INTI NUSA MANDIRI
DOI: https://doi.org/10.33480/inti.v20i2.7814

VOL. 20. NO. 2 FEBRUARI 2026
P-ISSN: 0216-6933 | E-ISSN: 2685-807X
Diterbitkan Oleh: LPPM Universitas Nusa Mandiri

CLUSTERING WILAYAH KEMISKINAN MULTIDIMENSI DI INDONESIA MENGGUNAKAN ALGORITMA FUZZY C-MEANS DAN OPTICS

Nicholas Eugene Supardi[1]; Teny Handhayani [2*]; Irvan Lewenusa[3]

Fakultas Teknologi Informasi[1,2,3]
Universitas Tarumanagara, Indonesia [1,2,3]
https://untar.ac.id/[1,2,3]
nicholas.535220102@stu.untar.ac.id[1], tenyh@fti.untar.ac.id[2*], irvanl@fti.untar.ac.id[3]

(*) Corresponding Author

Abstract—*A multidimensional approach to mapping poverty in Indonesia necessitates the use of the Human Development Index (HDI), Poverty Line, and Expenditure per Capita. This study aims to evaluate the performance of two clustering algorithms with distinct paradigms the centroid-based Fuzzy C-Means (FCM) and the density-based OPTICS in profiling poverty across 501 regencies and cities. Experimental results indicate that OPTICS achieved high internal validity, with a Silhouette Score of 0.7301 and a Davies-Bouldin Index (DBI) of 0.329. Significantly, OPTICS identified 94% of the data as noise. This finding reveals a fundamental characteristic: the distribution of socio-economic data in Indonesia is highly heterogeneous and sparse, lacking inherently dense cluster structures. Conversely, FCM, employing a soft clustering approach, successfully accommodates the ambiguity of data boundaries and provides comprehensive segmentation across all regions. Despite yielding lower validity metrics (Silhouette Score 0.3894), FCM was selected as the final model because it satisfies the practical requirements of the application, which demands complete coverage mapping. This study concludes that a soft clustering approach is more applicable than density-based clustering for analyzing highly heterogeneous data such as that found in Indonesia.*

Keywords*: Application, Clustering, Fuzzy C-Means, OPTICS, Poverty*

Abstrak—Sebuah pendekatan multidimensi untuk memvisualisasikan kemiskinan di Indonesia memerlukan penggunaan Indeks Pembangunan Manusia (IPM), Garis Kemiskinan, dan Pengeluaran per Kapita. Tujuan dari penelitian ini adalah untuk mengevaluasi kinerja dua algoritma clustering dengan berbagai paradigma, yaitu *Fuzzy C-Means* (FCM) yang berbasis pusat dan OPTICS yang berbasis kepadatan, dalam menggambar profil kemiskinan di 501 kabupaten/kota. Validitas internal OPTICS sangat tinggi (*Silhouette Score* 0.73, DBI 0.32), menurut hasil eksperimen. Secara signifikan, OPTICS menghasilkan nilai validitas internal yang tinggi, namun mengklasifikasikan sekitar 94% wilayah kabupaten/kota sebagai *noise*, yaitu wilayah yang tidak membentuk *cluster* padat berdasarkan kriteria kepadatan. Temuan Ini menunjukkan bahwa distribusi data sosio-ekonomi di Indonesia sangat heterogen dan tersebar, sehingga pendekatan clustering berbasis kepadatan hanya mampu mengelompokan sebagian kecil wilayah dengan karakteristik ekstrem. Sebaliknya, FCM dengan pendekatan *soft clustering* dapat menangani ketidaktegasan batas antar-data dan melakukan segmentasi seluas seluruh wilayah. Meskipun memiliki skor validitas yang lebih rendah (*Silhouette Score* 0.38 dan DBI 1.01), FCM dipilih sebagai model terakhir karena memenuhi persyaratan dasar pembuatan kebijakan yang membutuhkan pemetaan menyeluruh. Studi ini menemukan bahwa pendekatan *soft clustering* lebih efektif daripada pendekatan *density-based clustering* untuk data dengan heterogenitas tinggi seperti di Indonesia.

Kata kunci: *Aplikasi, Klastering, Fuzzy C-Means, OPTICS, Kemiskinan*

PENDAHULUAN

Kesenjangan wilayah yang tajam merupakan salah satu tantangan utama dalam pembangunan di Indonesia. Kondisi sosial-ekonomi masyarakat sangat beragam karena wilayah Indonesia terdiri dari lebih dari 514 kabupaten/kota yang tersebar di ribuan pulau. Disparitas ini terjadi tidak hanya antarprovinsi, tetapi juga antarkabupaten dalam provinsi yang sama (Santoso & Anggraini, 2024) Dalam konteks agenda global, Tujuan Pembangunan Berkelanjutan (Sustainable Development Goals/SDGs) menuntut pemahaman yang komprehensif mengenai klasifikasi dan distribusi kemiskinan. Pencapaian SDG 1 (Tanpa Kemiskinan), khususnya target 1.2.2 yang bertujuan mengurangi setidaknya setengah proporsi laki-laki, perempuan, dan anak-anak dari segala usia yang hidup dalam kemiskinan multidimensi, sangat memerlukan pemetaan wilayah yang akurat dan berbasis data (Tawakal et al., 2025).

Clustering telah digunakan secara luas untuk mengumpulkan data kemiskinan, tetapi terdapat perbedaan besar dalam penelitian sebelumnya. (Faturahman & Hidayati, 2025) menggunakan *Fuzzy C-Means*, analisis mereka hanya mencakup kabupaten atau kota di Provinsi Jawa Tengah. Dalam hal ini, (Bahtiyar et al., 2024) membatasi analisis mereka pada lingkup Jawa Timur, dan (Oktaviani et al., 2024) berfokus pada Jawa Barat dengan menggunakan metode partisi keras *K-Means*. Di sisi lain, (Isah et al., 2024) melakukan penelitian dengan menggunakan metode *Fuzzy C-Means* dan *Fuzzy Probabilistic C-Means*, tetapi menggunakan unit analisis tingkat provinsi yang kurang mendetail untuk mengidentifikasi disparitas lokal dengan berbasis kepadatan.

Dalam literatur tersebut, keterbatasan wilayah studi berdampak langsung pada kualitas generalisasi hasil. Studi yang hanya mencangkup stau provinsi, pulau, atau kota besar cenderung menangkap karakteristik regional tertentu tanpa mempertimbangkan variasi sosial-ekonomi di skala nasional. Akibatnya, temuan tersebut menjadi kurang optimal ketika diterapkan pada tingkat kebijakan nasional karena kemungkinan bias data lokal dan kurangnya kemampuan model untuk memetakan pola yang lebih kompleks atau tersebar dalam data besar yang heterogen.

Penelitian ini menawarkan tiga elemen utama kebaruaan (*novelty*) untuk mengisi celah tersebut. Fokus utama penelitian ini adalah menganalisis bagaimana perbedaan paradigma clustering (density-based dan soft clustering) berprilaku ketika diterapkan pada data kemiskinan multidimensi yang berskala nasional dan sangat heterogen. Pertama, analisis telah diperluas ke tingkat nasional, mencakup 501 kabupaten/kota di seluruh Indonesia selama sepuluh tahun (2015–2024). Ini memberikan gambaran yang jauh lebih menyeluruh dibandingkan dengan studi regional sebelumnya. Kedua, Pendekatan komparatif antara algoritma *Fuzzy C-Means* (FCM) dan *OPTICS* kemudian digunakan. FCM dipilih karena kemampuan untuk menangani ketidakpastian data sosial melalui derajat keanggotaan (Darrell & Reswara, 2024), sedangkan OPTICS dipilih karena merupakan metode berbasis kepadatan yang memiliki kemampuan untuk mendeteksi klaster dan suara secara arbitrer (Nurhalizah et al., 2025). Ketiga, memasukkan hasil analisis ke dalam sistem aplikasi web interaktif yang membantu pengambilan keputusan. Tujuan utama penelitian ini adalah membuat profil kemiskinan yang menyeluruh, melakukan penilaian kinerja FCM dan OPTICS, dan menyediakan alat visualisasi data yang berguna.

Untuk mengevaluasi kualitas struktur pengelompokan yang dibuat tanpa label eksternal, pendekatan validasi internal digunakan. Dua metrik evaluasi utama dalam penelitian ini adalah *Silhouette Coefficient* dan *Davies-Bouldin Index* (DBI). *Silhouette Coefficient* mengukur kualitas kohesi dan separasi klaster dengan menilai seberapa mirip suatu objek dengan klasternya sendiri dibandingkan dengan klaster lain; nilai yang mendekati +1 menunjukkan penempatan objek yang sangat baik dalam klaster (Hasan, 2024). Sebaliknya, *Davies-Bouldin Index* (DBI) menilai rasio antara tingkat kepadatan di dalam klaster dan jarak antar klaster. Prinsipnya adalah bahwa nilai indeks yang lebih rendah menunjukkan skema pengelompokan yang lebih optimal karena klaster memiliki kepadatan tinggi dan sangat terpisah satu sama lain (Ying et al., 2024).

BAHAN DAN METODE

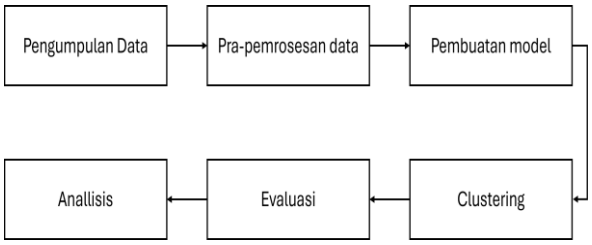
Gambar 1 menampilkan alur kerja penelitian. Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari situs resmi Badan Pusat Statistik (BPS) Republik Indonesia. Data diambil pada Triwulan II tahun 2025, dengan cakupan periode waktu 2015 - 2024. Tujuan dari pemilihan jangka waktu dan cakupan wilayah ini adalah untuk menghasilkan tren disparitas yang representatif dan menyeluruh secara nasional. Unit analisis dalam penelitian ini adalah 501 Kabupaten/Kota di Indonesia, sesuai dengan ketersediaan dan konsistensi data pada

INTI NUSA MANDIRI
DOI: https://doi.org/10.33480/inti.v20i2.7814

VOL. 20. NO. 2 FEBRUARI 2026
P-ISSN: 0216-6933 | E-ISSN: 2685-807X
Diterbitkan Oleh: LPPM Universitas Nusa Mandiri

tingkat administrasi tersebut. Terdapat tiga indikator utama dalam variabel penelitian yang menunjukkan berbagai aspek kemiskinan: Indeks Pembangunan Manusia (IPM) dalam skala 0-100 untuk mengukur kualitas hidup, Garis Kemiskinan (GK) dalam satuan rupiah per kapita per bulan sebagai ambang batas kebutuhan dasar, dan Pengeluaran per kapita dalam satuan ribu per tahun untuk menunjukkan daya beli riil masyarakat. Dengan demikian, total observasi awal yang dikumpulkan berjumlah 15.030 data (501 Kabupaten/kota x 10 tahun x 3 variabel).

Pada tahap awal, dataset masih mengandung beberapa data yang tidak digunakan, yaitu data agregat tingkat provinsi serta sejumlah kecil data yang memiliki nilai tidak lengkap (*missing values*). Data tingkat provinsi dikeluarkan karena penelitian ini secara spesifik berfokus pada analisis kabupaten/kota. Sementara itu, data dengan nilai yang tidak lengkap dihapus karena proporsinya sangat kecil dan tidak mempengaruhi representativitas data secara keseluruhan. Proses pengolahan data modeling dan evaluasi kualitas klaster dilakukan menggunakan bahasa Python dengan beberapa library seperti *scikit-learn* (modeling), *pandas* (data), *matplotlib dan Plotly* (grafik), lalu *folium* (peta).



Sumber: (Hasil Penelitian., 2025)
Gambar 1. Alur Penelitian

Sebelum digunakan untuk analisis, data mentah melalui beberapa tahapan pra-pemrosesan untuk memastikan validitas input. Tahap pertama adalah pembersihan data, yang menangani nilai yang hilang dan menghapus data gabungan tingkat provinsi, sehingga analisis terkonsentrasi pada tingkat kabupaten atau kota (Srirahayu & Pribadie, 2023). Selanjutnya, variabel Pengeluaran per Kapita diubah menjadi satuan. Ini berarti mengubah nilai tahunan menjadi nilai bulanan sehingga memiliki basis waktu yang sama dengan variabel Garis Kemiskinan. Penelitian ini menggunakan teknik Min-Max Scaler untuk normalisasi data karena adanya perbedaan skala yang sangat ekstrim antarvariabel. IPM berkisar dari puluhan hingga jutaan rupiah, sedangkan Garis Kemiskinan berkisar dari ratusan ribu hingga jutaan rupiah. Nilai data secara keseluruhan diubah menjadi rentang seragam mulai dari 0 hingga 1 (Shantal et al., 2023). Ini memastikan bahwa setiap variabel memberikan kontribusi bobot yang sama dalam perhitungan jarak algoritma dan mencegah bias yang dominan dari variabel dengan nilai nominal yang besar. Persamaan normalisasi berikut digunakan:

$$X_{Scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Untuk membandingkan kinerja, proses pengelompokan wilayah dilakukan menggunakan dua algoritma yang memiliki sifat yang berbeda. Pertama, algoritma soft clustering *Fuzzy C-Means* (FCM) memungkinkan setiap titik data memiliki tingkat keanggotaan parsial terhadap lebih dari satu klaster (Andrian et al., 2024). Metode ini berfungsi dengan meminimalkan fungsi objektif berikut:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} \mu_{ij}^{m} \left\| x_i - v_j \right\|^2 \quad (2)$$

Di mana N adalah jumlah data, C adalah jumlah klaster, μ_{ij}^{m} adalah derajat keanggotaan, dan $\left\| x_i - v_j \right\|$ adalah jarak geometris antara data ke-i dan pusat klaster ke-j. Untuk mengidentifikasi struktur klaster, algoritma kedua adalah OPTICS (*Ordering Points To Identify the Clustering Structure*), yang merupakan pengembangan dari DBSCAN. OPTICS menggunakan ide *Core Distance* dan *Reachability Distance* untuk mengidentifikasi struktur klaster. *Core Distance* adalah jarak minimum yang diperlukan untuk membentuk klaster padat, dan *Reachability Distance* adalah jarak yang diperlukan dari titik p ke titik o. Minpts adlaah jumlah minimum data agar terbentuk suatu klaster dan Xi adalah persentase penurunan kepadatan untuk dianggap sebagai klaster baru (Fan & Wang, 2024). *Distance* ini diukur dengan rumus:

$$Rd_{Xi,MinPts}(P,O) = \begin{cases} Undef, if |N_{Xi}(O)| < MinPts \\ \max\left(Cd_{Xi,MinPts}(O), d(O,P)\right), otherwise \end{cases} \quad (3)$$

Metode validasi internal tanpa label eksternal digunakan untuk mengevaluasi kualitas hasil clustering. *Silhouette Coefficien*t, metrik pertama yang digunakan untuk mengukur seberapa mirip suatu objek dengan klasternya sendiri dengan klaster lain, digunakan (Paembonan & Abduh, 2021). Nilai silhouette s(i) untuk setiap titik data didapat dengan persamaan berikut:

VOL. 20. NO. 2 FEBRUARI 2026
P-ISSN: 0216-6933 | E-ISSN: 2685-807X
Diterbitkan Oleh: LPPM Universitas Nusa Mandiri

INTI NUSA MANDIRI
DOI: https://doi.org/10.33480/inti.v20i2.7814

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

Dengan a(i) merupakan jarak rata-rata di dalam klaster dan b(i) merupakan jarak rata-rata terdekat ke klaster lain. *Davies-Bouldin* (DBI) menilai rasio antara tingkat kepadatan dalam klaster dan jarak pemisahan antar klaster, berfungsi sebagai metrik kedua (Hasan, 2024). Rumus berikut digunakan untuk menghitung indeks ini:

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{j \neq i}\left(\frac{S_i+S_j}{d_{ij}}\right) \quad (5)$$

Di mana si adalah sebaran dalam klaster dan dij adalah jarak antara pusat klaster. Langkat terakhir adalah implementasi model klaster kedalam sebuah aplikasi berbasis web. Model pengembangan *Waterfall* dalam *Software Development Life Cycle* (SDLC), digunakan untuk mengimplementasikan komputasi dan pengembangan sistem aplikasi web (Widianto, 2024). Bahasa pemrograman Python digunakan, dengan kerangka kerja *Django* untuk sisi *back-end* dan *Vue.js* untuk antarmuka pengguna, serta pustaka *Folium* dan *Plotly* untuk visualisasi data interaktif.

Perlu ditegaskan bahwa penelitian ini tidak bertujuan untuk mengukur tingkat kemiskinan pada tahun berjalan, melainkan untuk menganalisis pola dan struktur *cluster* kemiskinan multidimensi berdasarkan karakteristik data sosial – ekonomi antarwilayah. Oleh karena itu, penggunaan data periode 2015 – 2024 dimaksudkan untuk mengangkap stabilitas pola, variasi struktural, dan tingkat hetegonitas data, bukan untuk menghasilkan nilai kemiskinan terkini. Pendekatan ini memungkinkan evaluasi algoritma *clustering* pada data yang lebih representatif dan menghindari bias temporer akibat fluktuasi jangka pendek. Dengan demikian, pemilihan rentang waktu historis dapat memperkuat validitas analisis pola *cluster*, khususnya dalam konteks perbandingan algoritma FCM dan OPTICS.

HASIL DAN PEMBAHASAN

Pengujian data dilakukan kepada masing masing variabel dan data gabungan. Dengan algoritma *Fuzzy C-Means* dan OPTICS. Eksperimen pada algoritma *Fuzzy C-Means* menggunakan jumlah *cluster* 2 – 10, untuk OPTICS eksperimen menggunakan parameter minpts 5, 10, 15, 25, 33, dan 50 lalu parameter Xi 0.1 dan 0.05. Rentang jumlah cluster 2 – 10 dipilih berdasarkan pertimbangan metodologis dan praktis. Secara konseptual, jumlah cluster minimum 2 cluster diperlukan untuk merepresentasikan pemisahan dasar antarwilayah, sedangkan jumlah cluster maksimum 10 cluster digunakan untuk menghindari segmentasi yang terlalu granular dan sulit diinterpretasikan dalam konteks pemetaan kebijakan publik [congming shi]. Eksperimen ini bertujuan untuk mencari nilai S*ilhouette* tertinggi dan DBI terendah:

Tabel 1. Pengujian dataset IPM FCM

Algo	Cluster	SS	DBI
FCM	2	0.50	0.68
FCM	3	0.58	0.52
FCM	4	0.54	0.50
FCM	5	0.52	0.51
FCM	6	0.51	0.52
FCM	7	0.51	0.52
FCM	8	0.49	0.55
FCM	9	0.48	0.56
FCM	10	0.49	0.55

Sumber : (Hasil Penelitian, 2025)

Tabel 2. Pengujian dataset IPM OPTICS

Algo	minpts	Xi	SS	DBI	noise
OPTICS	5	0.1	0.66	0.44	376
OPTICS	5	0.05	0.48	0.61	243
OPTICS	10	0.05	0.78	0.25	400
OPTICS	10	0.1	0.78	0.25	377
OPTICS	15	0.05	-	-	-
OPTICS	15	0.1	0.70	0.39	369
OPTICS	25	0.05	-	-	-
OPTICS	25	0.05	0.68	0.41	420
OPTICS	33	0.1	-	-	-
OPTICS	33	0.05	-	-	-
OPTICS	50	0.1	-	-	-
OPTICS	50	0.05	-	-	-

Sumber : (Hasil Penelitian, 2025)

Pada penelitian ini algoritma *Fuzzy C-Means* (FCM) menunjukkan kinerja yang cukup stabil pada rentang jumlah cluster 2–10, seperti yang ditunjukkan oleh hasil eksperimen yang ditunjukkan pada Tabel 1. Untuk variabel IPM dengan metode FCM, jumlah cluster optimal adalah 3. Nilai *Silhouette Score* (SS) tertinggi tercatat pada *cluster* 3 sebesar 0.58 dengan nilai DBI sebesar 0.52. Meskipun nilai DBI sedikit lebih rendah pada *cluster* 5, penurunan nilai *Silhouette* menjadi 0.54 menunjukkan bahwa pemisahan antar cluster tidak sebaik pada *cluster* 3.

Sebaliknya, hasil eksperimen algoritma OPTICS di Tabel 2 menunjukkan variasi yang signifikan dalam hasil bergantung pada parameter MinPts dan Xi. Konfigurasi metrik terbaik dicapai pada MinPts 10 dan Xi 0.1, yang menghasilkan Skor *Silhouette* yang sangat tinggi sebesar 0.78 dan Skor DBI yang sangat rendah sebesar 0.25. Meskipun

VOL. 20. NO. 2 FEBRUARI 2026
P-ISSN: 0216-6933 | E-ISSN: 2685-807X
Diterbitkan Oleh: LPPM Universitas Nusa Mandiri

algoritma OPTICS menghasilkan nilai *Silhouette Score* yang sangat tinggi dan nilai *Davies-Bouldin* Index yang rendah, interpretasi metrik ini perlu dilakukan secara hati-hati. Pada algoritma OPTICS, nilai *Silhouette Score* hanya dihitung berdasarkan sebagian kecil data yang berhasil membentuk *cluster* padat, sementara sebagian besar wilayah lainnya diklasifikasikan sebagai *noise*. Dalam penelitian ini, sekitar 94% wilayah kabupaten/kota dikategorikan sebagai *noise* oleh OPTICS. Dengan demikian, nilai *Silhouette Score* yang tinggi tidak mencerminkan kualitas pengelompokan secara menyeluruh, melainkan hanya menunjukkan kohesi lokal yang kuat pada sejumlah kecil wilayah yang memiliki kepadatan data tinggi. Kondisi ini menyebabkan terjadinya bias struktural dalam evaluasi, karena mayoritas wilayah administratif tidak ikut berkontribusi dalam perhitungan kualitas *cluster*. Berbeda dengan FCM, yang dapat memetakan seluruh data ke dalam *cluster*.

Tabel 3. Pengujian dataset Garis Kemiskinan FCM

Algo	Cluster	SS	DBI
FCM	2	0.57	0.63
FCM	3	0.50	0.61
FCM	4	0.46	0.63
FCM	5	0.45	0.66
FCM	6	0.40	0.72
FCM	7	0.38	0.76
FCM	8	0.39	0.77
FCM	9	0.36	0.84
FCM	10	0.57	0.63

Sumber : (Hasil Penelitian, 2025)

Tabel 4. Pengujian Data Garis Kemiskinan OPTICS

Algo	minpts	Xi	SS	DBI	noise
OPTICS	5	0.1	0.61	0.63	478
OPTICS	5	0.05	0.54	0.66	397
OPTICS	10	0.05	-	-	-
OPTICS	10	0.1	0.74	0.33	452
OPTICS	15	0.05	-	-	-
OPTICS	15	0.1	-	-	-
OPTICS	25	0.05	-	-	-
OPTICS	25	0.05	-	-	-
OPTICS	33	0.1	-	-	-
OPTICS	33	0.05	-	-	-
OPTICS	50	0.1	-	-	-
OPTICS	50	0.05	-	-	-

Sumber : (Hasil Penelitian, 2025)

Untuk mengidentifikasi pola pengelompokan standar biaya hidup minimum antarwilayah, pengujian tambahan dilakukan pada variabel Garis Kemiskinan. Hasil eksperimen algoritm*a Fuzzy C-Means* (FCM) menunjukkan nilai optimal pada jumlah *cluster* 2 dengan skor *Silhouette* tertinggi 0.57 dan DBI tertinggi 0.63, seperti yang ditunjukkan dalam Tabel 3. Ini menunjukkan bahwa data Garis Kemiskinan paling tepat terbagi menjadi dua kelompok besar, masing-masing menunjukkan daerah dengan standar kemiskinan rendah dan tinggi.

Selain itu, hasil eksperimen algoritma OPTICS di Tabel 4 menunjukkan bahwa konfigurasi parameter terbaik dicapai pada MinPts 10 dan Xi 0.1. Kombinasi ini menghasilkan evaluasi metrik yang sangat baik, dengan Skor *Silhouette* 0.74 dan DBI rendah 0.33. Namun, temuan ini diikuti oleh jumlah suara yang sangat tinggi sebanyak 452 titik data dari 501 data yang menunjukkan bahwa pola distribusi data Garis Kemiskinan sangat tersebar dan tidak memiliki struktur kepadatan alami yang kuat. Akibatnya, OPTICS tidak berhasil mengelompokkan mayoritas wilayah, sedangkan FCM lebih efektif karena dapat menggabungkan seluruh data ke dalam profil yang jelas.

Tabel 5. Pengujian Data Pengeluaran Perkapita FCM

Algo	Cluster	SS	DBI
FCM	2	0.52	0.67
FCM	3	0.53	0.57
FCM	4	0.51	0.57
FCM	5	0.51	0.54
FCM	6	0.52	0.54
FCM	7	0.50	0.55
FCM	8	0.49	0.56
FCM	9	0.47	0.58
FCM	10	0.45	0.61

Sumber : (Hasil Penelitian, 2025)

Tabel 6. Pengujian Data Pengeluaran Perkapita OPTICS

Algo	minpts	Xi	SS	DBI	noise
OPTICS	5	0.1	0.69	0.38	396
OPTICS	5	0.05	0.51	0.673927	277
OPTICS	10	0.05	0.95	0.06	456
OPTICS	10	0.1	0.70	0.42	369
OPTICS	15	0.05	0.89	0.14	461
OPTICS	15	0.1	0.78	0.27	419
OPTICS	25	0.05	-	-	-
OPTICS	25	0.05	-	-	-
OPTICS	33	0.1	-	-	-
OPTICS	33	0.05	-	-	-
OPTICS	50	0.1	-	-	-
OPTICS	50	0.05	-	-	-

Sumber : (Hasil Penelitian, 2025)

Kapasitas ekonomi dan daya beli masyarakat diukur dengan variabel pengeluaran per kapita. Menurut hasil Tabel 5, algoritma *Fuzzy C-Means* (FCM) menunjukkan kinerja yang stabil pada jumlah cluster 2 dengan S*ilhouette Score* sebesar 0,52 dan DBI sebesar 0,67. Namun, nilai *Silhouette* sedikit meningkat pada cluster 3 (0,53), pemilihan *cluster* 2 dianggap lebih efektif untuk memecah profil wilayah menjadi dua kategori utama: wilayah dengan daya beli tinggi dan wilayah dengan daya beli rendah.

VOL. 20. NO. 2 FEBRUARI 2026
P-ISSN: 0216-6933 | E-ISSN: 2685-807X
Diterbitkan Oleh: LPPM Universitas Nusa Mandiri

INTI NUSA MANDIRI
DOI: https://doi.org/10.33480/inti.v20i2.7814

Tabel 6 menunjukkan hasil eksperimen algoritma OPTICS yang menunjukkan fenomena ekstrim yang kontras dengan FCM. Konfigurasi parameter terbaik dicapai pada MinPts 10 dan Xi 0.05. Ini menghasilkan nilai Skor *Silhouette* yang sangat tinggi sebesar 0,95 dan nilai DBI yang sangat rendah sebesar 0,06. Meskipun demikian, metrik yang tampaknya sempurna ini terdiri dari 456 titik data *noise*. Ini menunjukkan bahwa OPTICS hanya dapat menemukan beberapa wilayah dengan profil pengeluaran yang sangat padat (sekitar 9% dari total data), sementara mayoritas wilayah dianggap sebagai anomali atau data yang tersebar.

Tabel 7. Pengujian Data Gabungan FCM

Algo	Cluster	SS	DBI
FCM	2	0.38	1.01
FCM	3	0.29	1.14
FCM	4	0.28	1.07
FCM	5	0.25	1.12
FCM	6	0.28	0.98
FCM	7	0.24	1.02
FCM	8	0.25	1.08
FCM	9	0.23	1.10
FCM	10	0.22	1.08

Sumber : (Hasil Penelitian, 2025)

Untuk menghasilkan profil wilayah yang menyeluruh, analisis dilakukan pada kumpulan data gabungan yang mencakup ketiga dimensi kemiskinan: IPM, Garis Kemiskinan, dan Pengeluaran per Kapita. Ini dilakukan setelah memahami karakteristik parsial dari setiap variabel. Dengan konfigurasi *cluster* 2, algoritma *Fuzzy C-Means* (FCM) menunjukkan stabilitas performa dengan nilai *Silhouette Score* 0,38 dan nilai *Davies-Bouldin Index* (DBI) 1.01. Meskipun konfigurasi lain, seperti *cluster* 6, menghasilkan DBI yang lebih rendah (0.98), nilai *Silhouette Score* turun signifikan menjadi 0.28, menunjukkan penurunan kualitas kohesi internal klaster. Oleh karena itu, jumlah klaster ideal untuk FCM adalah *cluster* 2. Ini dapat membagi wilayah Indonesia menjadi dua karakteristik dominan tanpa menghilangkan generalitas data.

Tabel 8. Pengujian Data Gabungan OPTICS

Algo	minpts	Xi	SS	DBI	noise
OPTICS	5	0.1	0.73	0.32	473
OPTICS	5	0.05	0.56	0.54	464
OPTICS	10	0.05	-	-	-
OPTICS	10	0.1	-	-	-
OPTICS	15	0.05	-	-	-
OPTICS	15	0.1	-	-	-
OPTICS	25	0.05	-	-	-
OPTICS	25	0.05	-	-	-
OPTICS	33	0.1	-	-	-
OPTICS	33	0.05	-	-	-
OPTICS	50	0.1	-	-	-
OPTICS	50	0.05	-	-	-

Sumber : (Hasil Penelitian, 2025)

Sebaliknya, Tabel 8 menunjukkan hasil eksperimen algoritma OPTICS yang menunjukkan fenomena yang berbeda. Konfigurasi parameter terbaik dicapai dengan MinPts 5 dan Xi 0.1. Ini menghasilkan metrik validasi internal yang sangat baik dengan Skor Silhouette 0.73 dan DBI 0.32. Namun, keunggulan statistik ini diimbangi oleh kelemahan utama dalam pemetaan wilayah: algoritma ini menganggap 473 titik data, yang merupakan sekitar 94% dari populasi, sebagai *noise*. Hal ini menunjukkan bahwa struktur data sosial-ekonomi Indonesia sangat heterogen dan tersebar, dan metode berbasis densitas seperti OPTICS tidak dapat membentuk klaster yang mencakup mayoritas wilayah. Karena metode FCM memiliki kemampuan untuk segmentasi data sebesar 100%, yang sesuai dengan tujuan pembuatan kebijakan yang inklusif, model ini dipilih sebagai model terbaik untuk memetakan seluruh kabupaten atau kota.

Hasil segmentasi model terbaik FCM *cluster* 2 dianalisis secara menyeluruh dengan menggunakan teknik visualisasi spasial dan statistik. Gambar 1 menunjukkan peta pemetaan klaster yang menunjukkan distribusi geografis yang berbeda. Klaster 1 cenderung terkonsentrasi di pusat pertumbuhan ekonomi, perkotaan besar, dan ibu kota provinsi, sedangkan Klaster 2 cenderung terkonsentrasi di kabupaten, daerah penyangga, dan wilayah timur Indonesia.

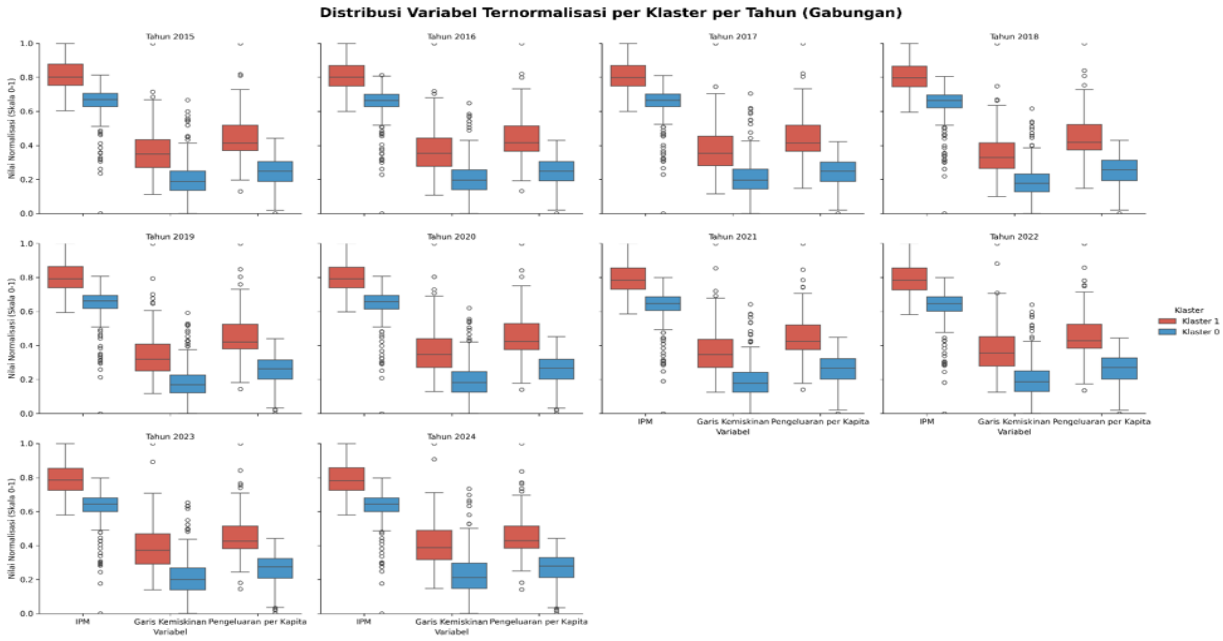
Sumber : (Hasil Penelitian, 2025)

Gambar 2. Pemetaan Cluster Wilaya Kabupaten/Kota

Gambar 2 menunjukkan distribusi variabel per tahun, yang memperkuat validitas pemisahan klaster ini. Kedua klaster menunjukkan separasi yang jelas pada variabel IPM dan pengeluaran per kapita, dengan area tumpang tindih yang minimal di grafik. Hal ini menunjukkan bahwa kedua kelompok

INTI NUSA MANDIRI
DOI: https://doi.org/10.33480/inti.v20i2.7814
VOL. 20. NO. 2 FEBRUARI 2026
P-ISSN: 0216-6933 | E-ISSN: 2685-807X
Diterbitkan Oleh: LPPM Universitas Nusa Mandiri

wilayah memiliki perbedaan substansial dalam hal kualitas hidup dan daya beli meskipun menggunakan pendekatan *fuzzy*. Selanjutnya, tren seluruh indikator untuk kedua klaster meningkat dari tahun 2015 hingga 2024, menurut analisis dinamika temporal yang ditunjukkan pada Gambar 3. Namun, temuan penting yang ditunjukkan oleh grafik tersebut adalah bahwa kesenjangan kesejahteraan antara Klaster 1 dan Klaster 2 tetap ada dan tidak menunjukkan penurunan yang signifikan dalam sepuluh tahun terakhir.



Sumber: (Hasil Penelitian, 2025)

Gambar 3. Box Plot Per Tahun Per Klaster

Pola yang konsisten menunjukkan karakteristik penting dari data kemiskinan Indonesia. Pola ini ditemukan melalui pengujian yang dilakukan baik terhadap data gabungan maupun variabel tunggal, seperti IPM, Garis Kemiskinan, dan Pengeluaran per Kapita. Pada seluruh skenario eksperimen, algoritma OPTICS secara konsisten menghasilkan metrik validasi internal *Silhouette Score* dan DBI yang jauh lebih baik daripada FCM. Sebagai contoh, OPTICS menghasilkan Skor *Silhouette* hampir sempurna sebesar 0,95 untuk variabel Pengeluaran per Kapita. Tingginya jumlah *noise*, yang mencapai 75% hingga 94% dari semua data, tetap mengikuti keunggulan statistik. Hal ini secara empiris menunjukkan bahwa struktur data sosial-ekonomi Indonesia sangat berbeda dan tersebar, atau tersebar. Akibatnya, tidak ada klaster padat alami yang meliputi mayoritas wilayah.

Algoritma *Fuzzy C-Means* (FCM), di sisi lain, memiliki kemampuan untuk melakukan segmentasi total pada seluruh skenario pengujian, yang menunjukkan stabilitas kinerja. Meskipun skor metrik FCM lebih rendah daripada OPTICS, mereka berhasil menggabungkan semua kabupaten dan kota ke dalam profil yang dapat ditafsirkan, dengan jumlah klaster optimal yang konsisten pada cluster 2 untuk data gabungan. Berdasarkan perbandingan komprehensif ini, disimpulkan bahwa FCM adalah metode yang paling sesuai untuk tujuan pemetaan kemiskinan nasional secara praktis, meskipun OPTICS lebih baik dalam mendeteksi anomali struktur data. Ini karena FCM mampu memberikan klasifikasi yang menyeluruh dan bermakna untuk setiap wilayah administratif tanpa menghilangkan informasi yang penting. Oleh karena itu, model FCM dengan *cluster* 2 dipilih untuk digunakan dalam sistem aplikasi web.

KESIMPULAN

Adanya perbedaan signifikan antara validitas dan utilitas praktis dalam evaluasi kualitas kelompok yang menggunakan dua metrik: *Silhouette Score* dan *Davies-Bouldin Index* (DBI). Meskipun algoritma OPTICS menghasilkan metrik yang sangat baik dengan skor *Silhouette* 0,7301 dan DBI 0,329, 94% data diklasifikasikan sebagai noise. Sebaliknya, metode Fuzzy C-Means (FCM) ternyata lebih cocok untuk tujuan pemetaan menyeluruh. Dengan jumlah *cluster* ideal adalah 2, FCM memberikan keseimbangan terbaik dengan

Silhouette Score 0,3894 dan DBI 1,0163, yang memungkinkan segmentasi seluruh dataset secara keseluruhan tanpa menghilangkan informasi wilayah. Profil kemiskinan di Indonesia terbagi menjadi dua pola utama berdasarkan hasil pengelompokan. Sementara klaster kedua terdiri dari wilayah berkembang yang dominan di kawasan pedalaman dan Indonesia Timur, ditandai dengan capaian pembangunan manusia dan ekonomi yang lebih rendah namun memiliki keuntungan standar biaya hidup yang lebih terjangkau. Kota besar dan pusat pertumbuhan ekonomi di Jawa dan Bali terdiri dari klaster pertama, yang merupakan wilayah dengan kesejahteraan tinggi (IPM dan pengeluaran tinggi) namun memiliki biaya hidup yang mahal. Temuan ini mendukung kenyataan bahwa ada disparitas spasial yang persisten antarwilayah. Untuk mempertajam profil kemiskinan multidimensi, penelitian lanjutan dapat mengembangkan sistem ini dengan menambahkan variabel infrastruktur atau sosial lainnya. Selain itu, metode pembelajaran mesin prediktif dapat digunakan untuk memproyeksikan tren pergeseran antar klaster di masa depan. Ini akan membantu dalam perencanaan kebijakan pengentasan kemiskinan yang lebih proaktif dan tepat sasaran.

REFERENSI

Andrian, G., Arisandi, D., & Handhayani, T. (2024). Clustering Data Meteorologi Wilayah Indonesia Timur Dengan Metode K-Means Dan Fuzzy C-Means. *INTI Nusa Mandiri*, *18*(2), 100–106. https://doi.org/10.33480/inti.v18i2.5039

Bahtiyar, R., Lestanti, S., & Nur Budiman, S. (2024). Penerapan Metode Fuzzy C-Means Cluster Ing Untuk Pengelompokan Kabupaten Di Jawa Timur Berdasarkan Indikator Kesejahteraan Rakyat. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(5), 3701–3710. https://doi.org/10.36040/jati.v7i5.7812

Darrell, F., & Reswara, W. (2024). *Analisis Algoritma K-Means Dan Fuzzy C-Means Untuk Persentase Rata - Rata Pengeluaran Untuk Makanan Dan Bukan Makanan*. *19*(1), 13–21.

Fan, Y., & Wang, M. (2024). Specification Mining Based on the Ordering Points to Identify the Clustering Structure Clustering Algorithm and Model Checking. *Algorithms*, *17*(1), 28. https://doi.org/10.3390/a17010028

Faturahman, R. D., & Hidayati, N. (2025). Implementasi Fuzzy C-Means Dalam Pengelompokan Tingkat Kemiskinan Pada Kabupaten/Kota Di Provinsi Jawa TengaH. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, *10*(1), 137–149. https://doi.org/10.29100/jipi.v10i1.5747

Hasan, Y. (2024). Pengukuran Silhouette Score Dan Davies-Bouldin Index Pada Hasil Cluster K-Means Dan Dbscan. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(3S1). https://doi.org/10.23960/jitet.v12i3S1.5001

Isah, I., Estri, M. N., & Larasati, N. (2024). Perbandingan Metode Fuzzy C-Means Dan Fuzzy Probabilistic C-Means Dalam Pengelompokan Provinsi Di Indonesia Berdasarkan Indeks Pembangunan Manusia Tahun 2022. *PROSIDING SEMINAR NASIONAL SAINS DATA*, *4*(1), 832–841. https://doi.org/10.33005/senada.v4i1.346

Nurhalizah, C. D., Damaliana, A. T., & Prasetya, D. A. (2025). OPTICS-Based Clustering of East Java Regencies and Cities by Divorce Factors. *Journal of Information Systems and Technology Research*, *4*(3), 113–123. https://doi.org/10.55537/jistr.v4i3.1227

Oktaviani, N., Fauzan, A., & Widyastuti, G. (2024). Pengelompokan Kabupaten/Kota di Jawa Barat Berdasarkan Tingkat Kesejahteraan Masyarakat Menggunakan K-Means Cluster. *Emerging Statistics and Data Science Journal*, *2*(2), 290–301. https://doi.org/10.20885/esds.vol2.iss.2.art22

Paembonan, S., & Abduh, H. (2021). Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat. *PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik*, *6*(2), 48. https://doi.org/10.51557/pt_jiit.v6i2.659

Santoso, E., & Anggraini, S. (2024). DISPARITAS PEMBANGUNAN ANTAR WILAYAH DI INDONESIA: MODEL DATA PANEL. *CERMIN: Jurnal Penelitian*, *8*(2), 355. https://doi.org/10.36841/cermin_unars.v8i2.5468

Shantal, M., Othman, Z., & Bakar, A. A. (2023). A Novel Approach for Data Feature Weighting Using Correlation Coefficients and Min–Max Normalization. *Symmetry*, *15*(12), 2185. https://doi.org/10.3390/sym15122185

Srirahayu, A., & Pribadie, L. S. (2023). Review Paper Data Mining Klasifikasi Data Mining. *Jurnal Ilmiah Informatika Global*, *14*(1). https://doi.org/10.36982/jiig.v14i1.2981

Tawakal, I., Effendi, M. M., & Majid, A. M. (2025). ANALISIS TINGKAT KEMISKINAN DENGAN ALGORITMA K-MEANS MENGGUNAKAN RAPIDMINER DI TINGKAT KOTA KABUPATEN DI JAWA TENGAH. *Journal of Information System Management (JOISM)*, *7*(1), 112–119. https://doi.org/10.24076/joism.2025v7i1.2084

Widianto, F. (2024). Implementasi Model SDLC Dalam Perancangan Sistem Informasi Manajemen Perpustakaan Berbasis Web. *Sistematis : Jurnal Ilmiah Sistem Informasi*, *1*(1), 60–68. https://doi.org/10.69533/4w86eq90

Ying, R., Taniguchi, T., Nabeta, K., Ishigami, K., Kikuchi, N., & Okamoto, S. (2024). Clustering First-order reversal curve diagram using the Gaussian mixture model and the Davies–Bouldin index. *Japanese Journal of Applied Physics*, *63*(11), 11SP04. https://doi.org/10.35848/1347-4065/ad892d