

Pengelompokan Kabupaten/Kota di Provinsi Banten Menggunakan K-Means Clustering Berbasis Indikator Pembangunan Multidimensi untuk Penentuan Prioritas Program Daerah

Clustering of Regencies and Cities in Banten Province Using K-Means Clustering Based on Multidimensional Development Indicators for Regional Program Priority Setting

Widyawati^{a,1,*}, Muhamad Oki Astrabuwono^{a,2}

^a*Sistem Informasi, Universitas Pamulang, Serang, Indonesia*
^b*Teknik Industri, Universitas Sultan Ageng Tirtayasa, Serang, Indonesia*
¹*dosen03354@unpam.ac.id; ²oki.astrabuwono@gmail.com*
**corresponding author*

Informasi Artikel	ABSTRAK
<p>Diserahkan : 16 Maret 2026 Diterima : 1 Mei 2026 Direvisi : 15 Mei 2026 Diterbitkan : 15 Mei 2026</p> <p>Kata Kunci: K-Means Clustering Indikator Pembangunan Daerah Pengelompokan Wilayah PDRB ADHK Provinsi Banten</p>	<p>Penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Provinsi Banten berdasarkan indikator pembangunan multidimensi, menganalisis karakteristik setiap cluster, serta memberikan rekomendasi awal mengenai prioritas program daerah. Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan data sekunder yang bersumber dari publikasi resmi Badan Pusat Statistik Provinsi Banten tahun 2024. Objek penelitian mencakup 8 kabupaten/kota di Provinsi Banten dengan enam variabel, yaitu garis kemiskinan, tingkat pengangguran terbuka, pertumbuhan ekonomi, produk domestik regional bruto atas dasar harga konstan (PDRB ADHK), rata-rata lama sekolah, dan umur harapan hidup. Data dianalisis melalui tahapan normalisasi Min-Max dan implementasi algoritma K-Means Clustering menggunakan Python. Hasil penelitian menunjukkan bahwa kabupaten/kota di Provinsi Banten terbagi ke dalam tiga cluster, yaitu wilayah prioritas intervensi, wilayah berkembang, dan wilayah maju. Hasil evaluasi menunjukkan nilai Silhouette Score sebesar 0.3216 yang mengindikasikan kualitas cluster yang cukup. Hasil pengelompokan ini menegaskan adanya perbedaan karakteristik pembangunan antarwilayah, sehingga kebijakan pembangunan daerah tidak dapat diterapkan secara seragam. Penelitian ini berkontribusi dalam menyediakan pemetaan pembangunan daerah berbasis data multidimensi yang dapat digunakan sebagai dasar awal untuk mendukung penyusunan kebijakan pembangunan yang lebih terarah, objektif, dan sesuai dengan kebutuhan masing-masing wilayah.</p>
<p>Keywords: <i>K-Means Clustering Regional Development Indicators Regional Clustering GRDP at Constant Prices Banten Province</i></p>	<p>ABSTRACT</p> <p>This study aims to cluster regencies and cities in Banten Province based on multidimensional development indicators, analyze the characteristics of each cluster, and provide initial recommendations for regional development priorities. A descriptive quantitative approach was employed using secondary data obtained from the official publications of Statistics Indonesia (BPS) of Banten Province in 2024. The study covers eight regencies and cities with six variables, namely poverty line, open unemployment rate, economic growth, gross regional domestic product at constant prices (GRDP), mean years of schooling, and life expectancy. The analysis was conducted through Min-Max normalization and implementation of the K-Means clustering algorithm using Python. The results indicate that the regencies and cities in Banten Province can be grouped into three clusters: priority development areas, developing (transitional) areas, and advanced areas. The evaluation results show a Silhouette Score of 0.3216, indicating a moderate clustering structure. These findings reveal significant disparities in development characteristics across regions, suggesting that uniform development policies may not be effective. This study contributes by providing a multidimensional, data-driven regional mapping approach that can support more targeted, objective, and differentiated policy formulation aligned with the specific needs of each region.</p>

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



I. Pendahuluan

Pembangunan daerah merupakan proses strategis yang bertujuan untuk meningkatkan kesejahteraan masyarakat secara merata. Namun, dalam praktiknya, setiap wilayah memiliki karakteristik sosial, ekonomi, pendidikan, dan kesehatan yang berbeda, sehingga pendekatan pembangunan tidak dapat diseragamkan. Perbedaan kondisi tersebut menuntut adanya kebijakan yang lebih adaptif dan berbasis data agar program pembangunan dapat dirancang sesuai dengan kebutuhan spesifik masing-masing wilayah. Provinsi Banten sebagai salah satu wilayah dengan dinamika pembangunan yang tinggi menunjukkan variasi capaian pembangunan antarkabupaten/kota pada sejumlah indikator penting, seperti tingkat kemiskinan, pengangguran, dan kualitas hidup masyarakat [1].

Meskipun ketersediaan data indikator pembangunan cukup melimpah, pemanfaatannya dalam mendukung pengambilan keputusan masih belum optimal. Kebijakan pembangunan daerah sering kali disusun secara agregat tanpa mempertimbangkan kemiripan karakteristik antarwilayah secara sistematis. Akibatnya, terdapat risiko ketidaktepatan sasaran program karena perbedaan kebutuhan antarwilayah tidak teridentifikasi secara jelas. Oleh karena itu, diperlukan suatu pendekatan analitis yang mampu mengelompokkan wilayah berdasarkan kemiripan indikator pembangunan sehingga dapat memberikan gambaran yang lebih objektif mengenai pola kesamaan dan perbedaan kondisi pembangunan daerah.

Sejumlah penelitian terdahulu menunjukkan bahwa metode *clustering*, khususnya K-Means, efektif digunakan untuk mengelompokkan wilayah berdasarkan indikator pembangunan [2]–[5]. Namun, sebagian besar penelitian masih berfokus pada indikator tunggal atau indikator sektoral tertentu, seperti Indeks Pembangunan Manusia (IPM), sehingga belum mampu menggambarkan kondisi Pembangunan secara multidimensi [2]–[5]. Selain itu, hasil pengelompokan yang dihasilkan umumnya masih terbatas pada klasifikasi statistik dan belum secara eksplisit diarahkan untuk mendukung penentuan prioritas program pembangunan daerah.

Berdasarkan hal tersebut, penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Provinsi Banten berdasarkan indikator pembangunan multidimensi, menganalisis karakteristik setiap cluster, serta memberikan rekomendasi awal mengenai prioritas program daerah. Penelitian ini diharapkan dapat memberikan kontribusi dalam bentuk pemetaan wilayah berbasis data yang lebih komprehensif dan operasional, sehingga dapat mendukung penyusunan kebijakan pembangunan daerah yang lebih tepat sasaran, objektif, dan sesuai dengan kebutuhan masing-masing wilayah.

II. Metode

2.1. Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif deskriptif. Pendekatan kuantitatif dipilih karena penelitian berfokus pada pengolahan data numerik yang berasal dari indikator pembangunan daerah, sedangkan sifat deskriptif digunakan untuk menggambarkan pola pengelompokan kabupaten/kota berdasarkan tingkat kemiripan karakteristik pembangunan yang dimiliki. Data yang dianalisis dalam penelitian ini merupakan data sekunder yang diperoleh dari publikasi resmi Badan Pusat Statistik Provinsi Banten tahun 2024, sehingga proses penelitian lebih difokuskan pada tahap kompilasi, pengolahan, dan interpretasi data untuk menghasilkan informasi pengelompokan wilayah yang mendukung analisis kebijakan daerah.

2.2. Objek Penelitian

Objek dalam penelitian ini adalah seluruh kabupaten/kota yang berada di Provinsi Banten, yaitu sebanyak 8 wilayah administratif yang terdiri atas 4 kabupaten dan 4 kota. Masing-masing kabupaten/kota digunakan sebagai unit analisis dalam proses pengelompokan, karena setiap wilayah memiliki nilai yang berbeda pada indikator pembangunan yang diteliti. Pemilihan kabupaten/kota di Provinsi Banten sebagai objek penelitian didasarkan pada adanya variasi karakteristik pembangunan antarwilayah, sehingga memungkinkan dilakukan analisis *clustering* untuk mengidentifikasi pola kemiripan maupun perbedaan kondisi pembangunan daerah.

2.3. Variabel Penelitian

Variabel dalam penelitian ini terdiri atas enam indikator pembangunan yang merepresentasikan dimensi sosial, ekonomi, pendidikan, dan kesehatan, yaitu garis kemiskinan, tingkat pengangguran terbuka, pertumbuhan ekonomi, PDRB ADHK, rata-rata lama sekolah, dan umur harapan hidup. Pemilihan variabel didasarkan pada pendekatan multidimensi, karena kondisi pembangunan daerah tidak dapat diukur hanya dengan satu indikator, melainkan perlu dilihat secara komprehensif.

Garis kemiskinan dan tingkat pengangguran terbuka digunakan untuk menggambarkan kondisi sosial dan ketenagakerjaan, sementara pertumbuhan ekonomi dan PDRB ADHK merepresentasikan kinerja serta kapasitas ekonomi daerah. Rata-rata lama sekolah mencerminkan tingkat pendidikan masyarakat, sedangkan umur harapan hidup digunakan sebagai indikator kesehatan dan kualitas hidup.

Nama kabupaten/kota digunakan sebagai identitas unit analisis, sedangkan nilai keenam indikator tersebut menjadi dasar dalam proses clustering untuk mengidentifikasi pola kemiripan karakteristik pembangunan antardaerah di Provinsi Banten.

2.4. Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari publikasi resmi Badan Pusat Statistik (BPS) Provinsi Banten tahun 2024 [6]. Data tersebut mencakup indikator kemiskinan, ketenagakerjaan, pertumbuhan ekonomi, PDRB ADHK, pendidikan, dan kesehatan pada tingkat kabupaten/kota [6]–[11].

Data dihimpun dari beberapa tabel statistik resmi BPS dan disusun menjadi satu himpunan data penelitian sehingga setiap kabupaten/kota memiliki nilai lengkap pada seluruh indikator yang digunakan. Penggunaan data tahun 2024 bertujuan untuk merepresentasikan kondisi pembangunan daerah yang mutakhir, sehingga hasil analisis dapat memberikan gambaran yang lebih relevan dan objektif dalam mendukung penentuan prioritas program daerah.

2.5. Tahap Pengolahan Data

Pengolahan data dalam penelitian ini dilakukan melalui beberapa tahapan, yaitu pengumpulan data, seleksi variabel, pembersihan data, normalisasi, clustering, dan interpretasi hasil. Data indikator pembangunan kabupaten/kota di Provinsi Banten diperoleh dari publikasi resmi BPS tahun 2024 [11], kemudian diseleksi berdasarkan relevansinya terhadap dimensi sosial, ekonomi, pendidikan, dan kesehatan.

Selanjutnya, dilakukan pembersihan data untuk memastikan seluruh variabel dalam bentuk numerik yang konsisten dan tidak terdapat data yang hilang. Data kemudian dinormalisasi menggunakan metode Min-Max untuk menyetarakan rentang nilai antarvariabel [12]. Setelah itu, dilakukan penentuan jumlah cluster berdasarkan evaluasi teknis, yang kemudian digunakan dalam proses K-Means Clustering untuk mengelompokkan wilayah berdasarkan kemiripan karakteristik indikator pembangunan [13].

Tahap akhir adalah interpretasi hasil clustering dengan menganalisis kecenderungan nilai indikator pada setiap cluster, yang selanjutnya digunakan sebagai dasar dalam memberikan gambaran awal mengenai prioritas program pembangunan daerah.

2.6. Normalisasi Data

Sebelum proses *clustering* dilakukan, data terlebih dahulu melalui tahap normalisasi agar setiap variabel berada pada skala yang sebanding. Tahap ini diperlukan karena variabel yang digunakan dalam penelitian memiliki rentang nilai yang berbeda-beda, misalnya nilai PDRB ADHK cenderung jauh lebih besar dibandingkan TPT, RLS, atau UHH. Apabila data langsung diolah tanpa normalisasi, maka variabel dengan rentang nilai terbesar berpotensi memberikan pengaruh yang lebih dominan dalam proses pengelompokan. Artikel Okereke dkk. (2023) menjelaskan alasan normalisasi dalam *clustering*, yaitu untuk mengurangi pengaruh perbedaan magnitudo data terhadap pembentukan *cluster*, dan dalam studi tersebut digunakan *MinMax normalization* sebelum proses K-means [14]. Oleh karena itu, penelitian ini menggunakan metode *Min-Max Normalization* untuk mentransformasi data ke dalam rentang nilai 0 sampai 1 [12], [15]. Rumus normalisasi yang digunakan adalah sebagai berikut:

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

dengan X' sebagai nilai hasil normalisasi, X sebagai nilai data asli, X_{min} sebagai nilai minimum, dan X_{max} sebagai nilai maksimum pada variabel yang bersangkutan. Melalui proses ini, seluruh variabel diubah ke dalam rentang yang seragam sehingga setiap indikator memiliki kontribusi yang lebih seimbang dalam proses *clustering* [12].

Penggunaan normalisasi Min-Max dalam penelitian ini juga didasarkan pada kesesuaiannya dengan algoritma K-Means, yang menggunakan ukuran jarak dalam proses pengelompokan. Karena K-Means sensitif terhadap skala data, maka penyetaraan rentang nilai antarvariabel menjadi langkah penting agar hasil cluster lebih representatif dan tidak bias terhadap variabel tertentu yang memiliki skala lebih besar. Hasil penelitian menunjukkan bahwa pada dataset dengan fitur yang memiliki satuan berbeda, penerapan *scaling* sebelum K-means menghasilkan kinerja *clustering* yang lebih baik dibandingkan penggunaan data mentah [15].

2.7. Algoritma K-Means

Algoritma K-Means merupakan salah satu metode *clustering* nonhierarki yang digunakan untuk mengelompokkan data ke dalam sejumlah *cluster* berdasarkan tingkat kemiripan karakteristik yang dimiliki. Prinsip dasar algoritma ini adalah membagi sekumpulan objek ke dalam kkk kelompok, sehingga objek dalam cluster yang sama memiliki tingkat kedekatan yang lebih tinggi dibandingkan dengan objek pada *cluster* lain. Dalam konteks penelitian ini, K-Means digunakan untuk mengelompokkan kabupaten/kota di Provinsi Banten berdasarkan nilai indikator pembangunan yang dimiliki, sehingga wilayah dengan karakteristik pembangunan

yang serupa dapat ditempatkan dalam kelompok yang sama. Secara umum, K-Means merupakan salah satu algoritma clustering yang banyak digunakan karena relatif sederhana, efisien, dan mudah diimplementasikan pada data numerik [16], [17].

Proses K-Means dimulai dengan menentukan jumlah *cluster* (k) yang akan dibentuk, kemudian memilih pusat *cluster* awal atau *centroid*. Setelah *centroid* awal ditetapkan, langkah berikutnya adalah menghitung jarak setiap objek terhadap seluruh *centroid* yang tersedia. Setiap objek kemudian ditempatkan ke dalam *cluster* dengan jarak terdekat. Setelah seluruh objek memperoleh *cluster* sementara, *centroid* baru dihitung berdasarkan rata-rata nilai seluruh anggota pada masing-masing *cluster*. Proses ini dilakukan secara berulang hingga tidak terjadi lagi perubahan keanggotaan *cluster* atau posisi *centroid* telah stabil. Dengan demikian, K-Means bekerja secara iteratif sampai diperoleh pembentukan *cluster* yang relatif konvergen [17], [18].

Dalam penelitian ini, ukuran kedekatan antarobjek ditentukan menggunakan *Euclidean Distance*, karena ukuran ini merupakan pendekatan yang umum digunakan dalam K-Means untuk data numerik [17]. Rumus jarak *Euclidean* yang digunakan adalah:

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

dengan (x, y) sebagai jarak antara objek x dan *centroid* y , x_i sebagai nilai objek pada variabel $ke - i$, y_i sebagai nilai *centroid* pada variabel $ke - i$, dan n sebagai jumlah variabel. Melalui perhitungan jarak tersebut, setiap objek akan dialokasikan ke *cluster* yang memiliki *centroid* terdekat.

Pemilihan algoritma K-Means dalam penelitian ini didasarkan pada kesesuaiannya dengan tujuan penelitian, yaitu mengelompokkan wilayah berdasarkan indikator pembangunan yang berbentuk data numerik. Selain itu, K-Means juga banyak digunakan dalam penelitian pengelompokan wilayah karena mampu memberikan hasil *cluster* yang mudah diinterpretasikan. Meskipun demikian, K-Means memiliki sensitivitas terhadap pemilihan *centroid* awal dan skala data, sehingga tahap normalisasi dan penentuan jumlah *cluster* menjadi bagian penting dalam proses analisis. Kajian Ikotun dkk. (2023) menegaskan bahwa K-means tetap menjadi salah satu algoritma *clustering* paling populer, tetapi kinerjanya sangat dipengaruhi oleh inisialisasi *centroid*, pemilihan jumlah *cluster*, dan karakteristik data yang dianalisis [16].

2.8. Penentuan Jumlah Cluster (*Silhouette Score*)

Penentuan jumlah *cluster* merupakan tahapan penting dalam algoritma K-Means karena akan memengaruhi kualitas hasil pengelompokan. Dalam penelitian ini, jumlah *cluster* ditentukan melalui pendekatan evaluasi kuantitatif menggunakan metode *Silhouette Score*. Metode ini digunakan untuk mengukur tingkat kesesuaian setiap objek terhadap *cluster* tempatnya berada dibandingkan dengan *cluster* lain, sehingga dapat memberikan indikasi kualitas struktur *cluster* yang terbentuk.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3)$$

Secara matematis, nilai *Silhouette Score* untuk setiap objek dihitung berdasarkan selisih antara rata-rata jarak intra-*cluster* (a) dan rata-rata jarak terdekat ke *cluster* lain (b), yang kemudian dinormalisasi dalam rentang nilai -1 hingga 1. Nilai yang mendekati 1 menunjukkan bahwa objek terkelompok dengan baik, sedangkan nilai mendekati 0 menunjukkan adanya tumpang tindih antar *cluster*, dan nilai negatif mengindikasikan kemungkinan kesalahan pengelompokan.

2.9. Simulasi Manual

Simulasi manual dilakukan untuk menggambarkan mekanisme dasar algoritma K-Means pada iterasi awal, mulai dari penentuan *centroid* awal hingga penghitungan jarak menggunakan *Euclidean Distance*. Simulasi ini digunakan sebagai ilustrasi konseptual proses pembentukan *cluster* sebelum implementasi dilakukan secara komputasional menggunakan Python.

2.10. Implementasi Python

Setelah simulasi manual dilakukan, proses *clustering* pada penelitian ini selanjutnya diimplementasikan menggunakan Python. Penggunaan Python dipilih karena dapat membantu proses pengolahan data dan perhitungan clustering menjadi lebih cepat dan terstruktur. Dalam penelitian ini, Python digunakan untuk membaca data, melakukan normalisasi, menjalankan algoritma K-Means, serta menampilkan hasil *cluster* dalam bentuk tabel dan grafik [17], [18], [22], [23].

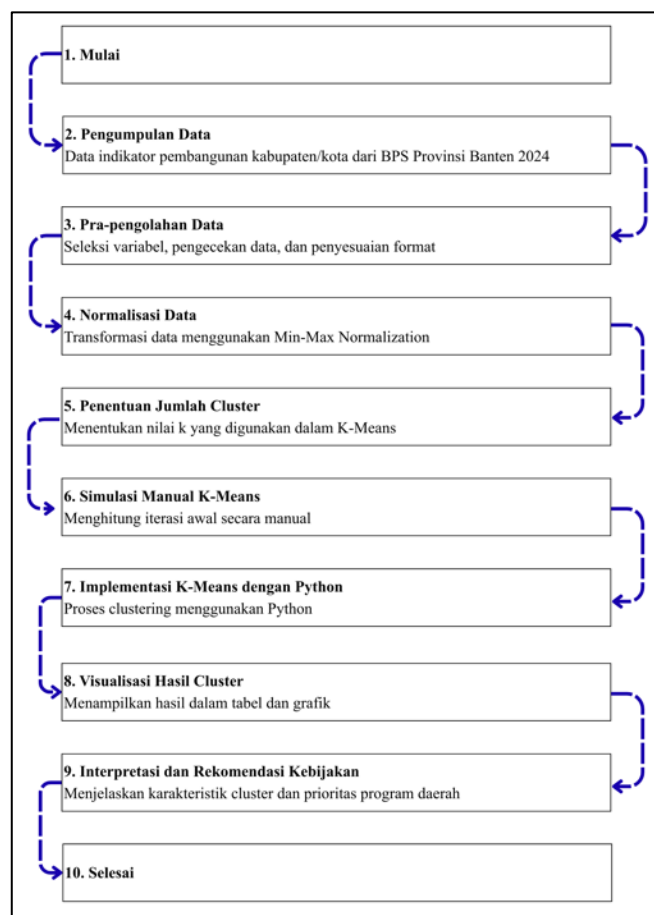
Pustaka yang digunakan dalam penelitian ini meliputi *pandas* untuk mengolah data, *scikit-learn* untuk proses normalisasi dan K-Means, serta *matplotlib* untuk visualisasi hasil. Tahapan implementasi dimulai dari memasukkan data ke dalam Python, kemudian melakukan penyesuaian format data, normalisasi, proses

clustering, dan pemberian label *cluster* pada setiap kabupaten/kota [18]. Hasil akhir kemudian divisualisasikan agar pola pengelompokan wilayah dapat dilihat dengan lebih jelas. Dengan demikian, penggunaan Python dalam penelitian ini berfungsi untuk mendukung keluaran *clustering* yang lebih informatif [17], [24].

2.11. Diagram Alur Penelitian

Diagram alur penelitian digunakan untuk menggambarkan tahapan penelitian secara sistematis, mulai dari pengumpulan data hingga interpretasi hasil *clustering*. Dalam penelitian ini, alur dimulai dari pengumpulan data indikator pembangunan kabupaten/kota di Provinsi Banten yang bersumber dari BPS Provinsi Banten tahun 2024. Data yang telah dikumpulkan kemudian melalui tahap pra-pengolahan, yang meliputi seleksi variabel, pengecekan kelengkapan data, dan penyesuaian format numerik. Selanjutnya, data dinormalisasi menggunakan metode Min-Max agar seluruh variabel memiliki skala yang sebanding. Setelah itu dilakukan penentuan jumlah *cluster*, kemudian proses *clustering* dilakukan melalui dua pendekatan, yaitu simulasi manual pada iterasi awal dan implementasi menggunakan Python. Hasil *clustering* yang diperoleh selanjutnya divisualisasikan dalam bentuk tabel dan grafik, kemudian diinterpretasikan untuk mengetahui karakteristik masing-masing *cluster* sebagai dasar penentuan prioritas program daerah.

Diagram alur penelitian menggambarkan tahapan penelitian secara sistematis mulai dari pengumpulan data sampai interpretasi hasil *clustering*. Berikut merupakan diagram alur dari penelitian terlampir pada gambar 1:



Gambar 1. Diagram Alur Penelitian

Diagram alur penelitian pada Gambar 1 menunjukkan tahapan penelitian yang dimulai dari pengumpulan data indikator pembangunan kabupaten/kota di Provinsi Banten, dilanjutkan dengan pra-pengolahan data, normalisasi, penentuan jumlah *cluster*, simulasi manual K-Means, implementasi menggunakan Python, visualisasi hasil *cluster*, hingga interpretasi hasil sebagai dasar rekomendasi kebijakan daerah.

III. Hasil dan Pembahasan

3.1 Deskripsi Data Penelitian

Data penelitian ini berupa data sekunder kuantitatif yang tertera pada Tabel 1 serta mencakup 8 kabupaten/kota di Provinsi Banten dengan 6 indikator pembangunan, yaitu garis kemiskinan, tingkat pengangguran terbuka (TPT), pertumbuhan ekonomi, PDRB ADHK, rata-rata lama sekolah (RLS), dan umur harapan hidup (UHH). Secara umum, data menunjukkan adanya perbedaan karakteristik antar kabupaten/kota di Provinsi Banten.

Tabel 1. Data Penelitian

Kabupaten/Kota	Garis Kemiskinan	TPT	Pertumbuhan Ekonomi	PDRB ADHK	RLS	UHH
Pandeglang	479.886	8,09	2,53	21.928.594,66	7,5	65,79
Lebak	440.705	6,23	2,8	23.377.240,36	6,78	68,34
Kabupaten Tangerang	632.941	6,06	5,18	113.998.210,70	9,16	70,88
Kabupaten Serang	449.291	9,18	4,8	63.255.334,10	8,03	65,9
Kota Tangerang	785.113	5,92	5,57	125.060.319,31	11,46	72,46
Kota Cilegon	663.533	6,08	4,82	88.628.114,51	10,4	67,66
Kota Serang	540.786	7,12	4,74	26.863.076,79	8,94	69,27
Kota Tangerang Selatan	828.278	5,09	5,36	73.052.841,63	12,1	73,31

3.2 Hasil Pra-pengolahan Data

Tahap pra-pengolahan data dilakukan untuk memastikan data siap digunakan dalam proses clustering. Pemeriksaan menunjukkan bahwa seluruh data pada delapan kabupaten/kota lengkap dan tidak terdapat data duplikasi. Selanjutnya dilakukan penyesuaian format numerik, seperti penghapusan tanda pemisah ribuan dan penyeragaman pemisah desimal, agar seluruh variabel dapat diproses secara numerik pada tahap normalisasi dan clustering menggunakan algoritma K-Means.

3.3 Hasil Normalisasi

Setelah tahap pra-pengolahan selesai dilakukan, data dinormalisasi menggunakan metode Min-Max untuk menyetarakan rentang nilai antarvariabel. Proses ini bertujuan agar tidak terdapat variabel yang mendominasi proses clustering akibat perbedaan skala data. Hasil normalisasi ditunjukkan pada Tabel 2.

Tabel 2. Hasil Normalisasi Min - Max

kabupaten_kota	garis_kemiskinan	tpt	pertumbuhan_ekonomi	pdrb_adhk	rls	uhh
Pandeglang	0,10	0,73	0,00	0,00	0,14	0,00
Lebak	0,00	0,28	0,09	0,01	0,00	0,34
Kabupaten Tangerang	0,50	0,24	0,87	0,89	0,45	0,68
Kabupaten Serang	0,02	1,00	0,75	0,40	0,23	0,01
Kota Tangerang	0,89	0,20	1,00	1,00	0,88	0,89
Kota Cilegon	0,57	0,24	0,75	0,65	0,68	0,25
Kota Serang	0,26	0,50	0,73	0,05	0,41	0,46
Kota Tangerang Selatan	1,00	0,00	0,93	0,50	1,00	1,00

3.4 Hasil Implementasi Python

1. Inisiasi Pemilihan Variabel

```
# -----
# 1. Input data (SUDAH NORMALISASI)
# -----
data = {
    'kabupaten_kota': [
        'Pandeglang',
        'Lebak',
        'Kabupaten Tangerang',
        'Kabupaten Serang',
        'Kota Tangerang',
        'Kota Cilegon',
        'Kota Serang',
        'Kota Tangerang Selatan'
    ],
    'garis_kemiskinan': [0.10, 0.00, 0.50, 0.02, 0.89, 0.57, 0.26, 1.00],
    'tpt': [0.73, 0.28, 0.24, 1.00, 0.20, 0.24, 0.50, 0.00],
    'pertumbuhan_ekonomi': [0.00, 0.09, 0.87, 0.75, 1.00, 0.75, 0.73, 0.93],
    'pdrb_adhk': [0.00, 0.01, 0.89, 0.40, 1.00, 0.65, 0.05, 0.50],
    'rls': [0.14, 0.00, 0.45, 0.23, 0.88, 0.68, 0.41, 1.00],
    'uhh': [0.00, 0.34, 0.68, 0.01, 0.89, 0.25, 0.46, 1.00],
}

df = pd.DataFrame(data)
```

Gambar 2. Inisiasi Pemilihan Variable

Proses clustering dilakukan menggunakan algoritma K-Means dengan jumlah *cluster* sebanyak tiga. Variabel yang digunakan dalam proses ini meliputi garis kemiskinan, tingkat pengangguran terbuka (TPT), pertumbuhan ekonomi, PDRB ADHK, rata-rata lama sekolah (RLS), dan umur harapan hidup (UHH). Implementasi dilakukan menggunakan Python dengan *library* scikit-learn.

2. Implementasi K-Means Menggunakan Python

```
# -----
# 2. Ambil variabel numerik
# -----
X = df[['garis_kemiskinan', 'tpt', 'pertumbuhan_ekonomi', 'pdrb_adhk', 'rls', 'uhh']]

# -----
# 3. K-Means Clustering
# -----
kmeans = KMeans(n_clusters=3, random_state=42, n_init=10)
labels = kmeans.fit_predict(X)

df['cluster'] = labels
```

Gambar 3. Implementasi K-Means

Hasil *clustering* menunjukkan bahwa Kota Cilegon, Kabupaten Tangerang, Kota Tangerang, Kota Tangerang Selatan tergolong ke dalam *Cluster* 1. Kabupaten Serang dan Kota Serang termasuk ke dalam *Cluster* 2, sedangkan Pandeglang dan Lebak termasuk ke dalam *Cluster* 3.

3. Perhitungan Silhouette Score

```
# -----
# 4. Silhouette Score
# -----
score = silhouette_score(X, labels)
print("Silhouette Score:", round(score, 4))
```

Silhouette Score: 0.3216

Gambar 4. Perhitungan *Silhouette Score*

Nilai *Silhouette Score* yang diperoleh sebesar 0,3216. Nilai ini berada pada kategori cukup (*moderate structure*), yang menunjukkan bahwa struktur *cluster* yang terbentuk telah memiliki pemisahan yang memadai, meskipun masih terdapat sebagian objek yang berada pada batas antar *cluster*. Hal ini wajar mengingat jumlah objek penelitian relatif kecil dan karakteristik antarwilayah memiliki kemiripan pada beberapa indikator. Dengan demikian, hasil clustering tetap dapat digunakan untuk tujuan analisis dan interpretasi kebijakan.

4. Penentuan Centroid Akhir

```
# -----
# 5. Centroid Akhir
# -----
centroids = kmeans.cluster_centers_

centroid_df = pd.DataFrame(
    centroids,
    columns=X.columns
)

print("\nCentroid Akhir:")
print(centroid_df.round(3))
```

Gambar 5. Penentuan Centroid Akhir

Nilai *centroid* akhir digunakan untuk melihat karakteristik masing-masing *cluster* berdasarkan enam indikator pembangunan. Perbedaan nilai *centroid* menunjukkan adanya variasi kondisi sosial-ekonomi antarcluster.

Tabel 4. Centroid Akhir Setiap Cluster

Cluster	Garis Kemiskinan	TPT	Pertumbuhan Ekonomi	PDRB ADHK	RLS	UHH
C1	0,740	0,170	0,888	0,760	0,752	0,705
C2	0,140	0,750	0,740	0,225	0,320	0,235
C3	0,050	0,505	0,045	0,005	0,070	0,170

5. Visualisasi Cluster

```

# -----
# 6. Mapping label cluster (biar rapi di paper)
# -----
mapping = {0: 'C1', 1: 'C2', 2: 'C3'}
df['cluster_label'] = df['cluster'].map(mapping)

# -----
# 7. PCA (visualisasi 2D)
# -----
pca = PCA(n_components=2)
X_pca = pca.fit_transform(X)

df['PCA1'] = X_pca[:, 0]
df['PCA2'] = X_pca[:, 1]

# -----
# 8. Warna cluster
# -----
warna_cluster = {
    'C1': 'blue',
    'C2': 'orange',
    'C3': 'green'
}

# -----
# 9. Plot
# -----
plt.figure(figsize=(10, 6))

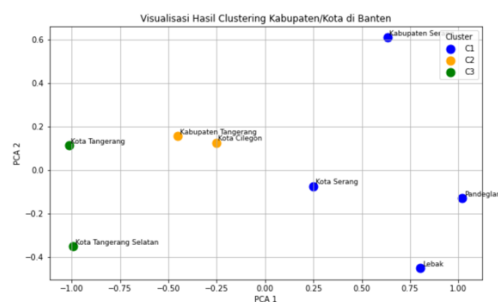
for cluster in df['cluster_label'].unique():
    subset = df[df['cluster_label'] == cluster]
    plt.scatter(
        subset['PCA1'],
        subset['PCA2'],
        label=cluster,
        s=120,
        color=warna_cluster[cluster]
    )

# Label nama daerah
for i, row in df.iterrows():
    plt.text(
        row['PCA1'] + 0.01,
        row['PCA2'] + 0.01,
        row['kabupaten_kota'],
        fontsize=9
    )

plt.title('Visualisasi Hasil Clustering Kabupaten/Kota di Banten')
plt.xlabel('PCA 1')
plt.ylabel('PCA 2')
plt.legend(title='Cluster')
plt.grid(True)

```

Gambar 6. Proses Clustering



Gambar 7. Visualisasi Cluster

Untuk mempermudah interpretasi hasil *clustering*, data divisualisasikan dalam dua dimensi menggunakan *Principal Component Analysis* (PCA). Visualisasi ini menunjukkan sebaran objek dan kedekatan antarwilayah dalam *cluster* yang sama.

3.5 Pembahasan Substantif per Cluster

Berdasarkan hasil pengelompokan menggunakan metode K-Means dengan tiga *cluster*, kabupaten/kota di Provinsi Banten terbagi ke dalam tiga kelompok yang menunjukkan tingkat kemiripan karakteristik pembangunan berdasarkan variabel garis kemiskinan, tingkat pengangguran terbuka (TPT), pertumbuhan ekonomi, PDRB ADHK, rata-rata lama sekolah (RLS), dan umur harapan hidup (UHH). Cluster 1 dikategorikan sebagai wilayah maju, Cluster 2 sebagai wilayah berkembang atau transisi, dan Cluster 3 sebagai wilayah prioritas intervensi pembangunan. Pengelompokan ini menunjukkan bahwa kondisi pembangunan antarwilayah di Provinsi Banten tidak bersifat homogen, melainkan terbagi ke dalam kelompok wilayah prioritas, wilayah berkembang, dan wilayah maju.

Cluster 1 terdiri atas Kabupaten Tangerang, Kota Tangerang, Kota Cilegon, dan Kota Tangerang Selatan menunjukkan nilai tinggi pada indikator pertumbuhan ekonomi (0,888), PDRB ADHK (0,760), rata-rata lama sekolah (0,752), dan umur harapan hidup (0,705), serta tingkat pengangguran yang relatif rendah (0,170). Hal ini mengindikasikan bahwa wilayah dalam *cluster* ini memiliki kapasitas ekonomi yang kuat, kualitas sumber daya manusia yang baik, serta kondisi kesehatan yang relatif lebih tinggi dibandingkan *cluster* lainnya. Oleh karena itu, *cluster* ini dapat dikategorikan sebagai wilayah maju.

Implikasi kebijakan *Cluster* 1 sebaiknya diarahkan pada penguatan daya saing dan transformasi ekonomi menuju sektor bernilai tambah tinggi. Pemerintah daerah dapat mendorong pengembangan industri berbasis teknologi, ekonomi digital, serta inovasi berbasis pengetahuan melalui kolaborasi antara pemerintah, sektor swasta, dan institusi pendidikan. Selain itu, penguatan ekosistem kewirausahaan, hilirisasi industri, serta peningkatan investasi juga menjadi strategi penting untuk menjaga momentum pertumbuhan. Dari sisi ketenagakerjaan, fokus kebijakan dapat diarahkan pada peningkatan kualitas tenaga kerja melalui pengembangan keterampilan lanjutan (*advanced skills*) agar mampu beradaptasi dengan kebutuhan industri modern. Dengan demikian, wilayah dalam *cluster* ini dapat berperan sebagai pusat pertumbuhan ekonomi (*growth pole*) yang mendorong perkembangan wilayah di sekitarnya.

Dalam konteks Provinsi Banten, penguatan daya saing pada *cluster* ini dapat dikaitkan dengan pengembangan kawasan industri dan jasa di Tangerang Raya serta optimalisasi kawasan industri manufaktur dan baja di Kota Cilegon. Kedekatan wilayah dengan Jakarta dan dukungan infrastruktur nasional juga menjadi faktor penting dalam mendorong transformasi ekonomi berbasis industri dan teknologi.

Cluster 2 terdiri atas Kabupaten Serang dan Kota Serang. *Cluster 2* ditandai oleh nilai tingkat pengangguran terbuka yang tinggi (0,750), sementara indikator lainnya berada pada tingkat menengah. Hal ini menunjukkan bahwa meskipun terdapat aktivitas ekonomi yang cukup berkembang, masih terdapat tantangan dalam penyerapan tenaga kerja dan peningkatan kualitas sumber daya manusia. Oleh karena itu, *cluster* ini dikategorikan sebagai wilayah berkembang atau transisi.

Kebijakan pembangunan pada *cluster* ini perlu difokuskan pada peningkatan penyerapan tenaga kerja dan penguatan keterampilan angkatan kerja. Program pelatihan vokasi yang berbasis kebutuhan industri lokal menjadi langkah strategis untuk mengurangi kesenjangan antara keterampilan tenaga kerja dan permintaan pasar. Selain itu, pemerintah daerah dapat mendorong pengembangan sektor ekonomi padat karya serta memperkuat konektivitas antara dunia pendidikan dan dunia industri melalui skema *link and match*. Intervensi lain yang penting adalah peningkatan produktivitas usaha mikro, kecil, dan menengah (UMKM) serta penciptaan iklim investasi yang kondusif. Dengan strategi tersebut, *cluster* ini diharapkan dapat bertransformasi menjadi wilayah yang lebih kompetitif dan naik kelas menuju kelompok wilayah maju.

Cluster 3 memiliki nilai yang relatif rendah pada hampir seluruh indikator, terutama pada pertumbuhan ekonomi (0,045), PDRB ADHK (0,005), rata-rata lama sekolah (0,070), dan umur harapan hidup (0,170). Hal ini menunjukkan bahwa wilayah dalam *cluster* ini masih menghadapi berbagai keterbatasan dalam aspek ekonomi, pendidikan, dan kesehatan. Oleh karena itu, *cluster* ini dikategorikan sebagai wilayah prioritas intervensi pembangunan.

Kebijakan pembangunan pada *cluster* ini perlu difokuskan pada pemenuhan kebutuhan dasar dan pengurangan ketimpangan pembangunan. Prioritas utama meliputi peningkatan akses dan kualitas pendidikan dasar, penguatan layanan kesehatan primer, serta pengurangan kemiskinan melalui program perlindungan sosial yang tepat sasaran. Selain itu, pengembangan ekonomi lokal berbasis potensi wilayah, seperti sektor pertanian, perikanan, atau pariwisata, dapat menjadi strategi untuk mendorong pertumbuhan ekonomi secara bertahap. Peningkatan infrastruktur dasar, seperti jalan, sanitasi, dan akses layanan publik, juga menjadi faktor kunci dalam mendukung percepatan pembangunan. Dengan pendekatan ini, wilayah dalam *cluster* ini diharapkan dapat mengejar ketertinggalan dan secara bertahap bertransisi menuju kelompok wilayah berkembang.

Pada wilayah seperti Pandeglang dan Lebak, pengembangan ekonomi lokal dapat diarahkan pada optimalisasi sektor pertanian, perikanan, dan pariwisata berbasis potensi alam daerah. Selain itu, peningkatan konektivitas infrastruktur menuju kawasan selatan Banten menjadi faktor penting dalam mengurangi ketimpangan pembangunan antarwilayah.

Secara keseluruhan, hasil *clustering* ini menunjukkan bahwa setiap kelompok wilayah di Provinsi Banten memiliki karakteristik dan kebutuhan kebijakan yang berbeda. Dengan demikian, hasil pengelompokan ini dapat menjadi dasar yang lebih objektif dalam mendukung penentuan prioritas program pembangunan daerah sesuai dengan karakteristik masing-masing *cluster*.

3.6 Implikasi bagi Pemerintah Daerah

Hasil *clustering* kabupaten/kota di Provinsi Banten memberikan implikasi praktis yang penting bagi pemerintah daerah, khususnya dalam mendukung perencanaan pembangunan yang lebih terarah dan berbasis data. Bagi Bappeda, hasil pengelompokan ini dapat digunakan sebagai alat bantu untuk memetakan kondisi pembangunan wilayah berdasarkan kemiripan karakteristik sosial, ekonomi, pendidikan, dan kesehatan. Dengan adanya pembagian wilayah ke dalam beberapa *cluster*, pemerintah daerah dapat memahami bahwa setiap kabupaten/kota tidak berada pada kondisi yang sama, sehingga kebijakan pembangunan tidak seharusnya disusun secara seragam.

Berdasarkan hasil *cluster*, prioritas program daerah dapat dibedakan sesuai dengan karakteristik masing-masing kelompok wilayah. Wilayah yang berada pada *cluster* prioritas intervensi dapat difokuskan pada program peningkatan kualitas pendidikan, penguatan layanan kesehatan dasar, pengurangan pengangguran, penanggulangan kemiskinan, dan dorongan pertumbuhan ekonomi lokal. Wilayah pada *cluster* berkembang atau transisi dapat diarahkan pada penguatan daya saing daerah melalui peningkatan kualitas tenaga kerja, pengembangan sektor ekonomi unggulan, dan optimalisasi investasi. Sementara itu, wilayah pada *cluster* maju dapat difokuskan pada strategi mempertahankan capaian pembangunan, peningkatan inovasi, transformasi digital, dan penguatan peran sebagai pusat pertumbuhan wilayah.

3.7 Perbandingan dengan Penelitian Sebelumnya

Hasil penelitian ini menunjukkan perbedaan dibandingkan dengan penelitian sebelumnya yang umumnya menggunakan indikator tunggal, seperti Indeks Pembangunan Manusia (IPM), dalam proses pengelompokan wilayah [2], [3]. Studi berbasis IPM pada dasarnya lebih menekankan aspek pembangunan manusia, terutama pendidikan, kesehatan, dan standar hidup, sehingga hasil *cluster* yang diperoleh cenderung menggambarkan tingkat capaian pembangunan manusia secara umum.

Sementara itu, penelitian ini menggunakan indikator yang lebih multidimensi, yaitu garis kemiskinan, tingkat pengangguran terbuka, pertumbuhan ekonomi, PDRB ADHK, rata-rata lama sekolah, dan umur harapan hidup. Dengan pendekatan tersebut, hasil clustering tidak hanya merepresentasikan kualitas pembangunan manusia, tetapi juga menggambarkan kondisi ekonomi dan ketenagakerjaan daerah secara lebih komprehensif.

Perbedaan pendekatan tersebut menyebabkan adanya perbedaan interpretasi hasil cluster. Penelitian Puspita [2], misalnya, lebih menitikberatkan pada pengelompokan wilayah berdasarkan capaian pembangunan manusia. Dalam penelitian ini, Kabupaten Serang dan Kota Serang berada pada cluster berkembang atau transisi karena memiliki tingkat pengangguran terbuka yang relatif tinggi, meskipun beberapa indikator pembangunan lainnya berada pada kategori menengah. Hal ini menunjukkan bahwa wilayah dengan capaian pembangunan manusia yang relatif baik belum tentu memiliki kondisi ekonomi dan ketenagakerjaan yang sama.

Oleh karena itu, hasil clustering dalam penelitian ini dinilai lebih relevan untuk mendukung kebijakan lintas sektor. Jika penelitian berbasis IPM lebih sesuai untuk melihat capaian pembangunan manusia, maka hasil penelitian ini dapat digunakan secara lebih luas dalam mendukung perumusan kebijakan pada sektor pendidikan, kesehatan, ketenagakerjaan, pengentasan kemiskinan, pengembangan ekonomi daerah, dan peningkatan daya saing wilayah. Dengan demikian, penelitian ini memberikan nilai tambah dalam bentuk hasil pengelompokan yang lebih operasional dan lebih sesuai untuk kebutuhan perencanaan pembangunan daerah berbasis data.

IV. Kesimpulan dan saran

Penelitian ini berhasil mengelompokkan kabupaten/kota di Provinsi Banten berdasarkan indikator pembangunan multidimensi menggunakan metode *K-Means Clustering*. Hasil analisis menunjukkan bahwa delapan kabupaten/kota terbagi ke dalam tiga cluster, yaitu wilayah prioritas intervensi, wilayah berkembang atau transisi, dan wilayah maju. Temuan ini menegaskan bahwa kondisi pembangunan antarwilayah di Provinsi Banten tidak bersifat homogen, melainkan memiliki perbedaan karakteristik pada aspek kemiskinan, ketenagakerjaan, ekonomi, pendidikan, dan kesehatan. Analisis lebih lanjut terhadap karakteristik setiap *cluster* menunjukkan adanya perbedaan kebutuhan pembangunan yang spesifik. Wilayah dalam *cluster* prioritas intervensi memiliki capaian yang relatif rendah pada sebagian besar indikator, sehingga memerlukan perhatian pada pemenuhan kebutuhan dasar dan penguatan layanan publik. Wilayah berkembang menunjukkan kondisi menengah dengan tantangan utama pada penyerapan tenaga kerja dan peningkatan kualitas sumber daya manusia. Sementara itu, wilayah maju memiliki kinerja yang relatif tinggi pada sebagian besar indikator dan berpotensi menjadi pusat pertumbuhan ekonomi daerah. Dengan demikian, hasil *clustering* ini tidak hanya memberikan pemetaan kondisi pembangunan wilayah, tetapi juga menghasilkan diferensiasi arah kebijakan yang lebih spesifik dan berbasis data. Penelitian ini berkontribusi dalam menyediakan kerangka analitik yang lebih komprehensif melalui penggunaan indikator multidimensi, serta menunjukkan bahwa pendekatan *clustering* dapat digunakan sebagai alat pendukung dalam penyusunan prioritas program pembangunan daerah yang lebih terarah dan objektif. Namun demikian, penelitian ini memiliki keterbatasan pada jumlah objek yang relatif kecil serta penggunaan data yang bersifat *cross-sectional*, sehingga belum sepenuhnya menangkap dinamika perubahan pembangunan antarwilayah dari waktu ke waktu. Selain itu, hasil clustering juga dipengaruhi oleh pemilihan jumlah *cluster* dan metode yang digunakan. Berdasarkan hasil penelitian, pemerintah daerah dapat memanfaatkan hasil clustering ini sebagai salah satu dasar dalam menyusun kebijakan pembangunan yang lebih terarah sesuai dengan karakteristik masing-masing kelompok wilayah. Pendekatan diferensiasi kebijakan diperlukan agar intervensi yang dilakukan lebih efektif dan tepat sasaran, baik dalam peningkatan kualitas layanan dasar, penguatan ekonomi daerah, maupun peningkatan daya saing wilayah. Untuk penelitian selanjutnya, disarankan agar cakupan variabel diperluas dengan menambahkan indikator lain yang relevan, seperti tingkat kemiskinan ekstrem, ketimpangan pendapatan, infrastruktur dasar, atau indikator digitalisasi daerah, sehingga hasil analisis menjadi lebih komprehensif. Selain itu, penggunaan metode clustering lain, seperti *Hierarchical Clustering* atau DBSCAN, dapat dilakukan untuk membandingkan hasil pengelompokan. Penggunaan data time series juga disarankan agar analisis tidak hanya menggambarkan kondisi pada satu periode, tetapi juga mampu menunjukkan dinamika perubahan pembangunan antarwilayah secara lebih mendalam.

Daftar Pustaka

- [1] BPS Provinsi Banten, "Indeks Pembangunan Manusia (IPM) Menurut Kabupaten/Kota di Provinsi Banten," BPS Provinsi Banten. Accessed: Mar. 15, 2026. [Online]. Available: <https://banten.bps.go.id/id/statistics-table/2/NzMjMg%3D%3D/indeks-pembangunan-manusia---ipm-menurut-kabupaten-kota-di-provinsi-banten.html>
- [2] R. N. Puspita, "Analisis K-Means Cluster pada Kabupaten/Kota di Provinsi Banten Berdasarkan Indikator Indeks Pembangunan Manusia," *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, vol. 2, no. 3, pp. 267–281, Dec. 2021, doi: 10.46306/lb.v2i3.

- [3] M. N. N. Akmal, A. Yusuf, A. M. Munier, and M. T. R. Noor, "Analisis K-Means Cluster Kabupaten/Kota di Provinsi Kalimantan Selatan berdasarkan Indikator Indeks Pembangunan Manusia," *Komputika: Jurnal Sistem Komputer*, vol. 14, no. 2, pp. 125–131, Nov. 2025, doi: 10.34010/komputika.v14i2.16301.
- [4] G. F. Anwar, U. Khaira, and P. E. P. Utomo, "Clustering Wilayah Di Indonesia Berdasarkan Kualitas Pendidikan Menggunakan Algoritma Fuzzy C-Means," *Information System Journal*, vol. 8, no. 02, pp. 167–178, Nov. 2025, doi: 10.24076/infosjournal.2025v8i02.2442.
- [5] G. R. Wahyudi, R. Rahmaddeni, E. Dini, S. Adrianto, and R. Fadila, "Pengelompokan Kabupaten di Indonesia untuk Pemetaan Pendapatan Daerah Menggunakan Algoritma K-Means," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 3, pp. 1143–1151, Aug. 2025, doi: 10.57152/malcom.v5i3.2206.
- [6] Badan Pusat Statistik Provinsi Banten, "Jumlah dan Persentase Penduduk Miskin Menurut Kabupaten/Kota di Provinsi Banten, 2024," Badan Pusat Statistik Provinsi Banten. Accessed: Mar. 15, 2026. [Online]. Available: <https://banten.bps.go.id/id/statistics-table/3/UkVkwGJVZFNWakl6VWxKVFQwWjVWeTISZDNabVFUMdkjMw%3D%3D/jumlah-dan-persentase-penduduk-miskin-menurut-kabupaten-kota-di-provinsi-banten--2024.html>
- [7] Badan Pusat Statistik Provinsi Banten, "Umur Harapan Hidup Menurut Kabupaten/Kota di Provinsi Banten," Badan Pusat Statistik Provinsi Banten. Accessed: Mar. 15, 2026. [Online]. Available: <https://banten.bps.go.id/id/statistics-table/2/NzQjMg%3D%3D/umur-harapan-hidup-menurut-kabupaten-kota-di-provinsi-banten.html>
- [8] Badan Pusat Statistik Provinsi Banten, "Rata-Rata Lama Sekolah Menurut Kabupaten/Kota di Provinsi Banten," Badan Pusat Statistik Provinsi Banten. Accessed: Mar. 15, 2026. [Online]. Available: <https://banten.bps.go.id/id/statistics-table/2/NzYjMg%3D%3D/rata-rata-lama-sekolah-menurut-kabupaten-kota-di-provinsi-banten.html>
- [9] Badan Pusat Statistik Provinsi Banten, "PDRB ADHK Menurut Kabupaten/Kota di Provinsi Banten," Badan Pusat Statistik Provinsi Banten. Accessed: Mar. 15, 2026. [Online]. Available: <https://banten.bps.go.id/id/statistics-table/2/MzEyIzI%3D%3D/pdrb-adhk-menurut-kabupaten-kota-di-provinsi-banten.html>
- [10] Badan Pusat Statistik Provinsi Banten, "Laju Pertumbuhan PDRB ADHK Menurut Kab/Kota di Provinsi Banten," Badan Pusat Statistik Provinsi Banten. Accessed: Mar. 15, 2026. [Online]. Available: <https://banten.bps.go.id/id/statistics-table/2/MzEzIzI%3D%3D/laju-pertumbuhan-pdrb-adhk-menurut-kab-kota-di-provinsi-banten.html>
- [11] Badan Pusat Statistik Provinsi Banten, "Tingkat Pengangguran Terbuka (TPT) Menurut Kabupaten/Kota di Provinsi Banten," Badan Pusat Statistik Provinsi Banten. Accessed: Mar. 15, 2026. [Online]. Available: <https://banten.bps.go.id/id/statistics-table/2/MTU3IzI%3D%3D/tingkat-pengangguran-terbuka-tpt-menurut-kabupaten-kota.html>
- [12] Scikit-Learn Developers, "MinMaxScaler," scikit-learn. Accessed: Mar. 15, 2026. [Online]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>
- [13] Scikit-Learn Developers, "scikit-learn 1.8.0 documentation," scikit-learn. Accessed: Mar. 15, 2026. [Online]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>
- [14] G. E. Okereke, M. C. Bali, C. N. Okwueze, E. C. Ukekwe, S. C. Echezona, and C. I. Ugwu, "K-means clustering of electricity consumers using time-domain features from smart meter data," *Journal of Electrical Systems and Information Technology*, vol. 10, no. 1, pp. 1–8, Jan. 2023, doi: 10.1186/s43067-023-00068-3.
- [15] C. Wongoutong, "The impact of neglecting feature scaling in k-means clustering," *PLoS One*, vol. 19, no. 12, Dec. 2024, doi: 10.1371/journal.pone.0310839.
- [16] A. M. Ikotun, A. E. Ezugwu, A. M. Abu-Mahfouz, Y. Li, M. Masadeh, and A. A. Akinyelu, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf. Sci. (N. Y.)*, vol. 622, pp. 178–210, Apr. 2023, doi: <https://doi.org/10.1016/j.ins.2022.11.139>.
- [17] Scikit-learn Developers, "Dokumentasi clustering (2.3. Clustering)," scikit-learn. Accessed: Mar. 15, 2026. [Online]. Available: <https://scikit-learn.org/stable/modules/clustering.html>
- [18] Scikit-learn Developers, "Dokumentasi KMeans," scikit-learn. Accessed: Mar. 15, 2026. [Online]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>
- [19] W. Junthopas and C. Wongoutong, "Pre-Determining the Optimal Number of Clusters for k-Means Clustering Using the Parameters Package in R and Distance Metrics," *Applied Sciences (Switzerland)*, vol. 15, no. 21, pp. 1–20, Oct. 2025, doi: 10.3390/app152111372.

- [20] H. Irwandi, O. S. Sitompul, and S. Sutarman, "K-Means Performance Optimization Using Rank Order Centroid (ROC) And Braycurtis Distance," *Sinkron : Jurnal dan Penelitian Teknik Informatika*, vol. 6, no. 2, pp. 472–478, Apr. 2022, doi: 10.33395/sinkron.v7i2.11371.
- [21] M. T. H. Lubis and M. S. Hasibuan, "Measurement of Centroid Distance in Determining Stunting Clusters," *Journal La Multiapp*, vol. 5, no. 3, pp. 270–285, Aug. 2024, doi: 10.37899/journallamultiapp.v5i3.1479.
- [22] S. Developers, "K-means," Scikit-learn Documentation. Accessed: May 03, 2026. [Online]. Available: <https://scikit-learn.org/stable/modules/clustering.html#k-means>
- [23] F. Pedregosa FABIANPEDREGOSA *et al.*, "Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot," 2011. [Online]. Available: <http://scikit-learn.sourceforge.net>.
- [24] A. Farhan AlShammari, "Implementation of Clustering using K-Means in Python," *Int. J. Comput. Appl.*, vol. 186, no. 40, pp. 12–17, Sep. 2024, doi: 10.5120/ijca2024923990.