E-ISSN : 2962-9713
P-ISSN : 2963-928X

Perbandingan Metode *Mean Shift Clustering* Dan *Density Peaks Clustering* Dalam Pengelompokan Provinsi Berdasarkan Indikator Kesehatan Lingkungan Di Indonesia

Alvira Napu[1], Fahrezal Zubedi[2], La Ode Nashar[3]
[1]Program Studi Statistika, Jurusan Matematika, Universitas Negeri Gorontalo, alvira_s1statistika@mahaiswa.ung.ac.id
[2]Program Studi Statistika, Jurusan Matematika, Universitas Negeri Gorontalo, fahrezal@ung.ac.id
[3]Program Studi Statistika, Jurusan Matematika, Negeri Gorontalo, laode.nashar@ung.ac.id

Abstrak

Kesehatan lingkungan merupakan salah satu aspek penting dalam meningkatkan kualitas hidup masyarakat. Namun, capaian indikator kesehatan lingkungan masih menunjukkan perbedaan antarprovinsi di Indonesia sehingga diperlukan analisis pengelompokan untuk mengidentiikasi provinsi yang memiliki karakteristik serupa. Penelitian ini bertujuan membandingkan kinerja metode *Mean Shift Clustering* dan *Density Peaks Clustering* (DPC) dalam mengelompokkan provinsi berdasarkan indikator kesehatan lingkungan di Indonesia. Data yang digunakan merupakan data sekunder Profil Kesehatan Indonesia Tahun 2024 yang mencakup 38 provinsi dengan sembilan indikator kesehatan lingkungan. Analisis dilakukan menggunakan bahasa pemrograman *Python*. Penentuan parameter *bandwidth* pada *Mean Shift Clustering* menggunakan pendekatan *quantile*, sedangkan parameter *cut-off distance* pada DPC menggunakan pendekatan persentil. Kinerja kedua metode dievaluasi menggunakan *Silhouette Coeficient* (SC) dan *Davies-Bouldin Index* (DBI). Hasil penelitian menunjukkan bahwa *Mean Shift Clustering* menghasilkan dua klaster, sedangkan DPC menghasilkan dua klaster utama dan satu klaster *outlier* yang merepresentasikan provinsi dengan karakteristik berbeda dari klaster lainnya. DPC memperoleh nilai SC sebesar 0,7257 dan DBI sebesar 0,3342, sedangkan *Mean Shift Clustering* memperoleh nilai SC sebesar 0,7010 dan DBI sebesar 0,4440. Hasil tersebut menunjukkan bahwa DPC menghasilkan struktur klaster yang lebih kompak dan memiliki pemisahan antarklaster yang lebih baik, sehingga lebih sesuai digunakan untuk pengelompokan provinsi berdasarkan indikator kesehatan lingkungan di Indonesia.

Kata Kunci: *Mean Shift Clustering*, *Density Peaks Clustering, kesehatan lingkungan, Silhouette Coeficient*, *Davies-Bouldin Index*.

Abstract

Environmental health is a key aspect of improving people's quality of life. However, achievements in environmental health indicators still vary across provinces in Indonesia, necessitating cluster analysis to identify provinces with similar characteristics. This study aims to compare the performance of the Mean Shift Clustering and Density Peaks Clustering (DPC) methods in clustering provinces based on environmental health indicators in Indonesia. The data used were secondary data from the 2024 Indonesian Health Profile, covering 38 provinces with nine environmental health indicators. The analysis was conducted using the Python programming language. The bandwidth parameter in Mean Shift Clustering was determined using a quantile approach, while the

E-ISSN : 2962-9713
P-ISSN : 2963-928X

cut-off distance parameter in DPC was determined using a percentile approach. The performance of both methods was evaluated using the Silhouette Coefficient (SC) and the Davies-Bouldin Index (DBI). The results showed that Mean Shift Clustering produced two clusters, while DPC produced two main clusters and one outlier cluster representing provinces with characteristics different from the other clusters. DPC obtained an SC value of 0.7257 and a DBI value of 0.3342, while Mean Shift Clustering obtained an SC value of 0.7010 and a DBI value of 0.4440. These results indicate that DPC produces a more compact cluster structure and better separation between clusters, making it more suitable for grouping provinces based on environmental health indicators in Indonesia.

Keywords: *Mean Shift Clustering, Density Peaks Clustering, environmental health, Silhouette Coeficient*, *Davies-Bouldin Index*.

Corresponding Author:
Alvira Napu

Universitas Negeri Gorontalo; alvira_s1statistika@mahaiswa.ung.ac.id

PENDAHULUAN

Kesehatan lingkungan merupakan salah satu aspek penting yang berperan dalam upaya pencegahan penyakit dan peningkatan derajat kesehatan masyarakat. Pemerintah Indonesia melalui Kementerian Kesehatan menyelenggarakan program penyehatan lingkungan sebagai upaya pencegahan gangguan kesehatan yang bersumber dari faktor risiko lingkungan, baik dari aspek fisik, kimia, biologi, maupun sosial (Kementerian Kesehatan Indonesia, 2024). Lingkungan yang sehat dapat terwujud apabila berbagai faktor dalam lingkungan manusia tersusun dan dikelola dengan baik sehingga meningkatkan derajat kesehatan masyarakat (Diah Aliyatus Saidah, 2022).

Kondisi kesehatan lingkungan pada setiap provinsi di Indonesia memiliki keragaman yang berbeda, tercermin dari sembilan indikator yang digunakan oleh Kementerian Kesehatan Indonesia, yaitu akses air minum layak, akses sanitasi layak, status stop buang air besar sembarangan, pengawasan tempat fasilitas umum, tempat pengolahan pangan yang memenuhi syarat, pelaksanaan Kabupaten/Kota Sehat, pengelolaan limbah medis fasyankes, penerapan GERMAS, dan rumah layak huni. Kesamaan indikator tersebut dapat dijadikan dasar untuk mengelompokkan provinsi menggunakan analisis cluster, yaitu teknik analisis data eksploratif yang mengelompokkan objek berdasarkan tingkat kemiripan tanpa label kelas sebelumnya.

Seiring meningkatnya kompleksitas data, metode density-based clustering berkembang untuk mengakomodasi data berdimensi tinggi, distribusi tidak linier, serta keberadaan noise dan outlier (Erich Schubert, 2017). Di antara metode tersebut, Mean Shift Clustering dan Density Peaks Clustering (DPC) memiliki mekanisme penentuan pusat klaster yang berbeda dibandingkan metode density-based lainnya seperti DBSCAN dan OPTICS (Ricardo J. G. B. Campello, 2013). Mean Shift Clustering bekerja dengan menggeser titik data secara iteratif menuju area kepadatan tertinggi menggunakan kernel density estimation, sehingga mampu menemukan klaster dengan bentuk tidak beraturan tanpa perlu menentukan jumlah klaster di awal. Sementara itu, DPC menentukan pusat klaster berdasarkan dua

E-ISSN : 2962-9713
P-ISSN : 2963-928X

kriteria, yaitu kepadatan lokal yang tinggi dan jarak yang relatif jauh terhadap titik berkepadatan lebih tinggi (Rodriguez dan Laio, 2014).

Penelitian sebelumnya menunjukkan bahwa Mean Shift unggul dibandingkan K-Means dan DBSCAN dalam mengelompokkan data ancaman jaringan yang tidak merata (Cinderatama Toga Aldila, 2022), sementara DPC menunjukkan akurasi yang lebih tinggi dibandingkan metode clustering klasik pada berbagai dataset UCI (Xiuxi Wei a, 2023). Meskipun demikian, penelitian mengenai pengelompokan indikator kesehatan lingkungan di Indonesia masih lebih banyak menggunakan metode clustering konvensional seperti K-Means dan Partitioning Around Medoids, sedangkan kajian yang membandingkan Mean Shift Clustering dan DPC secara khusus pada data ini masih terbatas.

Berdasarkan uraian tersebut, penelitian ini bertujuan membandingkan kinerja metode Mean Shift Clustering dan Density Peaks Clustering dalam mengelompokkan 38 provinsi di Indonesia berdasarkan sembilan indikator kesehatan lingkungan Tahun 2024, sekaligus menentukan metode yang paling sesuai digunakan untuk keperluan tersebut. Hasil penelitian diharapkan dapat memberikan gambaran pola kesehatan lingkungan antarprovinsi serta menjadi dasar pendukung pengambilan keputusan dan perumusan kebijakan kesehatan lingkungan yang lebih tepat sasaran.

METODE

Pada Penelitian ini menggunakan pendekatan kuantitatif dengan data sekunder yang bersumber dari Profil Kesehatan Indonesia Tahun 2024, mencakup 38 provinsi sebagai objek dan sembilan indikator kesehatan lingkungan sebagai variabel, yaitu persentase rumah tangga dengan akses air minum layak, akses sanitasi layak, desa/kelurahan stop buang air besar sembarangan, tempat fasilitas umum yang diawasi sesuai standar, tempat pengolahan pangan yang memenuhi syarat, Kabupaten/Kota Sehat, fasyankes yang mengelola limbah medis sesuai standar, penerapan GERMAS, dan rumah tangga yang menempati rumah layak huni. Seluruh provinsi dijadikan sampel melalui teknik sampling jenuh. Analisis dilakukan menggunakan bahasa pemrograman Python melalui Google Colaboratory.

Mean Shift Clustering

Mean Shift Clustering bekerja dengan menggeser titik-titik data secara iteratif menuju area dengan kepadatan tinggi, sehingga mampu menemukan klaster dengan bentuk tidak beraturan, tahan terhadap outlier, serta bersifat bebas model karena tidak mengasumsikan bentuk klaster tertentu (Abdillah, 2024). Dalam penelitian ini, nilai bandwidth diestimasi menggunakan pendekatan. Quantile merupakan nilai proporsi dalam rentang 0 hingga 1 yang digunakan untuk mengestimasi bandwidth berdasarkan distribusi jarak antar data, sehingga nilai bandwidth yang diperoleh dapat menyesuaikan karakteristik penyebaran data. Proses pengelompokan Mean Shift mengikuti beberapa tahapan yaitu, sebagai berikut (Roid Fitrah Utari, 2024).

1. Penentuan nilai bandwidth diestimasi dengan pendekatan quantil dengan persamaan berikut.

$$h = \frac{1}{n}\sum_{i=1}^{n} d_i\ (\lfloor qn \rfloor)$$

Keterangan:

h	: Nilai estimasi *bandwidth*
q	: Nilai quantil

E-ISSN : 2962-9713
P-ISSN : 2963-928X

n : Jumlah dat
$d_{i(\lfloor qn \rfloor)}$: Jarak dari data ke-i menuju tetangga terdekat ke-$\lfloor qn \rfloor$

2. Menghitung nilai kernel Gaussian dengan Persamaan (Roid Fitrah Utari, 2024).

$$K\ (x - x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2h^2}\right)$$

Keterangan:
x : Pusat massa saat ini
x_i : Titik data ke-i
h : *Bandwidth*
$K\ (x - x_i)$: Nilai Kernel gaussian antara titik x_i dan pusat massa x

3. Menghitung vektor *Mean Shift* menggunakan Persamaan sebagai berikut (Awal Lidya Musaffak, 2025).

$$m(x) = \frac{\sum K(x - x_i)x_i}{\sum K(x - x_i)}$$

Perhitungan pusat massa baru (m(x)) dilakukan berdasarkan rata-rata berbobot titik-titik data dalam radius bandwidth, dengan bobot ditentukan oleh fungsi kernel. Pusat kemudian diperbarui menggunakan nilai (m(x)) dan proses ini dilakukan secara iteratif hingga perubahan posisi pusat semakin kecil atau mencapai kondisi konvergen (Abdillah, 2024). Setelah konvergen, titik-titik data yang menuju mode yang sama atau berdekatan dikelompokkan ke dalam satu klaster.

Density Peaks Clustering

Pada tahun 2014, Alex Rodriguez dan Alessandro Laio memperkenalkan algoritma Density Peak Clustering (DPC), yaitu algoritma berbasis kepadatan yang menentukan pusat klaster berdasarkan dua karakteristik utama, yaitu kepadatan lokal (p) dan jarak relatif (δ). Pusat klaster dipilih dari titik data yang memiliki nilai (p) dan (δ) tinggi melalui grafik keputusan (Min Wangl, 2017). Dalam penelitian ini, beberapa nilai percentile pada rentang 1% hingga 10% dari keseluruhan jarak yang telah dihitung dieksplorasi untuk menentukan nilai (dc). Rentang tersebut dipilih karena (dc) pada metode DPC umumnya ditetapkan sebagai persentase kecil dari seluruh pasangan jarak, sehingga dapat menghasilkan estimasi kepadatan yang sesuai dan mampu merepresentasikan struktur data dengan baik. Proses pengelompokan Density Peaks Clustering mengikuti beberapa tahapan utama yaitu sebagai berikut:

1. Hitung matriks jarak *euclidean* dengan persamaan berikut.

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{p} (x_{ik}) - x_{jk})^2}$$

2. Penentuan nilai dc diestimasi dengan pendekatan persentil dengan persamaan berikut (Ridwan Efendi, 2024).

$$d_c = d_{\lfloor n_d \times \frac{p}{100} \rfloor}$$

dengan banyaknya pasangan jarak dihitung menggunakan:

$$n_d = \binom{n}{2} = \frac{n(n-1)}{2}$$

Keterangan:
d_c : Nilai jarak batas (*cut-off distance*)
n_d : Jumlah seluruh pasangan jarak antar data
n : Jumlah data

E-ISSN : 2962-9713
P-ISSN : 2963-928X

p : Nilai persentil yang digunakan

3. Hitung nilai kepadatan lokal dengan persamaan berikut (Ridwan Efendi, 2024).

$$\rho_i = \sum_j \exp\left(-\frac{d(x_i, x_j)^2}{d_c^2}\right)$$

4. Nilai jarak lokal (δ) dihitung dengan mencari jarak $d(x_i, x_j)$ minimum antara titik x_i dan titik lain yang memiliki kepadatan yang lebih tinggi, dideinisikan sebagai berikut (Ridwan Efendi, 2024):

$$\delta_i = \begin{cases} \min\limits_{j:\rho_j>\rho_i} d(x_i, x_j), \\ \max\limits_j d(x_i, x_j) \end{cases}$$

5. Buat graik keputusan ρ dan δ untuk melihat pusat klaster dan outlier.
6. Pemberian label akhir klaster.

Metrik Evaluasi

Metrik evaluasi merupakan ukuran yang digunakan untuk menilai kualitas hasil pengelompokan yang dihasilkan oleh suatu metode klasterisasi. Hasil evaluasi tersebut dapat digunakan untuk membandingkan beberapa alternatif model atau parameter sehingga diperoleh struktur klaster yang paling optimal (Hasan, 2024). Pada penelitian ini, evaluasi kualitas klaster dilakukan menggunakan metrik evaluasi internal, yaitu.

1. *Silhouette Score* (SC)

SC digunakan untuk mengevaluasi kualitas hasil pengelompokan dengan mengukurseberapa baik suatu objek ditempatkan pada klaster yang sesuai. Silhouette Score untuk objek ke-i dideinisikan sebagai berikut (Rousseeuw, 1987).

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

Dengan a_i menyatakan rata-rata jarak objek ke-*i* terhadap seluruh objek lain dalam klaster yang sama, sedangkan b_i merupakan rata-rata jarak minimum objek ke-i terhadap objek-objek pada klaster lain yang paling dekat.

2. *Davies-Bouldin Index* (DBI)

Davies-Bouldin Index merupakan ukuran validitas internal klaster yang diperkenalkan oleh (Davies dan Bouldin, 1979). Suatu hasil pengelompokan dikatakan baik apabila setiap klaster memiliki anggota yang saling berdekatan dan pusat antar klaster terpisah sejauh mungkin (Muhammad Sholeh, 2023). Secara matematis ,Davies-Bouldin Index dirumuskan sebagai berikut (Davies dan Bouldin, 1979):

$$DBI = \frac{1}{K}\sum_{i=1}^{k} \max_{j\neq i}\left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)}\right)$$

dengan k merupakan jumlah klaster, σ_i dan σ_j menyatakan ukuran penyebaran klaster ke-i dan ke-j, sedangkan $d(c_i,c_j)$ adalah jarak antara pusat klaster ke-i dan ke-j. Nilai DBI yang lebih kecil menunjukkan kualitas pengelompokan yang lebih baik karena mengindikasikan bahwa setiap klaster memiliki tingkat kekompakan yang tinggi dan keterpisahan yang baik terhadap klaster lainnya (Yayan Sopyan, 2022).

HASIL DAN PEMBAHASAN

Mean Shift Clustering

E-ISSN : 2962-9713
P-ISSN : 2963-928X

Penentuan parameter bandwidth dilakukan melalui eksplorasi nilai quantile pada rentang 0,1–1,0. Semakin besar nilai quantile, semakin besar nilai bandwidth yang dihasilkan sehingga jumlah klaster yang terbentuk semakin sedikit. Hasil eksplorasi nilai quantile dan estimasi bandwidth ditunjukkan pada tabel berikut.

Tabel 1. Hasil Eksplorasi nilai quantile dan estimasi bandwidth

Quantil	Bandwidth	Jumlah Klaster	SC	DBI
0,1	22,7819	9	0,2632	0,7555
0,2	29,2950	6	0,3661	0,4168
0,3	41,1407	3	0,6780	0,3259
0,4	60,6218	3	0,6781	0,3259
0,5	64,7610	3	0,6781	0,3259
0,6	68,3113	3	0,6781	0,3259
O,7	77,9790	2	0,7010	0,4440
0,8	116,7612	1	NaN	NaN
0,9	126,0593	1	NaN	NaN
1,0	178,2143	1	NaN	NaN

Nilai quantile 0,7 dipilih sebagai parameter optimal karena menghasilkan SC tertinggi sebesar 0,7010 dengan bandwidth 77,979, Meskipun nilai DBI pada quantile 0,3 hingga 0,6 lebih rendah, nilai DBI pada quantile 0,7 masih tergolong rendah sehingga tetap menunjukkan tingkat kekompakan dalam klaster dan keterpisahan antar klaster yang baik.

Setelah diperoleh bandwidth optimal, nilai tersebut digunakan dalam kernel Gaussian untuk menentukan bobot setiap titik data berdasarkan kedekatannya dengan pusat saat ini, di mana titik yang lebih dekat memiliki bobot lebih besar. Bobot tersebut selanjutnya digunakan untuk menghitung pusat massa baru dan vektor Mean Shift sebagai arah serta besar perpindahan menuju daerah dengan kepadatan lebih tinggi. Pusat kemudian diperbarui secara iteratif hingga mencapai kondisi konvergen, kemudian titik-titik yang menuju mode yang sama atau berdekatan dikelompokkan ke dalam satu klaster. Berdasarkan proses tersebut, diperoleh hasil pengelompokan provinsi menggunakan metode Mean Shift Clustering yang disajikan pada tabel berikut.

Tabel 2. Hasil Pengelompokan Provinsi Menggunakan Mean Shift Clustering

Klaster	Jumlah	Provinsi
0	26	DKI Jakarta, Bali, Jawa Timur, DI Yogyakarta, Gorontalo, Nusa Tenggara Barat, Jawa Tengah, Jawa Barat, Kepulauan Riau, Sulawesi Utara, Banten, Riau, Sulawesi Selatan, Aceh, Kalimantan Utara, Kalimantan Timur, Maluku Utara, Sulawesi Tengah, Sumatera Selatan, Sumatera Barat, Lampung, Kepulauan Bangka Belitung, Jambi, Kalimantan Tengah, Kalimantan Selatan, Bengkulu.
1	12	Sulawesi Tenggara, Maluku, Sumatera Utara, Nusa Tenggara Timur, Papua, Papua Barat, Kalimantan Barat, Papua Barat Daya, Papua Tengah, Sulawesi Barat, Papua Selatan, Papua Pegunungan

Berdasarkan tabel 2, hasil pengelompokan menggunakan metode Mean Shift Clustering menghasilkan 2 cluster yang terbentuk berdasarkan kemiripan karakteristik data indikator kesehatan lingkungan.

E-ISSN : 2962-9713
P-ISSN : 2963-928X

Density Peaks Clustering

Tahap awal pada metode Density Peaks Clustering adalah menghitung jarak antar objek menggunakan jarak Euclidean sebagai dasar dalam menghitung kepadatan lokal setiap objek melalui parameter *cut-off distance* (dc). Nilai dc dilakukan melalui beberapa percobaan nilai persentil, yaitu dari 1% hingga 10% untuk memperoleh nilai dc yang optimal. Hasil eksplorasi nilai persentil dan estimasi nilai dc ditunjukkan pada tabel berikut.

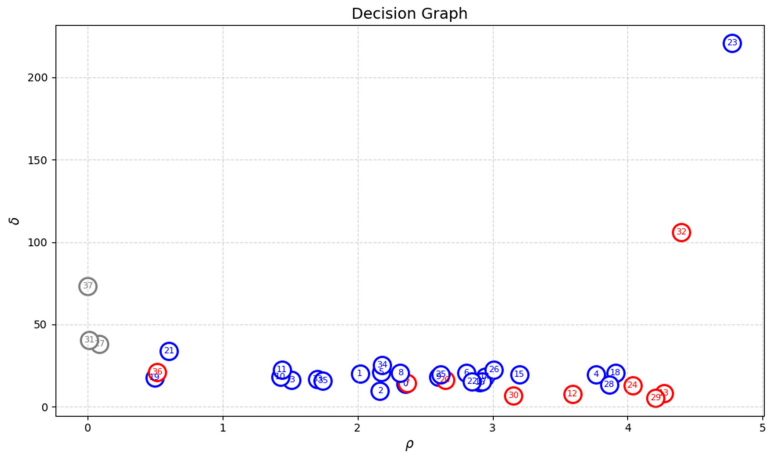
Tabel 3. Hasil Eksplorasi Nilai Percentile dan estimasi dc

Persentil	dc	Jumlah klaster	Jumlah *Outlier*	SC	DBI
1	13,3381	2	2	0,7194	0,3591
2	15,0225	2	2	0,7194	0,3591
3	16,3564	2	2	0,7194	0,3591
4	18,0318	2	3	0,7257	0,3342
5	19,1777	2	3	0,7257	0,3342
6	19,8580	2	1	0,7139	0,3886
7	20,6545	2	1	0,7139	0,3886
8	21,5023	2	1	0,7139	0,3886
9	22,3230	2	0	0,7010	0,4440
10	23,1462	2	0	0,7010	0,4440

Berdasarkan Tabel 4.6, persentil 4% dan 5% menghasilkan nilai SC tertinggi, yaitu sebesar 0,7257, serta nilai DBI yang sama dan relatif kecil, yaitu sebesar 0,3342, yang menunjukkan bahwa klaster memiliki tingkat kekompakan yang baik dan pemisahan antar klaster yang cukup jelas. Meskipun kedua persentil menghasilkan kualitas pengelompokan yang sama berdasarkan nilai SC dan DBI, penelitian ini memilih nilai dc dari persentil 4%, yaitu sebesar 18,0318, karena nilai dc yang lebih kecil lebih mampu mempertahankan karakteristik lokal data dengan memberikan pengaruh lebih besar pada objek-objek yang berada pada jarak terdekat. Dengan demikian, nilai dc sebesar 18,0318 dipilih sebagai parameter optimal.

Nilai dc tersebut digunakan untuk menghitung kepadatan lokal (ρ) dan jarak minimum (δ) setiap provinsi, yang kemudian divisualisasikan melalui grafik keputusan (ρ pada sumbu horizontal, δ pada sumbu vertikal) guna menentukan pusat klaster dan outlier. Adapun visualisasi *decision graph* dapat dilihat pada gambar berikut.

Gambar 1. Visualisasi *Decision graph*



E-ISSN : 2962-9713
P-ISSN : 2963-928X

Berdasarkan Gambar 4.3, sebagian besar provinsi terkonsentrasi pada bagian tengah grafik dengan nilai (\delta) relatif rendah, yang menunjukkan kedekatan dengan objek lain yang lebih padat. Data terbagi menjadi dua cluster utama, yaitu cluster pertama yang memiliki jumlah anggota lebih banyak dengan pusat pada titik 23 (Sumatera Barat), serta cluster kedua dengan jumlah anggota lebih sedikit dan pusat pada titik 32 (Sulawesi Barat). Kedua pusat cluster memiliki nilai (\rho) dan (\delta) relatif tinggi sehingga dianggap mewakili karakteristik kesehatan lingkungan masing-masing kelompok. Selain itu, terdapat tiga provinsi yang teridentifikasi sebagai outlier, yaitu Papua Barat, Papua Tengah, dan Papua Pegunungan, karena memiliki kepadatan lokal rendah dan posisi yang terpisah dari kedua kelompok utama. Adapun Pelabelan cluster yang dihasilkan oleh algoritma Density Peaks Clustering disajikan pada tabel berikut.

Tabel 4. Hasil Pengelompokan Provinsi Menggunakan DPC

Klaster	Jumlah	Provinsi
0	26	DKI Jakarta, Bali, Jawa Timur, DI Yogyakarta, Gorontalo, Nusa Tenggara Barat, Jawa Tengah, Jawa Barat, Kepulauan Riau, Sulawesi Utara, Banten, Riau, Sulawesi Selatan, Aceh, Kalimantan Utara, Kalimantan Timur, Maluku Utara, Sulawesi Tengah, Sumatera Selatan, Sumatera Barat, Lampung, Kepulauan Bangka Belitung, Jambi, Kalimantan Tengah, Kalimantan Selatan, Bengkulu.
1	9	Sulawesi Tenggara, Maluku, Sumatera Utara, Nusa Tenggara Timur, Papua, Kalimantan Barat, Papua Barat Daya, Sulawesi Barat, Papua Selatan.
-1	3	Papua Barat, Papua Tengah, Papua Pegunungan.

Berdasarkan Tabel 4.8, Hasil pengelompokan metode Density Peaks Clustering menunjukkan adanya heterogenitas karakteristik antarprovinsi yang tercermin dari terbentuknya dua klaster utama dan satu outlier.

Perbandingan Metode Mean Shift Clustering dan Density Peaks Clustering (DPC)

Setelah proses *clustering* menggunakan kedua metode selesai dilakukan, tahap selanjutnya adalah membandingkan hasil pengelompokan yang diperoleh berdasarkan metrik evaluasi, dengan tujuan mengetahui kualitas klaster yang dihasilkan sekaligus menentukan metode yang memberikan struktur pengelompokan terbaik untuk data indikator kesehatan lingkungan. Hasil perbandingan kedua metode disajikan pada tabel berikut.

Tabel 5. Perbandingan Hasil Evaluasi

Metode	SC	DBI
Mean Shift Clustering	0,7010	0,4440
Density Peaks Clustering	0,7257	0,3342

Berdasarkan tabel tersebut, Density Peaks Clustering memperoleh nilai SC sebesar 0,7257 dan DBI sebesar 0,3342, sedangkan Mean Shift Clustering memperoleh nilai SC sebesar 0,7010 dan DBI sebesar 0,4440. Nilai SC yang lebih tinggi pada DPC menunjukkan bahwa klaster yang dihasilkan memiliki tingkat kohesi dan separasi yang lebih baik dibandingkan Mean Shift Clustering. Sementara itu, nilai DBI yang lebih rendah pada DPC menandakan bahwa objek-

E-ISSN : 2962-9713
P-ISSN : 2963-928X

objek dalam klaster yang sama lebih homogen dan jarak antarklaster lebih terpisah dengan baik.

Selisih nilai SC antara kedua metode adalah sebesar 0,0247 atau sekitar 3,52%, sedangkan nilai DBI pada DPC lebih rendah sebesar 0,1098 dibandingkan *Mean Shift Clustering*. Secara keseluruhan, kedua metode menunjukkan kualitas pengelompokan yang tergolong baik karena sama-sama menghasilkan nilai SC di atas 0,70 dan nilai DBI yang relatif kecil. Meskipun selisih nilai evaluasi tersebut relatif kecil, kedua metrik secara konsisten menunjukkan bahwa metode DPC menghasilkan kualitas pengelompokan yang lebih baik dibanding *Mean Shift Clustering.*

SIMPULAN

Mean Shift Clustering menghasilkan dua klaster sedangkan Density Peaks Clustering menghasilkan dua klaster utama dan satu klaster outlier yang merepresentasikan provinsi dengan karakteristik berbeda secara signifikan. Berdasarkan metrik validasi internal, Density Peaks Clustering memperoleh nilai Silhouette Coefficient sebesar 0,7257 dan Davies-Bouldin Index sebesar 0,3342, lebih baik dibandingkan Mean Shift Clustering yang memperoleh SC sebesar 0,7010 dan DBI sebesar 0,4440. Hasil ini menunjukkan bahwa Density Peaks Clustering menghasilkan struktur klaster yang lebih kompak, memiliki pemisahan antarklaster yang lebih baik, serta mampu mengidentifikasi provinsi dengan karakteristik menyimpang, sehingga lebih direkomendasikan untuk pengelompokan data provinsi berdasarkan indikator kesehatan lingkungan di Indonesia.

DAFTAR PUSTAKA

Abdillah, A. R. (2024). Implementasi metode mean shift untuk clustering kejadian kecelakaan lalu lintas di kota makassar. . *UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG.*

Awal Lidya Musaffak, K. M. (2025). Penerapan metode mean shift clustering untuk mengelompokkan wilayah berdasarkan pengelolaan sampah. *Jurnal Mahasiswa Teknik Informatika*, doi: doi.org/10.36040/jati.v9i3.13777.

Cinderatama Toga Aldila, R. Z. (2022). Implementasi metode k-means, dbscan, dan meanshift untuk analisis jenis ancaman jaringan pada intrusion detection system. *Journal INOVTEK Polbeng*, doi: doi.org/10.35314/isi.v7i1.2336.

Davies, D. L. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(2):224–227.

Diah Aliyatus Saidah, R. S. (2022). Pengelompokan provinsi di indonesia berdasarkan indikator kesehatan lingkungan menggunakan metode partitioning around medoids dengan validasi indeks internal. *GAUSSIAN*, 11:302–312.

Erich Schubert, S. E. (2017). Dbscan revisited, revisited: Why and how you should (still) use dbscan. . *ACM Transactions on Data base Systems (TODS)*.

Hasan, Y. (2024). Pengukuran silhouette score dan davies-bouldin index pada hasil cluster k-means dan dbscan. *KAKIFIKOM (Kumpulan Artikel Karya Ilmiah Fakultas Ilmu Komputer*, Vol. 06 No. 01.

Indonesia, K. K. (2024). *Proil Kesehatan 2024. Kementerian.* Jakarta Selatan: Jl. H.R. Rasuna Said Blok X-5 Kav. 4-9.

Min Wang, F. M.-H.-X. (2017). Active learning through density clustering. *Expert Systems with Applications,*, 85:305–317. doi: doi.org/10.1016/j.eswa.2017.05.046.

E-ISSN : 2962-9713
P-ISSN : 2963-928X

Muhammad Sholeh, K. A. (2023). Perbandingan evaluasi metode davies bouldin, elbow dan silhouette pada model clustering dengan menggunakan algoritma k means. *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, Vol.8 No.1.

Ricardo J. G. B. Campello, D. M. (2013). Density-Based Clustering Based. *Lecture Notes in Computer Science, Springer, Berlin, Heidelberg.*

Ridwan Efendi, A. J. (2024). Penentuan pusat klaster secara otomatis pada algoritma density peaks clustering berbasis metode inter quartile range. *Jurnal Informatika dan Teknik Elektro Terapan*, doi: doi.org/10.23960/jitet.v12i3.4997.

Rodriguez, A. d. (2014). Clustering by fast search and ind of density peaks. *Journal of EMDR Practice and Research*, 344. doi: 10.1126/science.1242072.

Roid Fitrah Utari, F. I. (2024). Pengelompokan data pendistribusian listri menggunakan algoritma mean shift. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 1015–1023.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. . *Journal of Computational and Applied Mathematics*, 20:53–65.

Xiuxi Weia, M. (2023). An overviewon density peaks clustering. *Neurocomputing*, 554:126633.

Yayan Sopyan, A. D. (2022). Analisis algoritma k-means dan davies bouldin index dalam mencaricluster terbaik kasus perceraian di kabupaten kuningan. . *Building of Informatics, Technology and Science (BITS)*, Volume 4, No 3. doi: 10.47065/bits.v4i3.2697.