) menggunakan PSO

Pada tahap ini, PSO mencari jumlah k paling optimal pada K-Means berdasarkan nilai Silhouette Coefficient (SC). Parameter yang digunakan meliputi 15 partikel dan 100 iterasi. Setiap partikel merepresentasikan kandidat k yang diuji menggunakan K-Means dan dievaluasi dengan SC. Partikel menyimpan hasil terbaik individu (pbest) dan hasil terbaik global (gbest) sebagai acuan dalam memperbarui posisi. Proses ini dilakukan secara iteratif hingga konvergen atau mencapai batas iterasi, sehingga diperoleh nilai k dengan kualitas clustering terbaik.

1. Iterasi Algoritma PSO

Algoritma PSO melakukan proses iterasi untuk memperbarui kecepatan dan posisi setiap partikel berdasarkan nilai pbest dan gbest. Setiap partikel kemudian dievaluasi menggunakan nilai SC hingga diperoleh jumlah k yang optimal. Berikut merupakan hasil iterasi algoritma PSO.

Tabel 8. Hasil Iterasi Algoritma PSO

Iterasi	P1	P2	...	P15	Best k
1	k=6; v=1,54; SC=0,27	k=5; v=-0,15 SC=0,282	...	k=4; v=0,343 SC=0,243	k=5 SC=0,282
2	k=6; v=0,975 SC=0,268	k=4; v=-0,135 SC=0,243	...	k=4; v=0,954 SC=0,243	k=5 SC=0,282
3	k=6; v=0,496 SC=0,268	k=5; v=1,393 SC=0,282	...	k=4; v=0,888 SC=0,243	k=5 SC=0,282
⋮	⋮	⋮	⋮	⋮	⋮
99	k=6; v=0,275 SC=0,268	k=5; v=0,133 SC=0,282	...	k=3; v=-0,973 SC=0,232	k=5 SC=0,282
100	k=4; v=-1,036 SC=0,243	k=5; v=0,119 SC=0,282	...	k=5; v=2,795 SC=0,282	k=5 SC=0,282

Berdasarkan Tabel 8, setiap partikel mengalami pembaruan kecepatan (v) dan posisi (p) pada setiap iterasi berdasarkan nilai pbest dan gbest. Nilai v (velocity/kecepatan) menunjukkan besarnya perubahan posisi partikel dalam proses pencarian solusi optimal. Hasil iterasi menunjukkan bahwa best k secara konsisten berada pada k = 5 dengan nilai SC sebesar 0,282, sehingga PSO berhasil konvergen dan menghasilkan jumlah cluster optimal sebanyak 5 cluster.

2. Clustering K-Means dengan Jumlah Cluster (k) Optimal Hasil PSO

Jumlah k yang telah dioptimalkan menggunakan PSO selanjutnya diterapkan pada metode K-Means untuk mengelompokkan data yang telah distandarisasi berdasarkan jaraknya terhadap pusat klaster. Hasil pengelompokan tersebut ditampilkan pada tabel berikut ini:

Tabel 9. Hasil Cluster menggunakan K-Means dengan Optimasi PSO

Cluster	Provinsi
1	Nusa Tenggara Barat, Bengkulu, Bali, Sulawesi Tengah, Lampung, DI Yogyakarta, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, dan Nusa Tenggara Timur.
2	DKI Jakarta, Kalimantan Utara, Riau, Kepulauan Riau, dan Kalimantan Timur
3	Papua Tengah, Papua Pegunungan

Cluster	Provinsi
4	Sumatera Barat, Kalimantan Selatan, Banten, Maluku Utara, Aceh, Papua Barat Daya, Jambi, Sulawesi Utara, Kalimantan Tengah, Papua Selatan, Sumatera Utara, Kepulauan Bangka Belitung, Sulawesi Selatan, Kalimantan Barat, Sumatera Selatan, Maluku, Papua Barat.
5	Jawa Barat, Jawa Timur, Jawa Tengah

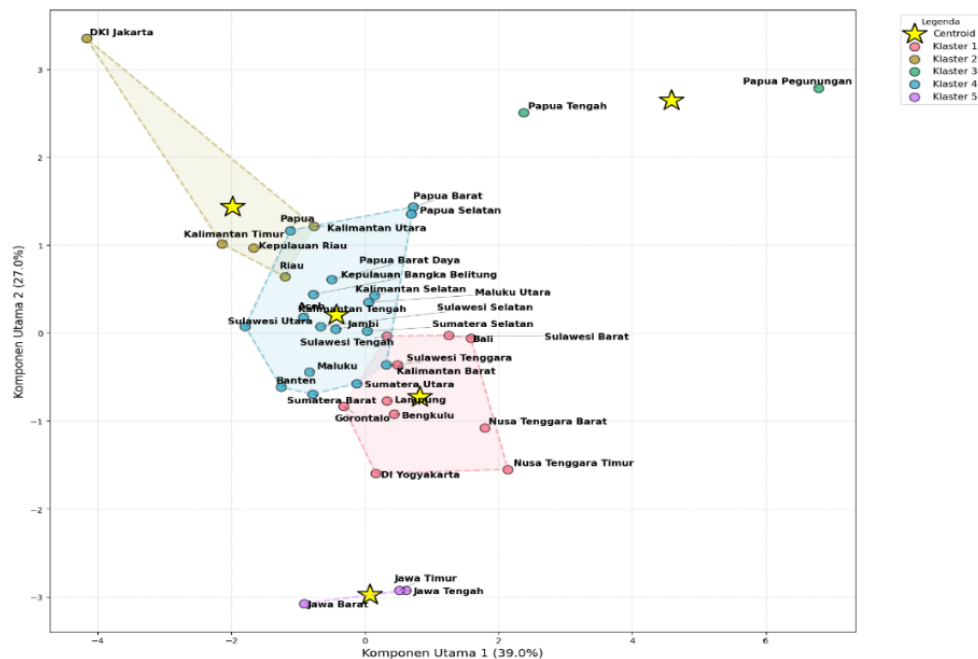
Tabel 9. Menunjukkan klasterisasi provinsi di Indonesia ke dalam 5 *cluster* dengan jumlah anggota yang berbeda pada setiap kelompok. Perbedaan ini menunjukkan adanya variasi karakteristik antar *cluster*, sehingga diperlukan analisis lebih lanjut untuk mengidentifikasi karakteristik setiap *cluster*. Karakteristik tiap klaster dijelaskan berdasarkan nilai rerata masing-masing variabel pada setiap kelompok, sebagaimana ditampilkan pada tabel berikut:

Tabel 10. Nilai Rata-rata Variabel pada Setiap *Cluster*

Cluster	X1	X2	X3	X4	X5	X6	X7
1	0,067	-0,190	-0,977	-0,479	-0,754	-0,475	0,630
2	1,013	-0,400	0,493	-0,073	0,922	2,151	-0,732
3	-3,110	-0,347	-1,680	3,068	1,438	-0,259	2,338
4	-0,031	-0,285	0,490	-0,059	0,274	-0,238	-0,442
5	0,343	3,241	0,616	0,029	-1,622	-0,398	0,215

Berdasarkan **Tabel 10**, Cluster 1 mencakup 10 provinsi di Sulawesi, Nusa Tenggara, dan sekitarnya dengan TPAK tinggi namun UMP dan PDRB per kapita rendah, mengindikasikan partisipasi kerja yang baik namun pendapatan belum optimal sehingga berpotensi menciptakan ketimpangan pendapatan. Cluster 2 meliputi DKI Jakarta, Kalimantan Timur, Riau, Kepulauan Riau, dan Kalimantan Utara dengan PDRB per kapita dan UMP tertinggi, namun tingkat pengangguran yang relatif tinggi berpotensi memperlebar ketimpangan pendapatan antarkelompok masyarakat. Cluster 3 terdiri dari Papua Tengah dan Papua Pegunungan dengan inflasi sangat tinggi dan IPM sangat rendah, mencerminkan tekanan daya beli sekaligus ketertinggalan pembangunan yang memperparah ketimpangan pendapatan. Cluster 4 mencakup 17 provinsi di Sumatera, Kalimantan, Sulawesi, Maluku, dan Papua dengan kemiskinan dan inflasi terkendali, namun pengangguran yang tinggi dan PDRB per kapita rendah mengindikasikan ketimpangan pendapatan yang masih perlu diperhatikan. Cluster 5 terdiri dari Jawa Barat, Jawa Tengah, dan Jawa Timur dengan jumlah penduduk miskin absolut tertinggi dan pengangguran cukup tinggi, mengindikasikan ketimpangan pendapatan yang besar akibat distribusi hasil pertumbuhan ekonomi yang belum merata.

3. Visualisasi Hasil Clustering



Gambar 4. Visualisasi Hasil Clustering

Gambar 4. menampilkan hasil visualisasi klusterisasi menggunakan PCA menunjukkan dua komponen utama, yaitu Komponen Utama 1 (39,0%) dan Komponen Utama 2 (27,0%). Hasilnya membentuk lima cluster yang terpisah cukup jelas. Cluster 1 mencakup Lampung, Bengkulu, DI Yogyakarta, Bali, NTB, dan NTT; Cluster 2 didominasi Kalimantan Timur, Kepulauan Riau, serta DKI Jakarta sebagai outlier; Cluster 3 terdiri dari Papua Tengah dan Papua Pegunungan; Cluster 4 merupakan kelompok terbesar yang mencakup berbagai wilayah di Indonesia; sedangkan Cluster 5 terdiri atas Jawa Barat, Jawa Tengah, dan Jawa Timur. Secara umum, cluster 2, 3, dan 5 tampak lebih terpisah, sementara cluster 1 dan 4 cenderung terkonsentrasi di pusat sebaran.

3.5. Pembentukan Cluster K-Means dengan Optimasi Jumlah Cluster (k) menggunakan FA

Pada tahap ini, Firefly Algorithm (FA) digunakan untuk mengoptimasi jumlah cluster (k) pada K-Means berdasarkan nilai Silhouette Coefficient (SC). Parameter yang digunakan meliputi 15 firefly dan 100 iterasi, serta parameter α , β , dan γ . Setiap firefly merepresentasikan kandidat k yang dievaluasi menggunakan K-Means dan SC sebagai nilai fitness. Firefly dengan fitness lebih rendah bergerak menuju yang lebih tinggi disertai komponen acak untuk eksplorasi. Proses iteratif ini menghasilkan nilai k optimal dengan kualitas clustering terbaik.

1. Iterasi Firefly Algorithm

Tabel 11. Hasil Iterasi Firefly ALgorithm

Iterasi	Nilai SC (k)				SC (Best k)
	F1	F2	...	F15	
1	0,171 (3)	0,232 (3)	...	0,232 (3)	0,359 (3)
2	0,232 (3)	0,398 (3)	...	0,232 (3)	0,398 (3)
3	0,388 (3)	0,243 (4)	...	0,398 (3)	0,398 (3)
4	0,232 (3)	0,322 (4)	...	0,232 (3)	0,398 (3)
5	0,398 (3)	0,398 (3)	...	0,216 (4)	0,398 (3)
⋮	⋮	⋮	⋮	⋮	⋮
99	0,232 (3)	0,218 (3)	...	0,206 (3)	0,398 (3)
100	0,359 (3)	0,388 (4)	...	0,218 (3)	0,398 (3)

Berdasarkan **Tabel 11**, nilai fitness (SC) mengalami perubahan pada setiap iterasi Firefly Algorithm dengan jumlah cluster yang didominasi oleh $k = 3$ dan sebagian kecil $k = 4$. Nilai SC tertinggi mencapai sekitar 0,398 dan konsisten menjadi fitness terbaik pada sebagian besar iterasi. Hasil ini menunjukkan bahwa algoritma telah konvergen dan berhasil mempertahankan solusi terbaik. Secara keseluruhan, $k = 3$ merupakan jumlah cluster yang paling dominan serta menghasilkan nilai fitness terbaik, sehingga dipilih sebagai jumlah cluster optimal.

2. Clustering K-Means dengan Jumlah Cluster (k) Optimal Hasil FA

Tahap selanjutnya adalah dilakukan pengelompokan menggunakan metode K-Means dengan jumlah k optimal dari hasil optimasi FA. Pengelompokan dilakukan terhadap data yang telah distandarisasi berdasarkan jarak setiap objek ke pusat cluster sehingga terbentuk klaster yang homogen secara internal dan heterogen antar cluster. Berikut hasil pengelompokan disajikan pada tabel:

Tabel 12. Hasil Cluster menggunakan K-Means dengan Optimasi FA

Cluster	Provinsi
1	Jawa Timur, Bali, Nusa Tenggara Timur, Bengkulu, Jawa Barat, DI Yogyakarta, Nusa Tenggara Barat, dan Jawa Tengah.
2	Riau, Kalimantan Timur, Papua Barat Daya, Sumatera Selatan, Aceh, Maluku Utara, Sulawesi Tengah, Lampung, DKI Jakarta, Kalimantan Selatan, Papua, Kepulauan Riau, Gorontalo, Sumatera Barat, Kalimantan Barat, Sulawesi Utara, Papua Barat, Banten, Kalimantan Utara, Maluku, Sulawesi Selatan, Papua Selatan, Jambi, Sulawesi Tenggara, Kepulauan Bangka Belitung, Kalimantan Tengah, Sulawesi Barat, Sumatera Utara, dan Papua Tengah.
3	Papua Pegunungan

Dari **Tabel 12**, dapat kita lihat klasterisasi 38 provinsi menggunakan optimasi (FA) pada $k = 3$ menghasilkan tiga klaster. Klaster 2 merupakan kelompok terbesar yang mencakup sebagian besar provinsi di Sumatera, Kalimantan, dan Indonesia Timur. Klaster 1 terdiri atas delapan provinsi di Jawa, Bali, dan Nusa Tenggara, sedangkan Klaster 3 hanya beranggotakan Papua Pegunungan. Nilai rata-rata setiap variabel pada masing-masing klaster disajikan pada tabel berikut.

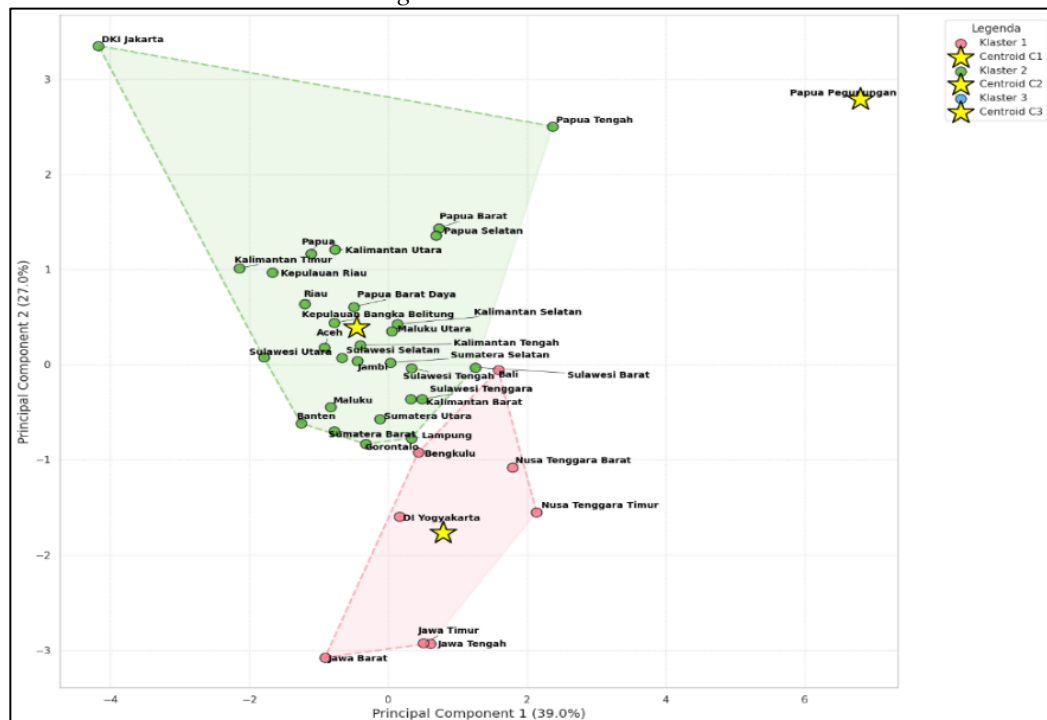
Tabel 13. Nilai Rata-rata Variabel pada Setiap Cluster

Cluster	X1	X2	X3	X4	X5	X6	X7
1	0,352	1,144	-0,465	-0,126	-1,296	-0,551	0,807
2	0,032	-0,304	0,204	-0,111	0,308	0,189	-0,364
3	-3,732	-0,324	-2,193	4,241	1,438	-1,079	4,111

Dari **Tabel 13**, terlihat Cluster 1 yang didominasi provinsi di Pulau Jawa, Bali, dan Nusa Tenggara menunjukkan IPM dan TPAK yang relatif tinggi, namun disertai UMP dan PDRB per kapita yang rendah. Kondisi ini mengindikasikan bahwa meskipun partisipasi angkatan kerja cukup baik, rendahnya upah dan aktivitas ekonomi berpotensi menciptakan ketimpangan pendapatan antara kelompok pekerja berpendapatan tinggi dan rendah. Cluster 2 yang mencakup sebagian besar provinsi di Sumatera, Kalimantan, Sulawesi, Maluku, dan Papua memiliki karakteristik mendekati rata-rata nasional, namun heterogenitas kondisi sosial ekonomi dalam kelompok ini mengindikasikan adanya variasi tingkat ketimpangan pendapatan yang cukup beragam antarprovinsi dalam cluster yang sama. Cluster 3 yang hanya terdiri dari Papua Pegunungan memiliki karakteristik paling kontras, yaitu inflasi dan TPAK sangat tinggi namun IPM sangat rendah. Tingginya inflasi yang tidak diimbangi kualitas pembangunan manusia yang memadai menc-

erminkan kondisi ketimpangan pendapatan yang paling parah dibandingkan cluster lainnya, karena kenaikan harga barang dan jasa justru menekan daya beli masyarakat yang sebagian besar masih berada dalam keterbatasan akses terhadap layanan dasar.

3. Visualisasi Hasil Clustering



Gambar 5. Visualisasi Hasil Clustering

Gambar diatas menampilkan visualisasi klasterisasi menggunakan PCA menunjukkan bahwa Dim1 (Principal Component 1) menjelaskan 39,0% dan Dim2 (Principal Component 2) sebesar 27,0% variasi data. Kedua komponen ini secara kumulatif merepresentasikan 66,0% informasi dari data asli. Hasilnya membentuk tiga cluster dengan pola sebaran yang berbeda. Cluster 2 merupakan kelompok terbesar dengan sebaran yang luas. Cluster 1 berkumpul di bagian bawah grafik dan didominasi wilayah Jawa, Bali, serta Nusa Tenggara. Sementara cluster 3 terpisah dari yang lain dan hanya berisi Papua Pegunungan. Secara keseluruhan, pemisahan antar kelompok cukup jelas.

3.6. Evaluasi dan Perbandingan Hasil Clustering K-Means dengan Optimasi Jumlah cluster (k) menggunakan PSO dan FA

Tabel 14. Nilai Silhouette Coefficient Berdasarkan Pendekatan K-Means

Pendekatan K-Means	Silhouette Coefficient (SC)
K-Means	0,23
K-Means Optimasi PSO	0,28
K-Means Optimasi FA	0,39

Berdasarkan Tabel 14, nilai Silhouette Coefficient (SC) menunjukkan adanya peningkatan kinerja pada setiap metode. K-Means tanpa optimasi menghasilkan nilai SC sebesar 0,23 yang menunjukkan kualitas cluster rendah. Setelah dilakukan optimasi menggunakan PSO, nilai SC meningkat menjadi 0,28, dan juga dilakukan optimasi menggunakan FA, nilai SC kembali meningkat menjadi 0,39. Hasil ini menunjukkan bahwa penerapan algoritma optimasi mampu meningkatkan kualitas clustering, dengan FA memberikan performa terbaik dalam menghasilkan pemisahan cluster yang lebih optimal.

4. Kesimpulan

Hasil penelitian membuktikan bahwa pengintegrasian algoritma Particle Swarm Optimization (PSO) maupun Firefly Algorithm (FA) ke dalam metode K-Means secara signifikan mampu meningkatkan kualitas pengelompokan dibandingkan K-Means yang dijalankan tanpa optimasi. PSO menghasilkan konfigurasi optimal dengan 4 kelompok dan nilai Silhouette Coefficient sebesar 0,40, sedangkan FA mencapai hasil yang lebih baik dengan 3 kelompok optimal dan nilai Silhouette Coefficient sebesar 0,39. Temuan ini menegaskan bahwa penerapan metode optimasi mampu meningkatkan mutu klusterisasi secara keseluruhan, dengan FA unggul dalam menghasilkan jumlah cluster yang lebih proporsional dan pemisahan antarkelompok yang lebih tegas. Di sisi lain, hasil pengelompokan mengungkap adanya perbedaan karakteristik yang nyata antara satu kelompok dengan kelompok lainnya, yang mencerminkan keragaman pola ketimpangan pendapatan di berbagai provinsi Indonesia. Dengan demikian, penerapan metode optimasi—terutama FA—berpotensi menghasilkan pengelompokan yang lebih akurat dan representatif. Penelitian selanjutnya disarankan untuk mengeksplorasi kombinasi algoritma optimasi lain, memperluas cakupan indikator yang digunakan, atau mengaplikasikan pendekatan serupa pada wilayah dan kasus yang berbeda sebagai upaya untuk menguji konsistensi serta meningkatkan performa hasil klusterisasi.

Referensi

- [1] Wahyuni, S. dan Andriyani, D. 2022. Pengaruh inflasi, jumlah penduduk dan pertumbuhan ekonomi terhadap ketimpangan pendapatan di provinsi aceh. *Jurnal Ekonomi Regional Unimal*, 05(1). <https://doi.org/10.29103/jeru.v5i1.7919>
- [2] Farhan, M. dan Sugianto. 2022. Analisis faktor-faktor yang mempengaruhi Tingkat ketimpangan pendapatan di pulau jawa. *Jurnal Ilmiah Bidang Sosial, Ekonomi, Budaya, Teknologi, dan Pendidikan*, 1(4). <https://publish.ojs-indonesia.com/index.php/SIBATIK/article/view/29>
- [3] Maharani, A., Lica, Raihan, A. F., Hidayatullah, M. R., dan Putra, A. R. 2025. Analisis kesejahteraan dan ketimpangan pendapatan (gini ratio) terhadap tingkat pengangguran di kepulauan bangka belitung. *Jurnal Mirai Management*, 10(1):514–526. <https://doi.org/10.37531/mirai.v10i1.9258>
- [4] Badan Pusat Statistik (BPS), 2024
- [5] C. Frahita dan S. F. Nurhayati, Analisis faktor-faktor yang mempengaruhi ketimpangan pendapatan di provinsi daerah istimewa yogyakarta, tahun 2017-2023. *JIMEA | Jurnal Ilmiah MEA (Manajemen, Ekonomi, dan Akuntansi)*, vol. 9, no. 1, 1847, 2025.
- [6] G. R. Wahyudi, R. Rahmaddeni, E. Dini, S. Adrianto, dan R. Fadila, *Clustering* regencies in indonesia for regional mapping using the k-means algorithm pengelompokan kabupaten di indonesia untuk pemetaan daerah menggunakan algoritma k-means. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 3, 2025. <https://doi.org/10.57152/malcom.v5i3.2206>
- [7] N. Nugroho dan F. Adhinata, Penggunaan metode k-means dan k-means++ sebagai *clustering* data covid-19 di pulau jawa. *Teknika*, vol. 11, pp. 170–179, 2022. <https://doi.org/10.34148/teknika.v11i3.502>
- [8] K. Yuliawan, Metode k-means *clustering* untuk mengelompokan indeks pemabangunan manusia (ipm) provinsi papua. *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 2, 2025. <https://doi.org/10.36040/jati.v9i2.13243>
- [9] H. Harliana, R. M. H. Bhakti, O. S. Bachri, dan F. S. Efendi, Optimasi k-means dengan *Particle Swarm Optimization* pada pengelompokan daerah stunting. *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, vol. 3, no. 2, pp. 95–101, 2021. <https://doi.org/10.46772/intech.v3i02.457>
- [10] D. Warianta, P. Astagina, R. Julianto, dan F. Arini, Optimalisasi k-means menggunakan algoritma firefly untuk segmentasi pelanggan pada e-commerce. *JURNAL FASILKOM*, vol. 14, pp. 775–785, 2024. <https://doi.org/10.37859/jf.v14i3.8287>
- [11] D. Dahnia, Penerapan metode *clustering* k-means terhadap dosen berdasarkan publikasi jurnal nasional dan terakreditasi sinta. *JSI: Jurnal Sistem Informasi (E-Journal)*, vol. 16, pp. 219–236, 2024. <https://doi.org/10.18495/jsi.v16i1.163>
- [12] R. Awaliah, Analisis *clustering* untuk mengelompokkan tingkat kesejahteraan kabupaten/kota berdasarkan sosial ekonomi rumah tangga di wilayah provinsi sulawesi selatan. Master's thesis, Universitas Islam Negeri (UIN) Alauddin Makassar, 2018.
- [13] A. Waruwu, M. Yetri, dan F. Setiawan, Implementasi data mining dalam mengelompokkan data penduduk kurang mampu menggunakan metode k-means *clustering*. *JURNAL SISTEM INFORMASI TGD*, vol. 2, no. 6, pp. 945–955, 2023. <https://doi.org/10.53513/jursi.v2i6.8965>

-
- [14] B. A. Suryani, M. N. Hayati, dan 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*, 6(1), 2023. <https://doi.org/10.36087/jrp.v6i1.139>
- [15] A. Sidqiy dan, S. Amar , Faktor – faktor yang mempengaruhi pemerataan pendapatan di indonesia. *Ecosains: Jurnal Ilmiah Ekonomi dan Pembangunan*, vol. 9, no. 1, pp. 48-57, 2020. <https://doi.org/10.24036/ecosains.11549957.00>