TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1653-1667
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10939

Klasterisasi Indikator Data Kemiskinan Kabupaten/Kota di Indonesia Menggunakan K-Means dengan Normalisasi Z-Score dan Evaluasi DBI

Elfonda Dandi Kurnia Putra*, Galet Guntoro Setiaji, Ahmad Rifa'i, Astrid Novita Putri

Fakultas Teknologi Informasi dan Komunikasi, Program Studi Teknik Informatika, Universitas Semarang, Semarang, Indonesia
Email: [1,*]dandiputra426@gmail.com, [2]galet@usm.ac.id, [3]rifai@usm.ac.id, [4]astrid@usm.ac.id
Email Penulis Korespondensi: dandiputra426@gmail.com

Abstrak–Penelitian ini bertujuan untuk mengoptimasi pengelompokan data tingkat kemiskinan di Indonesia menggunakan metode *K-Means* klastering dengan pendekatan normalisasi *Z-Score* serta evaluasi menggunakan *Davies-Bouldin Index (DBI)*. Permasalahan utama dalam pengelompokan data kemiskinan adalah terdapat perbedaan skala antar variabel yang dapat mempengaruhi hasil klastering. Dengan adanya hal itu maka, dilakukan normalisasi data menggunakan metode *Z-Score* untuk menyamakan skala antar variabel. Data yang digunakan merupakan dataset sekunder yang terdiri dari 514 kab/kota di indonesia berjudul "klasifikasi tingkat kemiskinan di indonesia" yang diperoleh dari repository kaggle yang berlaku sebagai penyusun adalah Ermila, Marcelo, Faisal dino bahtiar dengan update sesuai yang tertera di web adalah 3 tahun lalu tepatnya pada tahun 2023. Data tersebut merupakan indikator kemiskinan yang telah melalui proses seleksi variabel, di mana variabel yang tidak relevan atau memiliki makna yang sama dan berpotensi menimbulkan bias tidak digunakan dalam proses analisis. Proses klastering dilakukan dengan dua skenario, yakni jumlah klaster dengan K=2 dan K=3 untuk membandingkan hasil pengelompokan. Hasil penelitian menunjukkan bahwa metode K-Means mampu mengelompokkan data secara efektif berdasarkan kategori tingkat kemiskinan. Hasil evaluasi menggunakan davies-boudin index menunjukan nilai DBI sebesar 0,9417 untuk k=2 dan 2,706 untuk k=3. Berdasarkan perbandingan klastering tersebut, skenario k=2 menghasilkan nilai DBI yang lebih rendah sehingga digunakan sebagai acuan untuk menentukan tingkat klasifikasinya. Kontribusi penelitian ini terletak pada pengembangan proses klasterisasi yang mengintegrasikan seleksi indikator, normalisasi Z-Score, algoritma K-Means, dan evaluasi Davies-Bouldin Index pada 514 kabupaten/kota di Indonesia untuk memperoleh pengelompokan wilayah berdasarkan kemiripan karakteristik indikator kemiskinan. Dengan demikian, konsep yang dibuat dalam pendekatan yang digunakan untuk proses penelitian ini mampu menghasilkan pengelompokan yang lebih optimal dan dapat dijadikan sebagai dasar dalam pengambilan keputusan penanggulangan kemiskinan secara efektif dan juga tepat sasaran.

Kata Kunci: K-Means; Z-Score; Klastering; Kemiskinan; *Davies-Bouldin Index*

Abstract–This study aims to optimize the clustering of poverty-level data in Indonesia using the *K-Means* clustering method with a *Z-Score* normalization approach and evaluation using the *Davies-Bouldin Index (DBI)*. The main problem in clustering poverty data is the difference in scales among variables, which can affect the clustering results. Therefore, data normalization using the *Z-Score* method was performed to equalize the scales among variables. The data used in this study are secondary data consisting of 514 regencies/cities in Indonesia entitled "Classification of Poverty Levels in Indonesia," obtained from the Kaggle repository. The dataset was compiled by Ermila, Marcelo, and Faisal Dino Bahtiar and was last updated three years ago, specifically in 2023, as stated on the website. The data consist of poverty indicators that have undergone a variable selection process, in which variables that are irrelevant or have similar meanings and may potentially cause bias were excluded from the analysis. The clustering process was conducted using two scenarios, namely K=2 and K=3, to compare the clustering results. The results show that the *K-Means* method can effectively cluster the data based on poverty-level categories. The evaluation using the *Davies-Bouldin Index* shows DBI values of 0.9417 for K=2 and 2.706 for K=3. Based on the comparison of the clustering results, the K=2 scenario produced a lower DBI value and was therefore used as a reference for determining the classification levels. The contribution of this study lies in the development of a clustering process that integrates indicator selection, Z-Score normalization, the K-Means algorithm, and Davies-Bouldin Index evaluation across 514 regencies and municipalities in Indonesia to obtain regional groupings based on similarities in the characteristics of poverty indicators. Thus, the approach developed in this study can produce more optimal clustering results and can be used as a basis for supporting decision-making in poverty alleviation efforts effectively and appropriately.

Keywords: *K-Means; Z-Score*; Klastering; Poverty; *Davies-Bouldin Index*

1. PENDAHULUAN

Kemiskinan masih menjadi salah satu persoalan yang dihadapi Indonesia. Kondisi kemiskinan tidak hanya berkaitan dengan pendapatan, tetapi juga dapat dilihat dari berbagai keterbatasan yang dialami masyarakat. Fuady et al. (2022) menunjukkan bahwa kemiskinan di Indonesia memiliki dimensi yang beragam sehingga pengukurannya tidak cukup hanya menggunakan satu ukuran. Hasil penelitian Kause & Fithriyah (2024) juga menunjukkan bahwa kemiskinan multidimensi berkaitan dengan karakteristik pendidikan, lokasi tempat tinggal, dan kondisi sosial rumah tangga. Hal tersebut menunjukkan bahwa pemahaman mengenai kemiskinan perlu mempertimbangkan beberapa indikator yang dapat menggambarkan kondisi suatu wilayah secara lebih menyeluruh.

Jawa Tengah merupakan salah satu wilayah yang masih menghadapi persoalan kemiskinan. Berdasarkan data Badan Pusat Statistik Provinsi Jawa Tengah, persentase penduduk miskin pada September 2022 mencapai 10,98 persen atau sekitar 3,86 juta orang. Angka tersebut meningkat sebesar 0,05 persen poin dibandingkan Maret 2022 yang berada pada angka 10,93 persen (Badan Pusat Statistik Provinsi Jawa Tengah, 2023). Kondisi tersebut menunjukkan bahwa persoalan kemiskinan masih menjadi bagian penting dalam pembangunan wilayah. Publikasi Provinsi Jawa Tengah Dalam Angka 2023 juga menunjukkan bahwa karakteristik sosial, ekonomi, ketenagakerjaan, dan kesejahteraan

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1653-1667
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10939

masyarakat merupakan bagian dari indikator yang dapat digunakan untuk menggambarkan kondisi wilayah (Badan Pusat Statistik Provinsi Jawa Tengah., 2023).

Perbedaan kondisi antarwilayah juga berkaitan dengan berbagai faktor sosial dan ekonomi. Sari et al. (2023), menemukan bahwa tingkat kemiskinan di Jawa Tengah berkaitan dengan faktor pendidikan, pengangguran, kondisi ekonomi, lokasi, dan Indeks Pembangunan Manusia (IPM). Penelitian Sabilillah & Robertus (2024) juga menunjukkan bahwa urbanisasi, modal manusia, dan pengangguran perlu diperhatikan dalam melihat kondisi kemiskinan perkotaan di Jawa Tengah. Dengan adanya berbagai indikator tersebut, kondisi setiap kabupaten/kota dapat memiliki karakteristik yang berbeda. Oleh karena itu, analisis yang mampu mengelompokkan wilayah berdasarkan kemiripan karakteristik dapat membantu memberikan gambaran yang lebih jelas mengenai kondisi kemiskinan antarwilayah.

Salah satu pendekatan yang dapat digunakan untuk melihat kemiripan karakteristik tersebut adalah klastering. Mayasari & Nugraha (2023) menerapkan algoritma *K-Means* untuk mengelompokkan kabupaten/kota di Jawa Tengah berdasarkan data kemiskinan tahun 2022. Penelitian tersebut menunjukkan bahwa *K-Means* dapat digunakan untuk membentuk kelompok kabupaten/kota berdasarkan karakteristik data kemiskinan. Namun, penelitian tersebut masih terbatas pada wilayah Jawa Tengah dan menggunakan ruang lingkup wilayah pada satu provinsi. Penelitian ini mengembangkan aspek tersebut dengan memperluas objek analisis menjadi 514 kabupaten/kota di seluruh Indonesia. Perluasan cakupan tersebut dilakukan untuk memperoleh gambaran karakteristik kemiskinan yang lebih luas dan memungkinkan perbandingan kondisi antarwilayah pada tingkat nasional. Penelitian Tawakal et al. (2025) juga menggunakan *K-Means* untuk menganalisis tingkat kemiskinan kabupaten/kota di Jawa Tengah menggunakan data BPS periode 2020–2024. Penelitian tersebut menghasilkan lima kelompok dan menggunakan *Davies-Bouldin Index (DBI)* untuk mengevaluasi hasil klastering. Penggunaan DBI pada penelitian tersebut menjadi dasar bahwa kualitas hasil klastering perlu dinilai setelah proses pengelompokan. Penelitian ini mengembangkan pendekatan tersebut dengan tidak hanya menggunakan DBI sebagai evaluasi, tetapi juga membandingkan dua skenario jumlah klaster, yaitu K=2 dan K=3. Perbandingan tersebut digunakan untuk menentukan skenario yang menghasilkan kualitas klaster lebih baik berdasarkan nilai DBI terkecil. Hasil tersebut memperlihatkan bahwa *K-Means* dapat digunakan untuk memberikan gambaran mengenai perbedaan karakteristik kemiskinan antarwilayah.

Penelitian Alif & Fahmi (2026) selanjutnya melakukan klasterisasi kabupaten/kota di Jawa Tengah menggunakan tujuh indikator sosial ekonomi, yaitu IPM, proporsi penduduk miskin ekstrem, penduduk sangat miskin, pengeluaran per kapita, upah minimum kabupaten/kota, tingkat pengangguran terbuka, dan jumlah rumah tidak layak huni. Penggunaan banyak indikator memberikan gambaran kondisi sosial ekonomi yang lebih luas, tetapi penelitian ini mengembangkan aspek tersebut dengan melakukan seleksi indikator berdasarkan relevansi dan kesamaan makna antarvariabel. Dari sejumlah indikator yang tersedia, penelitian ini menggunakan empat indikator utama, yaitu persentase penduduk miskin (P0), rata-rata lama sekolah (RLS), pengeluaran per kapita (PPK), dan indeks pembangunan manusia (IPM). Seleksi tersebut dilakukan agar indikator yang digunakan tetap mewakili karakteristik kemiskinan tanpa memasukkan variabel yang memiliki informasi serupa dan berpotensi memengaruhi hasil pengelompokan. Penelitian tersebut menunjukkan bahwa penggunaan beberapa indikator sosial ekonomi dapat memberikan gambaran yang lebih luas mengenai karakteristik wilayah yang dikelompokkan. Dengan demikian, penggunaan beberapa indikator dalam proses klastering menjadi relevan ketika tujuan penelitian adalah melihat kemiripan kondisi sosial ekonomi antarwilayah.

Meskipun *K-Means* telah digunakan pada beberapa penelitian mengenai kemiskinan, proses pengolahan data tetap perlu diperhatikan karena setiap indikator dapat memiliki satuan dan rentang nilai yang berbeda. Sinaga dan Sipayung (2026), misalnya, menggunakan *Standard Scaler* atau *Z-Score* sebelum proses *K-Means* pada pengelompokan provinsi berdasarkan indikator sosial ekonomi yang mencakup persentase penduduk miskin, IPM, dan tingkat pengangguran terbuka. Standardisasi tersebut digunakan agar perbedaan skala antarindikator tidak memberikan pengaruh yang tidak proporsional terhadap perhitungan jarak. Penelitian ini mengembangkan penggunaan standardisasi tersebut dengan menerapkan normalisasi Z-Score pada tingkat kabupaten/kota. Penggunaan unit analisis yang lebih rinci memungkinkan karakteristik kemiskinan diamati pada 514 kabupaten/kota, sehingga variasi kondisi antarwilayah dapat terlihat lebih spesifik dibandingkan pengelompokan pada tingkat provinsi. Dengan demikian, standardisasi data menjadi salah satu tahap yang relevan ketika beberapa indikator dengan skala berbeda digunakan secara bersamaan dalam *K-Means.*

Selain standardisasi, kualitas klaster yang terbentuk juga perlu dievaluasi. Pengelompokan data tidak hanya ditujukan untuk menghasilkan sejumlah kelompok, tetapi juga perlu diketahui apakah kelompok tersebut memiliki tingkat kekompakan dan pemisahan yang baik. Nanda Shalsadilla et al. (2023)menggunakan *Davies-Bouldin Index* untuk menentukan jumlah klaster optimum pada pengelompokan wilayah berdasarkan indikator kemiskinan. Penelitian tersebut menunjukkan bahwa jumlah klaster dapat dibandingkan berdasarkan nilai *DBI*, dengan nilai yang lebih kecil menunjukkan hasil pengelompokan yang lebih baik. Penelitian ini mengadopsi prinsip tersebut dengan menggunakan nilai DBI sebagai dasar evaluasi kualitas klaster, kemudian mengembangkannya melalui penggabungan beberapa tahapan pengolahan data, yaitu seleksi indikator, normalisasi Z-Score, proses K-Means, dan perbandingan skenario jumlah klaster. Dengan tahapan tersebut, evaluasi DBI dilakukan setelah karakteristik indikator diseleksi dan diseragamkan skalanya sehingga hasil pengelompokan dapat dibandingkan secara lebih terstruktur. Penggunaan *DBI* juga diterapkan oleh Tawakal et al. (2025) dalam mengevaluasi hasil *K-Means* pada data kemiskinan kabupaten/kota di Jawa Tengah.

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1653-1667
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10939

Berdasarkan perbandingan penelitian terdahulu, terdapat beberapa aspek yang masih dapat dikembangkan dalam klasterisasi data kemiskinan. Penelitian Mayasari dan Nugraha (2023) serta Tawakal et al. (2025) masih berfokus pada wilayah Jawa Tengah, sedangkan Sinaga dan Sipayung (2026) menggunakan provinsi sebagai unit analisis. Kondisi tersebut menunjukkan adanya ruang untuk memperluas analisis pada tingkat kabupaten/kota dengan cakupan seluruh Indonesia. Di sisi lain, penelitian Alif dan Fahmi (2026) menggunakan sejumlah indikator sosial ekonomi untuk menggambarkan kondisi wilayah, sedangkan penelitian ini mengembangkan proses tersebut melalui seleksi indikator berdasarkan relevansi dan kesamaan makna antarvariabel. Sementara itu, penelitian Nanda Shalsadilla et al. (2023) menunjukkan penggunaan DBI untuk menentukan jumlah klaster optimum, yang kemudian dikembangkan dalam penelitian ini dengan membandingkan hasil K-Means pada beberapa skenario jumlah klaster setelah dilakukan seleksi indikator dan normalisasi Z-Score.

Dengan demikian, kesenjangan penelitian yang menjadi dasar penelitian ini terletak pada belum diterapkannya secara terpadu perluasan analisis hingga 514 kabupaten/kota di Indonesia, seleksi indikator yang relevan, normalisasi Z-Score, proses K-Means, dan evaluasi DBI untuk membandingkan kualitas hasil pengelompokan. Penelitian ini mengembangkan penelitian terdahulu melalui penggabungan tahapan tersebut dalam satu alur analisis yang sistematis. Pengembangan tersebut diharapkan dapat menghasilkan pengelompokan wilayah yang lebih terukur berdasarkan kemiripan karakteristik indikator kemiskinan.

Kontribusi penelitian ini adalah mengembangkan pendekatan klasterisasi kemiskinan pada tingkat kabupaten/kota melalui integrasi seleksi indikator, normalisasi Z-Score, algoritma K-Means, dan evaluasi Davies-Bouldin Index. Pendekatan tersebut diterapkan pada 514 kabupaten/kota di Indonesia dan digunakan untuk membandingkan dua skenario jumlah klaster, yaitu K=2 dan K=3. Hasil perbandingan tersebut memberikan dasar kuantitatif dalam menentukan skenario pengelompokan yang memiliki kualitas klaster lebih baik berdasarkan nilai DBI.

2. METODOLOGI PENELITIAN

2.1 Data dan Sumber Data

Penelitian ini memakai data sekunder yang diambil dari *Kaggle*, tepatnya dari dataset berjudul "Klasifikasi Tingkat Kemiskinan di Indonesia." Dalam dataset itu, ada informasi data mengenai 514 kabupaten dan kota dari seluruh Indonesia. Indikator sosial dan ekonomi yang berhubungan dengan kemiskinan di setiap daerah juga tercatat di sana. Beberapa data yang tercantum antara lain persentase penduduk miskin (P0), rata-rata lama sekolah (RLS), pengeluaran per kapita (yang sudah disesuaikan) (PPK), indeks pembangunan manusia (IPM), usia harapan hidup (UHH), Persentase rumah tangga yang memiliki akses terhadap sanitasi layak (PRTS), Persentase rumah tangga yang memiliki akses terhadap air minum layak (PRTA), tingkat pengangguran terbuka (TPT), tingkat partisipasi tenaga kerja (TPAK), produk domestik regional bruto (PDRB). Alasan memilih dataset ini cukup jelas data yang tersedia sudah sangat lengkap di level kabupaten/kota dan indikator yang dipakai memang cocok untuk proses pengelompokan menggunakan *K-Means*. Semua data yang digunakan benar-benar berasal dari dataset itu tanpa mengubah nilainya, kecuali pada tahap pra-pemrosesan adanya proses penyeleksian terhadap variable yang akan digunakan buat analisis.

2.2 Objek Penelitian dan Parameter Analisis

a. Kemiskinan
   Kemiskinan dalam penelitian ini ialah berfokus pada kondisi sosial ekonomi 514 kab/kota di Indonesia. Objek ini berfokus pada 4 indikator utama yang sudah di leburkan beberapa variabel yang sama maknanya yakni, persentase penduduk miskin, rata-rata lama sekolah, pengeluaran per kapita, dan indeks pembangunan manusia.
b. Klastering
   Klastering meruopakan teknik dalam data mining yang berfungsi mengelompokan data ke dalam beberapa kategori. Pada penelitian ini, klastering digunakan untuk membagi kab/kota ke dalam kelompok tertentu berdasarkan kemiripan karakteristik indikator kemiskinannya, sehingga wilayah yang berada dalam satu kelompok memiliki kondisi sosial-ekonomi yang relatif serupa.
c. Z-Score (normalisasi data)
   Z-Score berfungsi untuk menyamakan skala dari seluruh variabel yang dianalisis. Dalam data kemiskinan, tiap indikator memiliki satuan yang beragam, misalnya nilai pengeluaran per kapita yang mencapai ribuan rupiah berbanding dengan persentase penduduk miskin yang angkanya jauh lebih kecil. Dengan Z-Score, semua data ditransformasikan ke skala standar yang seimbang, sehingga setiap indikator memberikan kontribusi yang adil dalam proses perhitungan K-Means.
d. K-Means (algoritma pengelompokan)
   K-Means merupakan algoritma utama yang bertugas membagi 514 kabupaten/kota ke dalam kelompok (*klaster*) berbasis perhitungan jarak Euclidean ke titik pusat (*centroid*) secara berulang. Dalam penelitian ini, pengujian K-Means dibagi menjadi dua skenario kelompok, yaitu skenario $K = 2$ dan $K = 3$.
e. Davies-Bouldin Index (evaluasi model)
   Davies-Bouldin Index digunakan sebagai tolok ukur untuk menilai seberapa bagus kualitas klaster yang terbentuk dari algoritma K-Means. DBI mengukur keseimbangan antara kekompakan anggota di dalam satu klaster (*SSW*) dan jarak pemisahan antar klaster (*SSB*), di mana semakin kecil nilai DBI maka semakin baik kelompok yang dihasilkan.

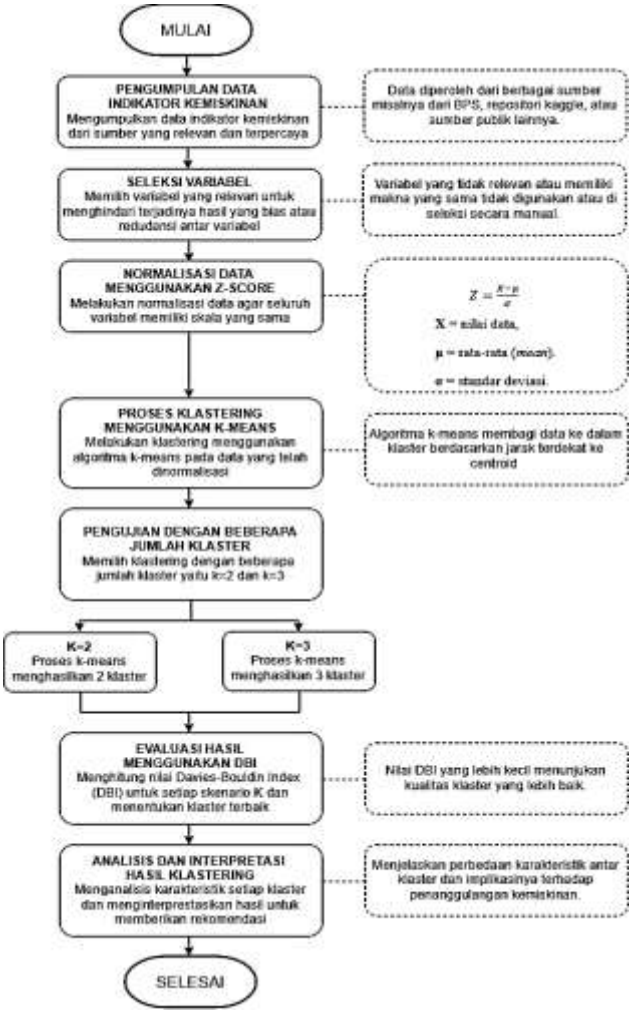
TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1653-1667
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10939

2.3 Kerangka Dasar Penelitian

Penelitian ini mengambil pendekatan kuantitatif. Analisisnya menggunakan data mining untuk mengelompokkan tingkat kemiskinan di Indonesia. Bentuk datanya adalah data sekunder, diambil dari *repositori Kaggle*, yang memang fokus pada indikator kemiskinan tadi. Variabel yang dipakai terdiri dari sejumlah indikator yang menggambarkan kondisi sosial ekonomi masyarakat. Tapi, tidak semua variabel dapat digunakan semua. Peneliti mengidentifikasi dulu variable yang akan digunakan untuk menghindari hasil yang bias. Jadi, hanya indikator yang benar-benar efektif yang dipakai untuk proses analisis. Metode utama yang dipilih adalah algoritma *K-Means.* Tujuannya tentu untuk mengelompokkan data berdasar kemiripan antar wilayah. Untuk menyamakan skala semua variabel, peneliti menggunakan normalisasi *Z-Score.* Setelah itu, dilakukan pengecekan kualitas hasil klastering-nya menggunakan *Davies-Bouldin Index (DBI).* Konseptual penelitiannya dimulai dari pengolahan data mentah, melakukan normalisasi, proses klastering, lalu evaluasi hasil yang terbaik.

2.4 Tahapan Penelitian

Tahapan penelitian (Nanda Shalsadilla et al., 2023) (Sinaga & Sipayung, 2026) dilakukan secara sistematis. Dapat dilihat pada Gambar 1 yang dimana terdapat flow chart yang digunakan sebagai acuan awal dalam melakukan penelitian sehingga hasil yang didapat lebih terstruktur.



Gambar 1. Tahapan penelitian.

a. Pengumpulan data indikator kemiskinan.
   Data yang digunakan merupakan data sekunder yang memuat indikator kemiskinan pada 514 kabupaten/kota di Indonesia. Data dikumpulkan dari sumber yang relevan dan terpercaya, kemudian disusun berdasarkan wilayah dan indikator yang digunakan dalam proses klasterisasi. Data tersebut menjadi dasar dalam menentukan karakteristik masing-masing kabupaten/kota berdasarkan indikator kemiskinan.
b. Seleksi variabel yang relevan untuk menghindari terjadinya hasil yang bias.
   Seleksi variabel dilakukan untuk menentukan indikator yang relevan dan dapat digunakan dalam proses klasterisasi. Variabel yang tidak relevan, memiliki informasi yang berulang, atau berpotensi menimbulkan redundansi antarvariabel dipertimbangkan untuk tidak digunakan. Tahap ini dilakukan agar variabel yang digunakan mampu merepresentasikan karakteristik indikator kemiskinan dan menghasilkan proses pengelompokan yang lebih optimal.

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1653-1667
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10939

c. Normalisasi data menggunakan *Z-Score*.
dengan Z merupakan nilai hasil normalisasi, X merupakan nilai data, μ merupakan nilai rata-rata (*mean*), dan σmerupakan standar deviasi. Hasil normalisasi menghasilkan data dalam skala yang relatif seragam sehingga setiap variabel dapat digunakan secara seimbang dalam perhitungan jarak pada algoritma K-Means.
d. Proses klastering menggunakan *K-Means*.
Proses klasterisasi dilakukan menggunakan algoritma *K-Means* untuk mengelompokkan kabupaten/kota berdasarkan kemiripan karakteristik indikator kemiskinan. Algoritma K-Means mengelompokkan data berdasarkan kedekatan jarak antara setiap data dengan pusat klaster (*centroid*). Dalam penelitian ini, data hasil normalisasi Z-Score digunakan sebagai masukan dalam proses perhitungan jarak dan pembentukan klaster.
e. Pengujian dengan beberapa jumlah klaster (K=2 dan K=3).
Pengujian dilakukan dengan menggunakan dua skenario jumlah klaster, yaitu K=2 dan K=3. Pada skenario K=2, seluruh kabupaten/kota dikelompokkan ke dalam dua klaster berdasarkan kedekatannya terhadap centroid masing-masing. Selanjutnya, proses yang sama dilakukan pada skenario K=3 dengan tiga centroid. Pengujian pada kedua skenario tersebut dilakukan untuk mengetahui konfigurasi jumlah klaster yang mampu memberikan pengelompokan wilayah yang lebih baik.
f. Evaluasi hasil menggunakan *DBI*.
Hasil klasterisasi pada K=2 dan K=3 selanjutnya dievaluasi menggunakan *Davies-Bouldin Index (DBI)*. Evaluasi ini digunakan untuk mengukur kualitas hasil pengelompokan berdasarkan tingkat kedekatan data dalam klaster dan keterpisahan antar-klaster. Nilai DBI yang lebih rendah menunjukkan hasil klasterisasi yang lebih baik, yaitu memiliki anggota yang lebih kompak dalam klaster dan lebih terpisah dari klaster lainnya. Nilai DBI dari skenario K=2 dan K=3 kemudian dibandingkan untuk menentukan jumlah klaster yang memberikan hasil terbaik.
g. Analisis dan interpretasi hasil klastering.
Tahap akhir dilakukan dengan menganalisis dan menginterpretasikan hasil klasterisasi berdasarkan karakteristik indikator kemiskinan pada masing-masing klaster. Analisis dilakukan dengan melihat pola dan perbedaan karakteristik antarwilayah yang tergabung dalam klaster yang sama. Hasil tersebut digunakan untuk memberikan gambaran mengenai pengelompokan kabupaten/kota berdasarkan kemiripan karakteristik indikator kemiskinan serta menjadi dasar dalam menarik kesimpulan penelitian.

2.5 Proses Normalisasi Data

Berdasarkan (Wongoutong, 2024), proses normalisasi merupakan tahap prapemrosesan yang penting pada algoritma *K-Means* dengan menstandarkan skala antar variable sehingga perbedaan skala atau satuan antarvariabel tidak menyebabkan variable tertentu untuk mendominasi perhitungan jarak dalam proses *K-Means* klastering. Algoritma ini menggunakan *Euclidean Distance* sebagai ukuran kedekatan antar data. Tanpa normalisasi, variabel yang memiliki rentang nilai lebih besar akan memberikan pengaruh yang lebih dominan terhadap hasil klastering. Rumus yang digunakan terdapat pada nomor urut (1)(2)(3).

$$Z = \frac{X-\mu}{\sigma} \quad (1)$$

Pada Persamaan (1), Xmerupakan nilai asli dari data pada suatu variabel, μmerupakan nilai rata-rata (*mean*), dan σmerupakan standar deviasi. Nilai Xdikurangi dengan rata-rata μ, kemudian hasilnya dibagi dengan standar deviasi σ. Perhitungan tersebut menghasilkan nilai Z-Score yang menunjukkan posisi suatu data terhadap nilai rata-rata variabelnya. Hasil normalisasi ini digunakan agar setiap variabel memiliki skala yang sebanding dalam proses klasterisasi.

$$\mu = \frac{\Sigma X_i}{n} \quad (2)$$

Pada Persamaan (2), μmerupakan nilai rata-rata dari suatu variabel, $\sum X_i$merupakan jumlah seluruh nilai data pada variabel tersebut, dan $n$merupakan jumlah data yang digunakan. Nilai rata-rata diperoleh dengan menjumlahkan seluruh nilai data kemudian membaginya dengan jumlah data. Nilai μyang diperoleh digunakan dalam perhitungan Z-Score pada proses normalisasi.

$$\sigma = \sqrt{\frac{\Sigma(X_i-\mu)^2}{N}} \quad (3)$$

Pada Persamaan (3), σmerupakan standar deviasi, $X_i$merupakan nilai data ke-i, μmerupakan nilai rata-rata, dan Nmerupakan jumlah data yang digunakan dalam perhitungan. Standar deviasi digunakan untuk mengetahui tingkat penyebaran data terhadap nilai rata-ratanya. Nilai standar deviasi tersebut selanjutnya digunakan dalam pembagian pada rumus Z-Score sehingga nilai setiap variabel dapat dinormalisasi pada skala yang sama. Proses ini bertujuan untuk menghindari dominasi variabel tertentu dalam perhitungan jarak pada algoritma *K-Means*.

2.6 Proses Klastering *K-Means*

Proses klastering menggunakan algoritma *K-Means* dilakukan secara iteratif dengan menentukan pusat klaster, menghitung jarak setiap data terhadap pusat klaster, menentukan keanggotaan klaster, dan memperbarui *centroid*

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1653-1667
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10939

berdasarkan anggota klaster. Tahapan tersebut dilakukan sampai diperoleh hasil pengelompokan yang memenuhi kondisi berhenti yang ditentukan (AlShammari, 2024).

a. Menentukan jumlah klaster (K).
   Jumlah klaster di tentukan sesuai dengan tujuan pengelompokan yang akan dilakukan.
b. Menentukan *centroid* awal.
   Centroid awal di tentukan sebagai pusat awal masing-masing klaster. Pada penelitian ini, *centroid* awal ditentukan secara acak (*random*) dari data yang digunakan.
c. Proses penentuan klaster.
   Digunakan untuk menentukan hasil nilai agar mengetahui masuk pada klaster mana kab/kota tersebut pada iterasi ke-1...ke-n. (Setiaji et al., 2025)

$$D_{Cn} = \sqrt{((Z_1 - C_{11})^2 + (Z_2 - C_{12})^2 + ((Z_3 - C_{13})^2 \ldots\ldots + (Z_n - C_{1n})^2)} \quad (4)$$

   Pada Persamaan (4), Dmerupakan jarak antara data dengan centroid awal, sedangkan $Z_1, Z_2, \ldots, Z_n$merupakan nilai atribut data yang telah dinormalisasi menggunakan Z-Score. Sementara itu, $C_{11}, C_{12}, \ldots, C_{1n}$merupakan nilai centroid awal pada masing-masing atribut. Jarak dihitung dengan membandingkan nilai setiap atribut pada data dengan nilai atribut pada centroid. Hasil perhitungan jarak digunakan untuk mengetahui centroid yang paling dekat dengan setiap data.
d. Menentukan centroid baru.
   Dihitung pada iterasi sebelumnya yang nantinya akan digunakan pada iterasi berikutnya.

$$D_{C1} = \sqrt{\{(Z_1 - NewC1_1)^2 + (Z_2 - NewC1_2)^2 + \cdots + (Z_n - NewC1_n)^2\}} \quad (5)$$

   Pada Persamaan (5), D_{C1}merupakan jarak data terhadap centroid Klaster 1. Nilai $Z_1, Z_2, \ldots, Z_n$merupakan nilai atribut data yang telah dinormalisasi, sedangkan $C_{1,1}, C_{1,2}, \ldots, C_{1,n}$merupakan nilai centroid Klaster 1 pada masing-masing atribut. Perhitungan ini digunakan untuk mengetahui tingkat kedekatan data terhadap centroid Klaster 1 pada setiap iterasi.

$$D_{C2} = \sqrt{\{(Z_1 - NewC2_1)^2 + (Z_2 - NewC2_2)^2 + \cdots + (Z_n - NewC2_n)^2\}} \quad (6)$$

   Pada Persamaan (6), D_{C2}merupakan jarak data terhadap centroid Klaster 2. Nilai $Z_1, Z_2, \ldots, Z_n$menunjukkan nilai atribut data yang telah dinormalisasi, sedangkan $C_{2,1}, C_{2,2}, \ldots, C_{2,n}$merupakan nilai centroid Klaster 2 pada masing-masing atribut. Nilai jarak yang diperoleh digunakan untuk membandingkan kedekatan data terhadap Klaster 2 dengan klaster lainnya.

$$D_{C3} = \sqrt{\{(Z_1 - NewC3_1)^2 + (Z_2 - NewC3_2)^2 + (Z_3 - NewC3_3)^2 \cdots + (Z_n - NewC3_n)^2\}} \quad (7)$$

e. Pada Persamaan (7), D_{C3}merupakan jarak data terhadap centroid Klaster 3. Nilai $Z_1, Z_2, \ldots, Z_n$merupakan nilai atribut data yang telah dinormalisasi, sedangkan $C_{3,1}, C_{3,2}, \ldots, C_{3,n}$merupakan nilai centroid Klaster 3 pada masing-masing atribut. Perhitungan tersebut digunakan untuk mengetahui jarak setiap data terhadap centroid Klaster 3 dan selanjutnya dibandingkan dengan jarak terhadap centroid lainnya.
f. Menghitung jarak data terhadap *centroid*.
   Jarak antara setiap data dengan masing-masing centroid dihitung menggunakan *Euclidean Distance*. Perhitungan ini digunakan untuk mengetahui kedekatan setiap data terhadap pusat klaster.
g. Menentukan keanggotaan klaster.
   Setiap data ditempatkan ke dalam klaster yang memiliki jarak *Euclidean* paling kecil terhadap *centroid.*
h. Memperbarui *centroid*.
   Centroid setiap klaster dihitung Kembali berdasarkan nilai rata- rata dari seluruh data yang menjadi anggota klaster tersebut.
i. Melakukan iterasi.
   Tahapan perhitungan jarak, penentuan keanggotaan klaster, dan pembaruan *centroid* dilakukan secara berulang sampai kondisi berhenti terpenuhi, yaitu ketika posisi *centroid* tidak mengalami perubahan yang berarti atau jumlah iterai maksimum yang telah ditentukan tercapai.

2.6.1 Perhitungan Jarak (*Euclidean Distance*)

Rumus perhitungan jarak:

$$D = \sqrt{((Z_1 - C_1)^2 + (Z_2 - C_2)^2 + \ldots + (Zn - Cn)^2)} \quad (8)$$

Pada Persamaan (8), D merupakan jarak antara data dengan centroid, Z_i merupakan nilai atribut ke-i dari data yang telah dinormalisasi, dan $C_i$merupakan nilai atribut ke-i dari centroid. Perhitungan dilakukan pada seluruh atribut yang digunakan dalam penelitian. Nilai jarak yang diperoleh digunakan untuk menentukan klaster yang paling dekat dengan data. Data kemudian ditempatkan pada klaster yang memiliki nilai jarak paling kecil.Perhitungan ini digunakan untuk menentukan kedekatan antar data terhadap *centroid*.(Pangestu & Fitriani, 2022).

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1653-1667
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10939

2.7 Evaluasi Menggunakan *Davies-Bouldin Index*

Davies-Bouldin Index digunakan untuk mengevaluasi kualitas hasil klastering. Evaluasi kualitas kelompok diuji menggunakan *Davies-Bouldin Index* (*DBI)* untuk memastikan bahwa klaster yang terbentuk memiliki tingkat separasi yang baik dan kohesi internal yang tinggi (Syahrul Adi Saputra et al., 2026) mengukur rasio antara jarak dalam klaster dan jarak antar klaster. Interpretasi nilai *DBI*:

a. Semakin kecil nilai *DBI* maka semakin baik kualitas klaster
b. Semakin besar nilai *DBI* maka klaster kurang optimal

DBI digunakan untuk menentukan jumlah klaster terbaik antara K=2 dan K=3 dalam penelitian ini (Nanda Shalsadilla et al., 2023).

$$SSW_i = \frac{\sum D_i}{N_i} \tag{9}$$

Pada Persamaan (9), SSW_imerupakan nilai rata-rata jarak anggota pada Klaster $i$terhadap centroid klaster tersebut. $\sum D_i$merupakan jumlah jarak seluruh data yang menjadi anggota Klaster $i$terhadap centroid, sedangkan $N_i$merupakan jumlah data pada Klaster i. Nilai SSW digunakan untuk melihat tingkat kedekatan anggota klaster terhadap centroid. Semakin kecil nilai SSW, semakin dekat anggota klaster dengan centroid sehingga klaster tersebut semakin kompak.

$$SSB_{\{ij\}} = \sqrt{\left\{ \backslash sum_{\{k=1\}}^{\{n\}} \left(C_{\{ik\}} - C_{\{jk\}}\right)^2 \right\}} \tag{10}$$

Pada Persamaan (10), SSB_{ij}merupakan jarak antara centroid Klaster $i$dan Klaster j. C_{ik}merupakan nilai centroid Klaster $i$pada atribut ke-k, sedangkan C_{jk}merupakan nilai centroid Klaster $j$pada atribut ke-k. Sementara itu, $n$menunjukkan jumlah atribut yang digunakan dalam perhitungan. Nilai SSB digunakan untuk melihat jarak atau tingkat pemisahan antara dua klaster berdasarkan posisi centroid masing-masing.

$$R_{\{ij\}} = \frac{SSW_i + SSW_j}{SSB_{\{ij\}}} \tag{11}$$

Pada Persamaan (11), R_{ij}merupakan nilai rasio antara Klaster $i$dan Klaster j, SSW_idan SSW_jmerupakan nilai rata-rata jarak internal dari masing-masing klaster, sedangkan SSB_{ij}merupakan jarak antara centroid kedua klaster. Nilai rasio tersebut digunakan untuk melihat hubungan antara kekompakan anggota klaster dan jarak antar-klaster. Nilai rasio yang lebih kecil menunjukkan kondisi klaster yang lebih baik karena jarak internal relatif lebih kecil dibandingkan dengan jarak antar-centroid. Nilai tersebut selanjutnya digunakan dalam perhitungan Davies-Bouldin Index.

3. HASIL DAN PEMBAHASAN

3.1 Normalisasi *Z-Score*

Normalisasi data menggunakan metode *Z-Score* dilakukan untuk menyetarakan skala antarvariabel sebelum proses klastering. Tahap ini diperlukan karena setiap variabel dapat memiliki satuan dan rentang nilai yang berbeda. Dengan melakukan standardisasi, setiap variabel ditransformasikan berdasarkan nilai rata-rata dan standar deviasi sehingga perbedaan skala antarvariabel tidak memberikan pengaruh yang terlalu besar dalam perhitungan jarak pada algoritma *K-Means.* Penggunaan *Standard Scaler* atau *Z-Score* sebelum proses *K-Means* juga telah diterapkan pada penelitian klastering berdasarkan indikator sosial ekonomi untuk memastikan perhitungan jarak antar data dapat dilakukan secara lebih tepat (Sinaga & Sipayung, 2026).

Selanjutnya, standar deviasi digunakan untuk mengetahui tingkat penyebaran data terhadap nilai rata-ratanya. Setelah proses standardisasi dilakukan, data hasil *Z-Score* digunakan sebagai input dalam proses perhitungan jarak pada algoritma *K-Means*, dapat dilihat pada Tabel 1.

Tabel 1. Perhitungan *Zscore* V1

Pro	Kab/Kota	Variabel	x	Mean(μ)	Deviasi(σ)	Zscore
Aceh	Simeulue	P0(V1)	18,98	10,6	7,6	1,10
		RLS(V2)	9,48	7,3	2,0	1,09
		PPK(V3)	Rp7.148	Rp10.325	Rp2.714	1,17
		IPM(V4)	66,41	69,9	6,5	-0,54

Tabel 1 adalah contoh penerapan normalisasi *ZScore* pada kab/kota Simeulue yang ada di provinsi Aceh. Dimana hasil dari V1 yakni presentase penduduk miskin (P0) menurut kab/kota, V2 yakni rata-rata lama sekolah penduduk 15+ (RLS), V3 yakni Pengeluaran per kapita, V4 yakni indeks Pembangunan manusia (IPM) dengan memunculkan nilai presentase, rata-rata, deviasi dan hasil perhitungan normalisasinya. Dengan cara ada pada rumus dengan nomer urutan (1)(2)(3).

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1653-1667
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10939

3.2 Analisis Variabel yang Tidak Digunakan

Variabel yang terdapat di dalam dataset utama sebelum diolah untuk proses normalisasi dan klastering tidak dapat digunakan semua. Hal ini disebabkan karena adanya persamaan nilai antar variable satu dengan yang lain, sehingga untuk menghindari adanya hasil yang bias maka di tetapkan eliminasi terhadap variable yang ada. Dengan mengeliminasi variabel tersebut, proses klastering menjadi lebih optimal dan menghasilkan kelompok yang lebih representatif, bisa dilihat pada Tabel 2 dan Tabel 3.

Tabel 2. Variabel yang digunakan

P0 (%)	RLS (th)	PPK (Rp)	IPM (rata)

Penjelasan untuk variable yang ada pada table 2 adalah variable 1 yaitu persentase penduduk miskin menurut kabupaten/kota (P0), variable 2 yaitu rata-rata lama sekolah penduduk 15+ (RLS), variable 3 yaitu Pengeluaran per kapita (PPK), variable 4 yaitu indeks Pembangunan manusia (IPM).

Tabel 3. Variabel yang tidak digunakan

UHH	PRTS	PRTA	TPT	TPAK	PDRB

Penjelasan untuk tabel 3 adalah tabel yang tidak digunakan karena memiliki kesamaan makna. usia harapan hidup (UHH) (th), Persentase rumah tangga yang memiliki akses terhadap sanitasi layak (PRTS), Persentase rumah tangga yang memiliki akses terhadap air minum layak (PRTA), tingkat pengangguran terbuka (TPT), tingkat partisipasi tenaga kerja (TPAK), produk domestik regional bruto (PDRB).

Dari table 2 dan 3 dapat dilihat bahwasanya terdapat persamaan nilai atau makna dari varibel yang tertera, sehingga dari total variabel yang terdapat pada data awal sebelum diolah dilakukan peleburan secara manual yakni dengan memilih diantara satu variable yang akan digunakan, dan untuk makna yang sama akan dileburkan atau tidak digunakan.

3.3 Hasil Klastering *K-Means*

a. Klastering dengan K=2
   Pada skenario K=2, data terbagi menjadi dua kelompok utama yaitu:
   1. Klaster dengan tingkat kemiskinan tinggi
   2. Klaster dengan tingkat kemiskinan rendah

   Hasil ini menunjukkan pemisahan yang cukup jelas antara wilayah dengan kondisi ekonomi yang berbeda secara signifikan, berikut adalah hasil *random centroid* dengan pembagian klaster pada K=2, dapat dilihat pada Tabel 4.

Tabel 4. Hasil random *centroid*

Hasil Random Centroid Awal					
Maluku Utara	Halmahera Selatan	-0,712	0,4	-1,173	-0,878
Sumatera Utara	Mandailing Natal	-0,146	0,665	-0,204	-0,417

Setelah melakukan proses normalisasi *ZScore*, selanjutnya pada Tabel 4 adalah proses penentuan *random centroid* yang dilakukan dengan skenario K=2, dimulai dengan melakukan random di kab/kota guna sebagai landasan awal penentuan *centroid* awal agar lebih netral dan hasilnya tidak bias. dengan maluku utara, Halmahera Selatan untuk C1/klaster 1 dan Sumatera utara, Mandailing Natal untuk C2/ klaster 2. Selanjutnya pada proses penentuan klaster dapat dilihat dari sampel data yang tertera pada Tabel 5.

Tabel 5. Proses penentuan klaster

Jenis Data	Provinsi	Kab/Kota	Zscore V1	Zscore V2	Zscore V3	Zscore V4
Data	Aceh	Simeulue	1,103	1,090	-Rp1,171	-0,537
C1	Maluku Utara	Halmahera Selatan	-0,711	0,400	-Rp1,173	-0,878
C2	Sumatera Utara	Mandailing Natal	-0,146	0,665	Rp-0,204	-0.417

Penentuan *centroid* awal dilakukan secara random dari data yang telah melalui proses standardisasi *Z-Score*. *Centroid* awal tersebut selanjutnya digunakan sebagai acuan dalam proses iterasi pertama untuk menghitung jarak setiap kabupaten/kota terhadap masing-masing *centroid.* Perhitungan jarak dilakukan dengan menggunakan *Euclidean Distance* berdasarkan nilai *Z-Score* pada seluruh variabel yang digunakan. Menurut tabel 5 data yang dipakai adalah salah satu data yang sudah melalui standarisasi *Z-Score* sebagai sampel. Lalu untuk perhitungannya dilakukan antara data dengan nilai dari C1, C2, dan C3. Jadi nanti akan muncul hasil perhitungannya dan akan teridentifikasi masuk ke dalam klaster mana kab/kota yang sudah dilakukan proses penentuan klaster. Setelah itu setiap kabupaten/kota kemudian ditempatkan ke dalam klaster yang memiliki jarak paling kecil terhadap centroid.begitu seterusnya di tiap kolomnya sampai dengan jumlah data yang tersedia (Ramadhana & Jambak, 2024) yakni 514 data kab/kota yang dilakukan pada perhitungan C1 dan C2 dengan rumus yang digunakan pada nomer urut (4).

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1653-1667
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10939

Tabel 6. Menentukan *centroid* baru

Klaster	Keterangan	Jumlah	V1	V2	V3	V4	Fungsi objektif
C1	Jumlah anggota						
	Jumlah X	52	73,939	38,385	71,873	83,287	
	New C1		1,422	0,738	1,382	1,601	
C2	Jumlah anggota						
	Jumlah X	462	287,503	355,445	314,963	279,026	
	New C2		0,622	0,769	0,681	0,604	

Proses selanjutnya adalah menghitung kab/kota mana yang masuk kedalam C1 maupun C2. Hal itu digunakan untuk mendapatkan centroid kedua yang dimana digunakan untuk acuan perhitungan I2 / iterasi 2 yang dilakukan sama dengan cara perhitungan pada I1 / iterasi 1 sampai dengan iterasi yang telah ditentukan yakni iterasi 10 atau hasil iterasi yang tidak berubah nilainya dari hasil iterasi sebelumnya. Dengan cara dari tabel 6 terdapat kolom jumlah x adalah jumlah keseluruhan *Zscore* dari kab/kota yang masuk di keanggotaan Cn. *New* C1 adalah hasil dari pembagian dari kolom jumlah x dengan Jumlah seluruh kab/kota yang terdapat di Cn yakni pada kolom jumlah anggota Cn, lalu dari hasil pembagian tersebut di temukan nilai pada kolom *New* Cn yang akan digunakan sebagai centroid kedua untuk perhitungan I2. rumus yang terdapat di nomer urut (5)(6).

Tabel 8. Fungsi objektif iterasi 1

Fungsi Objektif Awal	1000
Fungsi Objektif J	561,66
Fungsi Objektif	438,342

Jumlah fungsi objek C1 dan C2 itu digunakan untuk mengetahui nilai dari hasil perhitungan pada tiap iterasi ke-n dengan dijumlahkan dari fungsi objek C1 dan C2sehingga menjadi satu fungsi lalu hasil itu di kurangi dengan fungsi objektif awal yakni 1000 dengan rumus *absolute* agar nilainya yang muncul hanya positif. Lalu dari hasil pengurangan tersebut akan menghasilkan fungsi objektif yang dimana nilai tersebut adalah relatif tergantung nilai yang terdapat dari masing-masing iterasi yang akan digunakan sebagai acuan pengurangan untuk iterasi sesudahnya dan begitu seterusnya dengan nantinya menempati posisi yang saat ini di tempati oleh posisi fungsi objektif awal. Penerapannya dapat dilihat pada tabel 8.

b. Klastering dengan K=3

Pada skenario K=3, data terbagi menjadi tiga kelompok yaitu:

1. Klaster tingkat kemiskinan tinggi
2. Klaster tingkat kemiskinan sedang
3. Klaster tingkat kemiskinan rendah

Pembagian ini memberikan informasi yang lebih detail dibandingkan K=2 karena mampu menunjukkan tingkat variasi yang lebih spesifik, berikut adalah hasil dari *random centroid* dengan pembagian klaster pada K=3. Dapat dilihat pada Tabel 9.

Tabel 9. Hasil *random centroid*

Hasil Random Centroid Awal					
Papua	Deiyai	3,946	-2,025	-Rp2,083	-3,068
Kep. Bangka Belitung	Bangka Selatan	-0,909	-0,295	Rp0,532	-0,437
Papua	Keerom	0,711	0,360	-Rp0,515	-0,525

Setelah melakukan proses normalisasi *ZScore*, selanjutnya pada tabel 9 adalah proses penentuan *random centroid* yang dilakukan dengan skenario K=2, dimulai dengan melakukan random di kab/kota guna sebagai landasan awal penentuan *centroid* awal agar lebih netral dan hasilnya tidak bias. dengan maluku utara, Halmahera Selatan untuk C1/klaster 1 dan Sumatera utara, Mandailing Natal untuk C2/ klaster 2. Selanjutnya pada proses penentuan klaster dapat dilihat dari sampel data yang tertera pada Tabel 10.

Tabel 10. Proses penentuan klaster

Jenis Data	Prov.	Kab/Kota	Zscore V1	Zscore V2	Zscore V3	Zscore V4
Data	Aceh	Simeulue	1,103	1,090	-Rp1,171	-0,537
C1	Papua	Deiyai	3,946	-2,025	-Rp2,083	-3,068
C2	Kep. Bangka Belitung	Bangka Selatan	-0,909	-0,295	Rp0,532	-0,437
C3	Papua	Keerom	0,711	0,360	-Rp0,515	-0,525

Penentuan *centroid* awal dilakukan secara *random* dari data yang telah melalui proses standardisasi *Z-Score*. *Centroid* awal tersebut selanjutnya digunakan sebagai acuan dalam proses iterasi pertama untuk menghitung jarak setiap kabupaten/kota terhadap masing-masing *centroid*. Perhitungan jarak dilakukan dengan menggunakan *Euclidean Distance* berdasarkan nilai *Z-Score* pada seluruh variabel yang digunakan. Menurut tabel 10 data yang

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1653-1667
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10939

dipakai adalah salah satu data yang sudah melalui standarisasi *Z-Score* sebagai sampel. Lalu untuk perhitungannya dilakukan antara data dengan nilai dari C1, C2, dan C3. Jadi nanti akan muncul hasil perhitungannya dan akan teridentifikasi masuk ke dalam klaster mana kab/kota yang sudah dilakukan proses penentuan klaster. Setelah itu setiap kabupaten/kota kemudian ditempatkan ke dalam klaster yang memiliki jarak paling kecil terhadap centroid.begitu seterusnya di tiap kolomnya sampai dengan jumlah data yang tersedia (Ramadhana & Jambak, 2024) yakni 514 data kab/kota dengan rumus nomer urut (4).

Tabel 11. Menentukan *centroid* baru

Klaster	Keterangan	Jumlah	V1	V2	V3	V4	Fungsi Objektif
	Jumlah Anggota	21					
C1	Jumlah X		65,655	31,025	39,531	60,157	
	New C1		3,126	1,477	1,882	2,865	25,246
	Jumlah Anggota	197					
C2	Jumlah X		107,038	188,675	174,262	152,126	
	New C2		0,543	0,958	0,885	0,772	226,311
	Jumlah Anggota	296					
C3	Jumlah X		188,749	174,130	173,043	150,031	
	New C3		0,958	0,884	0,878	0,762	308,885

Proses selanjutnya dapat dilihat pada tabel 11 dengan menghitung kab/kota mana yang masuk kedalam C1, C2, atau C3. Hal itu digunakan untuk mendapatkan *centroid* kedua yang dimana digunakan untuk acuan perhitungan I2 / iterasi 2 yang dilakukan sama dengan cara perhitungan pada I1 / iterasi 1 sampai dengan iterasi yang telah ditentukan yakni iterasi 10 atau hasil iterasi yang tidak berubah nilainya dari hasil iterasi sebelumnya. Dengan cara dari kolom jumlah X yakni jumlah keseluruhan nilai dari kab/kota sesuai dengan variabelnya. Lalu pada kolom *New* Cn adalah hasil dari pembagian dari kolom X dengan Jumlah seluruh kab/kota yang terdapat di kolom jumlah anggota, lalu dari hasil pembagian tersebut di temukan nilai pada kolom *New* Cn yang akan digunakan sebagai centroid kedua untuk perhitungan I2. rumus yang terdapat di nomer urut (5)(6)(7).

Tabel 12. Fungsi objektif iterasi 1

Fungsi Objektif Awal	1000
Fungsi Objektif J	560,440
Fungsi Objektif	439,558

Jumlah fungsi objek C1, objek C2, dan objek C3 itu digunakan untuk mengetahui nilai dari hasil perhitungan pada tiap iterasi ke-n dengan dijumlahkan dari fungsi objek C1, C2, dan C3 sehingga menjadi satu fungsi lalu hasil itu di kurangi dengan fungsi objektif awal yakni 1000 dengan rumus *absolute* agar nilainya yang muncul hanya positif. Lalu dari hasil pengurangan tersebut akan menghasilkan fungsi objektif yang dimana nilai tersebut adalah relatif tergantung nilai yang terdapat dari masing-masing iterasi ke-n yang akan digunakan sebagai acuan pengurangan untuk iterasi sesudahnya dan begitu seterusnya sampai dengan maksimum iterasi yang ditentukan atau sampai nilai dari iterasi sebelumnya tidak berubah. Dapat dilihat penerapannya pada Tabel 12.

3.4 Evaluasi *Davies-Bouldin Index (DBI)*

Evaluasi menggunakan *DBI* dilakukan untuk menentukan kualitas hasil klastering. Interpretasi nilai *DBI* adalah sebagai berikut (Umagapi, et al., 2023):

a. Nilai DBI kecil maka klaster lebih baik
b. Nilai DBI besar maka klaster kurang optimal

Berdasarkan hasil perhitungan:

Tabel 13. Perhitungan *DBI* untuk K=2

Jumlah C1	71	Jumlah Jarak C1	63,414
Jumlah C2	443	Jumlah Jarak C2	417,127
Total	514		

Pada Tabel 13 terdapat kolom yang menunjukan jumlah keseluruhan dari data dan juga menunjukan masuk ke dalam klaster mana dari masing-masing kab/kota. Sementara itu didapatkan juga total fungsi objektif dari masing-masing klaster yang ditunjukan pada kolom jumlah jarak Cn.

Tabel 14. Perhitungan SSW

SSW 1	0,89
SSW 2	0,94

Sum of Squares Within (SSW) untuk k=2 seperti pada tabel 14, digunakan untuk mengukur tingkat penyebaran atau variasi data dalam klaster terhadap *centroidnya*. Nilai *SSW* yang lebih kecil menunjukkan bahwa anggota klaster

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1653-1667
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10939

memiliki jarak yang lebih dekat terhadap *centroid* sehingga klaster cenderung lebih kompak (Lutfiah et al., 2025). Dengan rumus di nomer (9).

Tabel 15. Perhitungan *SSB*

		Data ke i	
SSB		1	2
Data ke i	1	0,00	1,95
	2	1,95	0,00

Sum of Squares Between (SSB) untuk k=2 seperti pada tabel 15, merupakan ukuran yang digunakan untuk mengetahui tingkat separasi atau jarak antar klaster dengan menghitung jarak *Euclidean* antara *centroid* masing-masing klister (Mardhiatul Azmi et al., 2023). Pada penelitian ini diperoleh nilai *SSB* antara klaster 1 dan klaster 2 sebesar 1,95, sedangkan nilai diagonal bernilai 0 karena merupakan jarak *centroid* terhadap dirinya sendiri. Nilai *SSB* yang semakin besar menunjukkan bahwa jarak antar klaster semakin jauh sehingga pemisahan klaster menjadi semakin baik. Hal ini sesuai dengan konsep evaluasi *Davies–Bouldin Index*, yang mengutamakan klaster yang kompak dan memiliki separasi yang tinggi. Dengan rumusnya nomer (10).

Tabel 16. Perhitungan *DBI*

	R	Data ke i		Rmax	DBI
		1	2		
Data ke i	1	0,00	0,94	0,94	0,9417
	2	0,94	0,00	0,94	
Total klaster	2			1,88	

Tabel nilai R yang terdapat pada tabel 16, menunjukkan hasil perhitungan rasio antara penyebaran dalam klaster *SSW (within-klaster scatter)* dan jarak antar centroid *SSB(between-klaster separation*) untuk setiap pasangan klaster. Nilai R dihitung menggunakan rumus nomer urut (11). Selanjutnya, dari setiap baris dipilih nilai terbesar (Rmax) karena nilai tersebut menunjukkan klaster lain yang memiliki tingkat kemiripan paling tinggi terhadap klaster yang sedang dievaluasi. Berdasarkan hasil perhitungan diperoleh nilai Rmax untuk klaster 1 sebesar 0,94 dan klaster 2 sebesar 0,94. Nilai *Davies–Bouldin Index (DBI)* kemudian dihitung dengan merata-ratakan seluruh nilai Rmax, sehingga diperoleh nilai *DBI* sebesar 0,9417. Nilai *DBI* yang semakin kecil menunjukkan bahwa klaster yang terbentuk semakin kompak dan memiliki pemisahan yang semakin baik (Idrus et al., 2022).

Tabel 17. Perhitungan *DBI* untuk K=3

Jumlah C1	0	Jumlah Jarak C1	0
Jumlah C2	469	Jumlah Jarak C2	439,142
Jumlah C3	45	Jumlah Jarak C3	125,744
Total	514		

Pada tabel 17 terdapat kolom yang menunjukan jumlah keseluruhan dari data dan juga menunjukan masuk ke dalam klaster mana dari masing-masing kab/kota. Sementara itu didapatkan juga total fungsi objektif dari masing-masing klaster yang ditunjukan pada kolom jumlah jarak Cn.

Tabel 18. Perhitungan *SSW*

SSW 1	0,00
SSW 2	0,94
SSW 3	2,79

Sum of Squares Within (SSW) untuk k=3 seperti pada tabel 18, digunakan untuk mengukur tingkat penyebaran atau variasi data dalam klaster terhadap *centroid*nya. Nilai *SSW* yang lebih kecil menunjukkan bahwa anggota klaster memiliki jarak yang lebih dekat terhadap *centroid* sehingga klaster cenderung lebih kompak (Lutfiah et al., 2025).dengan rumus di nomer (9) dapat diperoleh nilai dari masing-masing *sum of squares within* dari tiap klaster dengan perolehan *SSW* 1 sebesar 0,00, *SSW* 2 sebesar 0,94, dan *SSW* 3 sebesar 2,79.

Tabel 19. Perhitungan *SSB*

		Data ke i		
SSB		1	2	3
Data ke i	1	0,00	3,73	4,52
	2	3,73	0,00	0,99
	3	4,52	0,99	0,00

Sum of Squares Between (SSB) untuk k=3 seperti pada tabel 19, merupakan ukuran yang digunakan untuk mengetahui tingkat separasi atau jarak antar klaster dengan menghitung jarak *Euclidean* antara *centroid* masing-masing

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1653-1667
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10939

klister (Mardhiatul Azmi et al., 2023). Berdasarkan hasil perhitungan diperoleh jarak *centroid* antara klaster 1 dan klaster 2 sebesar 3,73, antara klaster 1 dan klaster 3 sebesar 4,52, serta antara klaster 2 dan klaster 3 sebesar 0,99. Nilai terbesar terdapat pada pasangan klaster 1 dan klaster 3, sehingga kedua klaster tersebut memiliki tingkat pemisahan yang paling baik. Sebaliknya, pasangan klaster 2 dan klaster 3 memiliki jarak *centroid* terkecil sehingga keduanya merupakan klaster yang paling berdekatan. Dalam evaluasi menggunakan *Davies–Bouldin Index*, semakin besar jarak antar *centroid (separation*), maka semakin baik kualitas pemisahan klaster yang terbentuk. Hal ini sesuai dengan konsep evaluasi *Davies–Bouldin Index*, yang mengutamakan klaster yang kompak dan memiliki separasi yang tinggi. Dengan rumusnya nomer (10).

Tabel 20. Perhitungan *DBI*

	R	Data ke i			Rmax	DBI
		1	2	3		
Data ke i	1	0,00	0,25	0,62	0,62	2,706
	2	0,25	0,00	3,75	3,75	
	3	0,62	3,75	0,00	3,75	
Total klaster adalah 3					8,12	

Tabel nilai R yang terdapat pada Tabel 20, menunjukkan hasil perhitungan rasio antara penyebaran dalam klaster *SSW (within-klaster scatter)* dan jarak antar *centroid SSB (between-klaster separation)* untuk setiap pasangan klaster. Nilai R dihitung menggunakan rumus (11). Selanjutnya, dari setiap baris dipilih nilai terbesar (Rmax) yang menunjukkan klaster lain dengan tingkat kemiripan tertinggi terhadap klaster yang sedang dievaluasi. Berdasarkan hasil perhitungan diperoleh nilai Rmax untuk klaster 1 sebesar 0,62, klaster 2 sebesar 3,75, dan klaster 3 sebesar 3,75. Jumlah seluruh nilai Rmax adalah 8,12, sehingga nilai *Davies–Bouldin Index (DBI)* dihitung sebesar 2,706. Nilai *DBI* yang relatif besar dipengaruhi oleh tingginya nilai dan , yang menunjukkan bahwa klaster 2 dan klaster 3 memiliki pemisahan yang kurang baik dibandingkan pasangan klaster lainnya. Nilai *DBI* yang semakin kecil menunjukkan bahwa klaster yang terbentuk semakin kompak dan memiliki pemisahan yang semakin baik (Idrus et al., 2022).

3.5 Interpretasi Hasil

Evaluasi kualitas hasil klastering dilakukan menggunakan *Davies–Bouldin Index (DBI)* dengan membandingkan dua skenario jumlah klaster, yaitu K=2 dan K=3. *DBI* dipilih karena mampu mengevaluasi kualitas klaster berdasarkan dua karakteristik utama, yaitu kohesi *(within-klaster scatter)* dan separasi *(between-klaster separation).* Nilai *DBI* yang semakin kecil menunjukkan bahwa klaster yang dihasilkan memiliki tingkat kekompakan yang tinggi dan pemisahan antar klaster yang semakin baik.Tabel 8 Pembagian klaster berdasarkan K=2 kemiskinan rendah.

3.5.1 Interpretasi Hasil Klastering K=2

Berdasarkan hasil evaluasi diperoleh nilai *DBI* sebesar 0,9417 pada skenario K=2. Nilai tersebut diperoleh dari rata-rata nilai maksimum rasio kemiripan antar klaster (Rmax) sebesar 0,94 untuk masing-masing klaster. Nilai *SSW* pada kedua klaster masing-masing sebesar 0,89 dan 0,94, yang menunjukkan bahwa penyebaran data terhadap centroid relatif kecil sehingga anggota pada setiap klaster memiliki tingkat kemiripan yang tinggi. Selain itu, nilai *SSB* sebesar 1,95 menunjukkan bahwa centroid kedua klaster memiliki jarak yang cukup baik sehingga kedua klaster dapat dipisahkan secara jelas. Kombinasi antara nilai *SSW* yang rendah dan *SSB* yang cukup besar menghasilkan nilai R sebesar 0,94, sehingga rata-rata *DBI* yang diperoleh juga relatif kecil. Hal ini menunjukkan bahwa pembagian data menjadi dua kelompok menghasilkan klaster yang cukup kompak dengan tingkat pemisahan yang baik.

3.5.2 Interpretasi Hasil Klastering K=3

Pada skenario K=3 diperoleh nilai *DBI* sebesar 2,706, yang jauh lebih besar dibandingkan skenario K=2. Nilai tersebut diperoleh dari rata-rata Rmax sebesar 0,62, 3,75, dan 3,75 dengan jumlah keseluruhan sebesar 8,12. Hasil perhitungan menunjukkan bahwa klaster 1 memiliki nilai *SSW* sebesar 0,00, sedangkan klaster 2 dan klaster 3 masing-masing sebesar 0,94 dan 2,79. Nilai *SSW* pada klaster 3 merupakan yang terbesar sehingga menunjukkan bahwa penyebaran data pada klaster tersebut lebih tinggi dibandingkan klaster lainnya. Pada sisi separasi, nilai *SSB* menunjukkan bahwa jarak antara klaster 2 dan klaster 3 hanya sebesar 0,99, sedangkan jarak antara klaster 1 dengan klaster 2 sebesar 3,73 dan klaster 1 dengan klaster 3 sebesar 4,52. Hal ini menunjukkan bahwa klaster 2 dan klaster 3 memiliki *centroid* yang saling berdekatan sehingga tingkat pemisahannya relatif rendah. Akibat jarak antar *centroid* yang kecil tersebut, diperoleh nilai R23 dan R32 sebesar 3,75, yang merupakan nilai terbesar dibandingkan pasangan klaster lainnya. Nilai R yang tinggi mengindikasikan bahwa penyebaran data dalam kedua klaster relatif besar dibandingkan dengan jarak antar *centroid*, sehingga kedua klaster memiliki tingkat kemiripan yang tinggi dan sulit dibedakan secara jelas. Kondisi tersebut menyebabkan nilai rata-rata *DBI* meningkat menjadi 2,706.

3.5.3 Analisis Perbandingan K=2 dan K=3

Perbandingan hasil evaluasi menunjukkan adanya perbedaan kualitas klastering yang cukup signifikan. Berdasarkan hasil tersebut, pembagian data menjadi dua klaster menghasilkan kualitas klastering yang lebih baik dibandingkan pembagian menjadi tiga klaster. Dapat dilihat pada tabel 21, dalam hasil nilai *DBI* k=2 memiliki nilai yang kecil atau

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1653-1667
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10939

mendekati 0 sehingga dalam kualitasnya dapat dilihat dalam kolom interprestasinya, yang menunjukkan bahwa data dalam masing-masing klaster lebih homogen dan jarak antar klaster lebih terpisah.

Tabel 21. Perbandingan konsep K=2 dan K=3

Jumlah Klaster	Nilai DBI	Interpretasi
K=2	0,942	Klaster lebih kompak dan memiliki separasi yang baik sehingga hasilnya akan lebih akurat.
K=3	2,706	Klaster kurang optimal karena terdapat dua klaster yang memiliki jarak centroid sangat dekat sehingga kemungkinan hasil yang bias akan semakin besar.

Berikut adalah Tabel 22 dan Tabel 23 hasil pembagian klaster Sebagian data menggunakan skenario K=2 dengan kategori kemiskinan rendah dan tinggi:

Tabel 22. Pembagian klaster berdasarkan K=2 kemiskinan rendah

Provinsi	Kab/Kota	Klaster
Aceh	Simeulue	Kemiskinan Rendah
Aceh	Aceh Singkil	Kemiskinan Rendah
Aceh	Aceh Selatan	Kemiskinan Rendah
Aceh	Aceh Tenggara	Kemiskinan Rendah

Tabel 23. Pembagian klaster berdasarkan K=2 kemiskinan tinggi.

Provinsi	Kab/Kota	Klaster
Aceh	Kota Banda Aceh	Kemiskinan Tinggi
Aceh	Kota Langsa	Kemiskinan Tinggi
Sumatera Utara	Kota Lhokseumawe	Kemiskinan Tinggi
Sumatera Utara	Kota Pematang Siantar	Kemiskinan Tinggi

3.6 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma K-Means mampu mengelompokkan 514 kabupaten/kota di Indonesia berdasarkan karakteristik indikator kemiskinan. Berdasarkan hasil evaluasi menggunakan Davies-Bouldin Index (DBI), skenario K=2 menghasilkan nilai DBI sebesar 0,9417, sedangkan skenario K=3 menghasilkan nilai DBI sebesar 2,706. Nilai DBI yang lebih rendah pada K=2 menunjukkan bahwa pembentukan dua klaster memberikan tingkat kekompakan dan pemisahan klaster yang lebih baik dibandingkan dengan K=3.

Temuan penelitian ini sejalan dengan penelitian Nanda Shalsadilla et al. (2023) yang menggunakan Davies-Bouldin Index sebagai dasar untuk menentukan jumlah klaster yang optimal. Penelitian tersebut menunjukkan bahwa nilai DBI dapat digunakan untuk membandingkan kualitas beberapa skenario klaster, di mana nilai DBI yang lebih kecil menunjukkan hasil pengelompokan yang lebih baik. Prinsip tersebut juga diterapkan dalam penelitian ini melalui perbandingan hasil K-Means pada K=2 dan K=3.

Selain itu, penggunaan normalisasi Z-Score dalam penelitian ini memiliki kesamaan dengan penelitian Sinaga dan Sipayung (2026). Normalisasi dilakukan untuk menyetarakan skala antarvariabel sehingga perbedaan satuan tidak menyebabkan variabel tertentu lebih dominan dalam perhitungan jarak K-Means. Dengan demikian, penggunaan Z-Score dalam penelitian ini mendukung proses pengelompokan agar setiap indikator dapat memberikan kontribusi yang lebih seimbang terhadap pembentukan klaster.

Penelitian ini juga memiliki kesamaan dengan penelitian Tawakal et al. (2025) yang menerapkan DBI untuk mengevaluasi hasil K-Means pada data kemiskinan kabupaten/kota di Jawa Tengah. Namun, terdapat perbedaan pada cakupan wilayah penelitian. Penelitian terdahulu tersebut berfokus pada wilayah Jawa Tengah, sedangkan penelitian ini memperluas cakupan analisis hingga 514 kabupaten/kota di seluruh Indonesia. Perluasan cakupan tersebut memungkinkan karakteristik kemiskinan antardaerah di Indonesia dianalisis secara lebih luas.

Dibandingkan dengan penelitian Sinaga dan Sipayung (2026) yang menggunakan provinsi sebagai unit analisis, penelitian ini menggunakan kabupaten/kota sebagai unit analisis. Penggunaan unit analisis yang lebih terperinci memungkinkan perbedaan karakteristik kemiskinan antarwilayah dapat terlihat dengan lebih spesifik. Hal ini menjadi salah satu pengembangan dari penelitian terdahulu yang telah dilakukan.

Perbedaan lainnya terdapat pada penggabungan beberapa tahapan analisis. Penelitian ini tidak hanya menerapkan algoritma K-Means, tetapi juga melakukan seleksi indikator, normalisasi Z-Score, pengujian beberapa skenario jumlah klaster, serta evaluasi menggunakan Davies-Bouldin Index. Integrasi tahapan tersebut dilakukan untuk memperoleh hasil pengelompokan yang lebih terukur dan mengurangi kemungkinan penggunaan variabel yang tidak relevan atau memiliki informasi yang berulang.

Dengan demikian, hasil penelitian ini memperkuat temuan penelitian terdahulu bahwa K-Means dapat digunakan untuk melakukan pengelompokan wilayah berdasarkan karakteristik kemiskinan, sedangkan Davies-Bouldin Index dapat digunakan untuk menentukan kualitas dan jumlah klaster yang lebih optimal. Kontribusi penelitian ini terletak pada penerapan pendekatan tersebut pada 514 kabupaten/kota di Indonesia dengan mengintegrasikan seleksi indikator, normalisasi Z-Score, K-Means, dan evaluasi DBI. Hasil penelitian menunjukkan bahwa K=2 merupakan skenario yang

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1653-1667
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10939

lebih optimal karena menghasilkan nilai DBI sebesar 0,9417, lebih rendah dibandingkan K=3 yang menghasilkan nilai 2,706.

4. KESIMPULAN

Berdasarkan hasil evaluasi menggunakan *Davies–Bouldin Index*, dapat disimpulkan bahwa skenario K=2 merupakan jumlah klaster yang paling optimal untuk mengelompokkan data kemiskinan pada penelitian ini. Hal tersebut ditunjukkan oleh nilai *DBI* sebesar 0,9417, yang lebih kecil dibandingkan nilai *DBI* pada skenario K=3 sebesar 2,706. Nilai ini diperoleh berkat kombinasi tingkat penyebaran internal klaster (SSW) yang rendah serta jarak antar-centroid klaster (SSB) sebesar 1,95, yang menghasilkan separasi wilayah secara tegas antara kelompok kemiskinan tinggi dan rendah. Sebaliknya, pada skenario K=3 terjadi tumpang-tindih antara klaster 2 dan klaster 3 akibat jarak centroid yang terlampau dekat (0,99), yang memicu tingginya rasio kemiripan (Rmax 3,75) dan menurunkan kualitas pengelompokan. Dengan demikian, model K-Means dengan K=2 mampu merepresentasikan karakteristik data kemiskinan secara lebih efektif dan layak digunakan sebagai dasar dalam proses analisis maupun penyusunan rekomendasi kebijakan penanggulangan kemiskinan. Dengan demikian, model K-Means dengan K=2 mampu merepresentasikan karakteristik data kemiskinan secara lebih efektif dan layak digunakan sebagai dasar dalam proses analisis maupun penyusunan rekomendasi kebijakan penanggulangan kemiskinan. kontribusi penelitian tidak hanya terletak pada pembentukan kelompok wilayah, tetapi juga pada penerapan tahapan pengolahan dan evaluasi yang memungkinkan pemilihan skenario klaster berdasarkan ukuran kualitas yang terukur. Hasil tersebut dapat menjadi informasi pendukung dalam memahami kemiripan karakteristik kemiskinan antarwilayah dan membantu melihat kelompok wilayah yang memiliki kondisi indikator yang relatif serupa. Meski begitu, penelitian ini masih terbatas pada pengujian dua dan tiga klaster saja dengan indikator yang sudah disederhanakan. Untuk pengembangan selanjutnya bisa dengan menggunakan jumlah klaster yang lebih bervariasi, menambah indikator pendukung lainnya, atau membandingkan K-Means dengan metode pengelompokan lain agar gambaran pemetaan kemiskinan bisa semakin lengkap.

REFERENCES

Alif, Moh. F., & Fahmi, A. (2026). Klasterisasi Wilayah Kemiskinan Jawa Tengah Menggunakan K-Means Berbasis Indikator Sosial-Ekonomi. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *11*(3), 302–309. https://doi.org/10.25077/TEKNOSI.v11i3.

AlShammari, A. F. (2024). Implementation of Clustering using K-Means in Python. *International Journal of Computer Applications*, *186*(40), 12–17. https://doi.org/10.5120/ijca2024923990

Badan Pusat Statistik Provinsi Jawa Tengah. (2023). *Kemiskinan Provinsi Jawa Tengah September 2022*.

Badan Pusat Statistik Provinsi Jawa Tengah. (2023). *Provinsi Jawa Tengah dalam angka 2023*.

Fuady, M., Rafi, M., Fuady, F., & Aulia, D. F. (2022). Kemiskinan Multi Dimensi dan Indeks Pembangunan Manusia di Indonesia. *TATALOKA*, *24*(4), 330–337. https://doi.org/10.14710/TATALOKA.24.4.

Idrus, A., Tarihoran, N., Supriatna, U., Tohir, A., Suwarni, S., & Rahim, R. (2022). Distance Analysis Measuring for Clustering using K-Means and Davies Bouldin Index Algorithm. *TEM Journal*, 1871–1876. https://doi.org/10.18421/TEM114-55

Kause, J., & Fithriyah, F. (2024). Analisis Determinan Kemiskinan Multidimensi di Indonesia. *Jurnal Riset Ilmu Ekonomi*, *4*(2), 115–127. https://doi.org/10.23969/JRIE.V4I2.98

Lutfiah, F., Risfa, F., Muhammad Kevyn, R., Etis, S., & Ukasyah, A. (2025). Pendekatan Data Mining dalam Optimalisasi Margin Penjualan Adidas: Studi Klasterisasi dengan K-Means dan Fuzzy C-Means. *Euler: Jurnal Ilmiah Matematika, Sains Dan Teknologi*, *13*(2), 229–237. https://doi.org/10.37905/euler.v13i2.32417

Mardhiatul Azmi, Atus Amadi Putra, Dodi Vionanda, & Admi Salma. (2023). Comparison of the Performance of the K-Means and K-Medoids Algorithms in Grouping Regencies/Cities in Sumatera Based on Poverty Indicators. *UNP Journal of Statistics and Data Science*, *1*(2), 59–66. https://doi.org/10.24036/ujsds/vol1-iss2/25

Mayasari, S. N., & Nugraha, J. (2023). Implementasi K-Means cluster analysis untuk mengelompokkan kabupaten/kota berdasarkan data kemiskinan di Provinsi Jawa Tengah tahun 2022. *KONSTELASI: Konvergensi Teknologi Dan Sistem Informasi*, *3*(2), 317–329.

Nanda Shalsadilla, Shantika Martha, & Hendra Perdana. (2023). Penentuan Jumlah Cluster Optimum Menggunakan Davies Bouldin Index dalam Pengelompokan Wilayah Kemiskinan di Indonesia. *STATISTIKA Journal of Theoretical Statistics and Its Applications*, *23*(1), 63–72. https://doi.org/10.29313/statistika.v23i1.1743

Pangestu, M. S., & Fitriani, M. A. (2022). Perbandingan Perhitungan Jarak Euclidean Distance, Manhattan Distance, dan Cosine Similarity dalam Pengelompokan Data Bibit Padi Menggunakan Algoritma K-Means. *Sainteks*, *19*(2), 141. https://doi.org/10.30595/sainteks.v19i2.14495

Ramadhana, Y., & Jambak, M. I. (2024). Influence of Optimization of the k-Means Algorithm with Genetic Algorithm on the Results of High Dimension Data Clustering. *Indonesian Journal of Computer Science*, *13*(1). https://doi.org/10.33022/ijcs.v13i1.3634

Sabilillah, J., & Robertus, M. (2024). M. (2024). Pengaruh urbanisasi, modal manusia, dan pengangguran terhadap kemiskinan perkotaan di enam kota Provinsi Jawa Tengah tahun 2015–2021. *Diponegoro Journal of Economics*, *13*(3), 43–53.

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1653-1667
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10939

Sari, D. T., Khusna, N. I., & Wulandari, F. (2023). Analisis Tingkat Kemiskinan Di Provinsi Jawa Tengah: Suatu Kajian Berdasarkan Faktor Pendidikan, Sosial, Ekonomi, Lokasi Dan Indeks Pembangunan Manusia. *Jurnal PIPSI (Jurnal Pendidikan IPS Indonesia)*, *8*(1), 37–37. https://doi.org/10.26737/jpipsi.v8i1.3978

Setiaji, G. G., Gunata, K. P., & Setiarso, G. (2025). Optimasi Clustering K-Means Menggunakan Algoritma Genetika Dengan Data View Dan Like Di Tiktok. *Jurnal Transformatika*, *22*(2), 115–120. https://doi.org/10.26623/y2tedy77

Sinaga, C., & Sipayung, S. P. (2026). Implementasi K-Means clustering dengan Davies-Bouldin Index evaluation terhadap pengelompokan provinsi di Indonesia berdasarkan indikator sosioekonomi. *KAKIFIKOM*, *7*(2), 109–117.

Syahrul Adi Saputra, Galet Guntoro Setiaji, & Ahmad Rifa'i. (2026). Implementasi Sistem Informasi Kluster Penjualan Beras Menggunakan Algoritma K-Means. *Arcitech: Journal of Computer Science and Artificial Intelligence*, *6*(1), 145–162. https://doi.org/10.29240/arcitech.v6i1.17056

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

Umagapi, I. T., Umaternate, B., Hazriani, & Yuyun. (2023). Uji Kinerja K-Means Clustering Menggunakan Davies-Bouldin Index Pada Pengelompokan Data Prestasi Siswa. *Prosiding SISFOTEK*, *7*(1), 303–308.

Wongoutong, C. (2024). The impact of neglecting feature scaling in k-means clustering. *PLOS ONE*, *19*(12), e0310839. https://doi.org/10.1371/journal.pone.0310839