



## ***Clustering Regencies in Indonesia for Regional Mapping Using the K-Means Algorithm***

### **Pengelompokan Kabupaten di Indonesia untuk Pemetaan Daerah Menggunakan Algoritma K-Means**

**Gustri Romi Wahyudi<sup>1</sup>, Rahmaddeni<sup>2\*</sup>, Ema Dini<sup>3</sup>,  
Sukri Adrianto<sup>4</sup>, Rahmasari Fadila<sup>5</sup>**

<sup>1,2,3,5</sup>Teknik Informatika, Universitas Sains dan Teknologi Indonesia, Pekanbaru, Indonesia

<sup>4</sup>Sistem Informasi, Universitas Dumai, Dumai, Indonesia

E-Mail: <sup>1</sup>[gustriromiwhayudi@gmail.com](mailto:gustriromiwhayudi@gmail.com), <sup>2</sup>[rahmaddeni@usti.ac.id](mailto:rahmaddeni@usti.ac.id), <sup>3</sup>[emadini2301@gmail.com](mailto:emadini2301@gmail.com),  
<sup>4</sup>[sukriadrianto@gmail.com](mailto:sukriadrianto@gmail.com), <sup>5</sup>[rahmasarifadila2@gmail.com](mailto:rahmasarifadila2@gmail.com)

*Received Jun 28th 2025; Revised Jul 24th 2025; Accepted Jul 30th 2025; Available Online Jul 31th 2025, Published Aug 15th 2025*

*Corresponding Author: Gustri Romi Wahyudi*

*Copyright © 2025 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)*

#### **Abstract**

*Economic disparities among regions remain a critical issue in Indonesia's development, which can be observed through variations in Gross Regional Domestic Product (GRDP) at the district/city level. This study aims to map the economic conditions of Indonesian regions by clustering districts and cities based on GRDP data for 2024 using the K-Means algorithm. The data were obtained from the Central Bureau of Statistics (BPS) and underwent preprocessing, including data cleaning and Min-Max normalization. The number of clusters was determined as four groups (Very Low, Low, Medium, and High) based on quartile distribution, providing a more objective and representative classification of the data. The results indicate that most districts and cities fall into the Low and Medium categories, while the High category is dominated by regions with more advanced industrial and service sectors. Evaluation using the Silhouette Score produced a value of 0.778, signifying good clustering quality with clear separation between groups. These findings highlight the persistence of regional economic disparities. This research contributes to establishing a data-driven framework for classifying regional economies, which can serve as a reference for policymakers in designing more equitable development strategies and as a basis for future studies that integrate additional socio-economic variables.*

**Keywords:** *Clustering, District/ City, Economic Disparity, GRDP, K-Means, Regional Development*

#### **Abstrak**

Kesenjangan ekonomi antarwilayah merupakan permasalahan penting dalam pembangunan Indonesia, yang salah satunya dapat dilihat melalui variasi Produk Domestik Regional Bruto (PDRB) pada tingkat kabupaten/kota. Penelitian ini bertujuan untuk memetakan kondisi ekonomi daerah di Indonesia dengan mengelompokkan kabupaten/kota berdasarkan nilai PDRB tahun 2024 menggunakan algoritma K-Means *clustering*. Data penelitian diperoleh dari Badan Pusat Statistik (BPS), kemudian melalui tahap pra-pemrosesan berupa pembersihan data dan normalisasi *Min-Max*. Jumlah kluster ditentukan sebanyak empat kelompok (Sangat Rendah, Rendah, Sedang, dan Tinggi) dengan dasar pembagian kuartil, sehingga menghasilkan klasifikasi yang lebih objektif dan representatif terhadap distribusi data. Hasil penelitian menunjukkan bahwa mayoritas kabupaten/kota termasuk dalam kategori Rendah dan Sedang, sedangkan kategori Tinggi didominasi wilayah dengan basis industri dan jasa yang lebih maju. Evaluasi menggunakan *Silhouette Score* menghasilkan nilai 0,778, yang menandakan kualitas klusterisasi cukup baik dengan pemisahan antarkelompok yang jelas. Temuan ini mengindikasikan masih adanya ketimpangan distribusi ekonomi antarwilayah. Penelitian ini berkontribusi dalam menyediakan kerangka klasifikasi ekonomi daerah berbasis data kuantitatif yang dapat digunakan sebagai acuan dalam perumusan kebijakan pembangunan yang lebih merata serta menjadi dasar bagi penelitian lanjutan yang mengintegrasikan variabel sosial-ekonomi lainnya.

**Kata Kunci:** Kabupaten/ Kota, Ketimpangan Ekonomi, Klusterisasi, K-Means, PDRB, Pembangunan Wilayah

## 1. PENDAHULUAN

Indonesia Indonesia terdiri atas 38 provinsi dan 514 kabupaten/kota, dengan keragaman ekonomi yang luas di tiap wilayah. Produk Domestik Regional Bruto (PDRB) menjadi alat pengukur utama dalam menilai kapasitas ekonomi tiap daerah dalam menghasilkan barang dan jasa. Data Badan Pusat Statistik (BPS) menunjukkan bahwa PDRB nasional tahun 2023 mencapai Rp 20.892,4 triliun, meskipun distribusinya tidak merata provinsi seperti Daerah Khusus Ibukota (DKI) Jakarta dan Jawa Barat mendominasi kontribusi, sedangkan banyak kabupaten/kota di wilayah timur memiliki kontribusi yang sangat kecil [1]. Kondisi ini menegaskan urgensi pemetaan ekonomi wilayah agar strategi pembangunan dapat disusun dengan lebih tepat dan adil.

Pada Triwulan II 2024, pertumbuhan ekonomi nasional tercatat sebesar 5,05% *Year-on-Year* dengan Produk Domestik Bruto (PDB) mencapai Rp 5.536,5 triliun. Meskipun angka pertumbuhan tersebut menunjukkan stabilitas perekonomian nasional, distribusinya belum merata antar wilayah. Provinsi di Pulau Jawa bertanggung jawab atas sekitar 57% dari total PDRB nasional, sementara wilayah di luar Jawa, khususnya kawasan timur Indonesia, masih memberikan kontribusi yang relatif kecil. Ketimpangan ini mencerminkan adanya konsentrasi aktivitas ekonomi di wilayah tertentu yang berimplikasi pada kesenjangan pembangunan antar daerah. Oleh karena itu, penelitian ini berfokus pada penerapan metode K-Means *clustering* untuk mengelompokkan kabupaten/kota di Indonesia berdasarkan data PDRB 2024 secara objektif, serta menganalisis implikasi hasil klasterisasi terhadap strategi pembangunan daerah [2].

Permasalahan utama yang diangkat ialah bagaimana menyusun pemetaan ekonomi kabupaten/kota secara komprehensif berbasis data mutakhir (PDRB 2024), sehingga strategi pembangunan dapat disusun berdasarkan karakteristik ekonomi setiap wilayah. Oleh karena itu, penelitian ini berfokus pada penerapan metode K-Means *clustering* untuk mengelompokkan kabupaten/kota di Indonesia berdasarkan data PDRB 2024 secara objektif, serta menganalisis implikasi hasil klasterisasi terhadap strategi pembangunan daerah.

*Unsupervised Learning* seperti K-Means dipilih karena keunggulannya dalam efisiensi dan kemudahan interpretasi. Misalnya, studi oleh Purba dan Hartanto membandingkan K-Means dan Fuzzy C-Means dalam analisis *clustering* harga saham BRI dan menemukan keunggulan yang spesifik terhadap kondisi data tertentu [3]. Selain itu, Ningrum dan Ahadi menerapkan K-Means untuk klasifikasi laju PDRB kabupaten/kota di Jawa Timur dengan hasil yang memadai [4]. Penelitian lain juga menggunakan K-Means untuk mengelompokkan kabupaten/kota berdasarkan aspek sosial-ekonomi seperti pendidikan, kesehatan, dan pengeluaran per kapita, serta memanfaatkan *Davies–Bouldin Index* (DBI) sebagai metrik evaluasi kualitas klaster [5].

Metode alternatif lain yang efektif meliputi Fuzzy C-Means, yang digunakan untuk mengelompokkan tingkat kemiskinan di Kabupaten/Kota Jawa Tengah [6], serta dalam pengelompokan indikator sosial-ekonomi [7]. Hierarchical *clustering* khususnya metode Ward banyak diterapkan dalam kajian ekonomi lokal dan lingkungan [8], sedangkan *Self-Organizing Map* (SOM) telah dipakai secara adaptif dalam analisis wilayah terdampak Covid-19 [9].

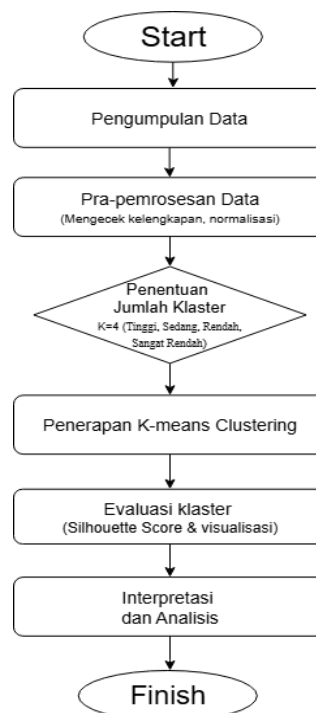
Penelitian ini menawarkan kebaruan yang jelas. Pertama, penggunaan data PDRB tahun 2024 menjadikannya sangat terkini. Kedua, hasil klasterisasi disajikan dalam empat kategori Sangat Rendah, Rendah, Sedang, dan Tinggi sehingga mudah ditafsirkan oleh pembuat kebijakan. Ketiga, visualisasi spasial hasil klasterisasi disertakan untuk membantu pengambil kebijakan melihat peta ekonomi secara langsung. Secara akademik, penelitian ini memperkaya literatur mengenai penerapan *unsupervised learning* untuk analisis ekonomi regional. Secara praktis, hasil penelitian ini diharapkan dapat menjadi acuan bagi para pembuat kebijakan dalam merumuskan strategi pembangunan daerah yang adil dan berbasis data.

## 2. METODE PENELITIAN

Penelitian ini menggunakan algoritma K-Means *clustering*, yang merupakan salah satu metode *unsupervised learning* untuk mengelompokkan data berdasarkan kesamaan karakteristik tertentu. Data yang dianalisis berupa PDRB tahun 2024 dari seluruh kabupaten dan kota di Indonesia. Tahapan yang dilalui dalam penelitian ini mencakup (lihat Gambar 1):

1. Pengumpulan Data  
Data PDRB kabupaten/kota diperoleh dari BPS tahun 2024.
2. Pra-pemrosesan Data (*Preprocessing*)  
Tahap ini meliputi pembersihan data, pengecekan data yang hilang, serta normalisasi agar data berada pada skala yang sama dan tidak mempengaruhi hasil klasterisasi
3. Penentuan Jumlah Klaster (k)  
Jumlah klaster ditentukan sebanyak empat klaster, yaitu Sangat Rendah, Rendah, Sedang, dan Tinggi. Penentuan jumlah klaster ini didasarkan pada tujuan kategorisasi yang memudahkan interpretasi dan mendukung perumusan kebijakan pembangunan wilayah.  
Klasifikasi tersebut dijelaskan sebagai berikut:
  - a. Sangat Rendah: Kabupaten/kota dengan nilai PDRB yang berada pada kuartil pertama (25% terbawah).

- b. Rendah: Kabupaten/kota dengan nilai PDRB pada kuartil kedua (26–50%).
  - c. Sedang: Kabupaten/kota dengan nilai PDRB pada kuartil ketiga (51–75%).
  - d. Tinggi: Kabupaten/kota dengan nilai PDRB pada kuartil keempat (25% teratas).
- Pembagian berdasarkan kuartil ini dilakukan agar pembagian kluster lebih objektif dan menggambarkan distribusi PDRB secara menyeluruh.
4. Penerapan Algoritma K-Means  
Setelah jumlah kluster ditetapkan, algoritma K-Means digunakan untuk mengelompokkan kabupaten/kota ke dalam empat kluster sesuai dengan klasifikasi yang telah dirancang sebelumnya.
  5. Evaluasi Hasil Klusterisasi  
Proses evaluasi dilakukan dengan menggunakan metrik *Silhouette Score* guna mengukur seberapa baik data terkelompok dan sejauh mana pemisahan antar kluster terbentuk secara optimal.
  6. Interpretasi dan Analisis Kluster  
Langkah ini bertujuan untuk menelaah hasil klusterisasi guna memahami karakteristik masing-masing kelompok wilayah serta implikasinya terhadap kondisi ekonomi regional.



**Gambar 1.** Metodologi Penelitian

### 2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari PDRB atas dasar harga berlaku tahun 2024, mencakup seluruh kabupaten dan kota di Indonesia. Sumber data diperoleh dari publikasi resmi BPS [1]. PDRB dipilih sebagai indikator utama karena mampu merepresentasikan tingkat kemampuan ekonomi suatu daerah dalam menghasilkan barang dan jasa. Nilai yang digunakan adalah total PDRB per kabupaten/kota selama satu tahun penuh, yakni tahun 2024.

### 2.2. Pra-pemrosesan Data

Tahapan pra-pemrosesan dilakukan untuk memastikan bahwa data telah siap dan sesuai untuk diolah menggunakan algoritma K-Means. Proses ini mencakup beberapa langkah utama sebagai berikut:

1. Pemeriksaan kelengkapan data: Menghapus atau memperbaiki data yang hilang (*missing values*).
2. Mengingat skala nilai PDRB sangat beragam antar wilayah, proses normalisasi dilakukan menggunakan metode *Min-Max*. Tujuan dari normalisasi ini adalah untuk menyetarakan seluruh nilai agar berada dalam rentang 0 hingga 1, sehingga tidak terjadi dominasi nilai yang terlalu besar dalam proses klusterisasi. Adapun rumus yang digunakan adalah metode *Min-Max scaling*.

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

Normalisasi diperlukan untuk menghindari dominasi variabel dengan skala lebih besar terhadap hasil klusterisasi [10].

### 2.3. Penentuan Jumlah Klaster

Penelitian ini menetapkan jumlah klaster sebanyak empat (4) berdasarkan pertimbangan kebijakan dan kebutuhan analisis ekonomi. Pembagian ini ditujukan untuk mengklasifikasikan kabupaten/kota ke dalam kategori Sangat Rendah, Rendah, Sedang, dan Tinggi. Penentuan jumlah klaster secara langsung tanpa metode eksploratif seperti *Elbow Method* dilakukan agar hasil klasifikasi sesuai dengan tujuan kebijakan pembangunan ekonomi [11]. Dengan demikian, tidak dilakukan analisis eksploratif tambahan untuk menentukan jumlah klaster.

### 2.4. Penerapan K-Means Clustering

K-Means merupakan teknik klasterisasi non-hierarki yang berfungsi untuk mengelompokkan data ke dalam sejumlah klaster berdasarkan kedekatannya terhadap pusat klaster (*centroid*). Tahapan dalam pelaksanaan algoritma K-Means antara lain meliputi:

1. Menetapkan jumlah klaster yang diinginkan, pada penelitian ini digunakan empat klaster ( $k = 4$ ).
2. Menentukan posisi awal *centroid* secara acak..
3. Mengelompokkan setiap data ke dalam klaster berdasarkan jarak Euclidean terdekat dengan *centroid*.
4. Memperbarui posisi *centroid* dengan menghitung rata-rata nilai dari seluruh anggota dalam masing-masing klaster.
5. Mengulangi proses pengelompokan dan pembaruan *centroid* hingga hasilnya konvergen atau jumlah iterasi maksimum tercapai [12].

Algoritma ini diimplementasikan menggunakan *Python* di platform *Google Colab* dengan library *scikit-learn*. Parameter *random state* diatur untuk memastikan replikasi hasil.

### 2.5. Visualisasi dan Analisis Klaster

Visualisasi hasil klasterisasi digunakan untuk mempermudah pemahaman pola data yang terbentuk. Salah satu teknik yang umum dipakai adalah *Principal Component Analysis* (PCA), yang berfungsi sebagai alat bantu untuk menampilkan data dalam dua dimensi sehingga hubungan antar klaster lebih mudah diamati [14]. Namun, penerapan PCA dalam konteks ini hanya bersifat komplementer dan tidak dimaksudkan sebagai metode reduksi dimensi utama.

Setelah klaster terbentuk, dilakukan analisis distribusi wilayah berdasarkan hasil pengelompokan untuk mengidentifikasi ciri khas pada tiap kategori ekonomi. Hal ini penting karena dapat memberikan gambaran mengenai kesenjangan maupun potensi yang ada di tiap daerah [13]. Selanjutnya, hasil klasterisasi dijadikan dasar dalam merumuskan rekomendasi kebijakan yang lebih tepat sasaran bagi pembangunan daerah, sehingga informasi yang diperoleh mampu mendukung perencanaan ekonomi yang lebih merata dan berkelanjutan [15].

### 2.6. Clustering

*Clustering* merupakan salah satu metode analisis data yang digunakan untuk mengelompokkan objek ke Klasterisasi merupakan teknik dalam analisis data yang digunakan untuk membagi objek ke dalam kelompok-kelompok berdasarkan tingkat kemiripan tertentu. Objek-objek dalam satu klaster memiliki ciri-ciri yang relatif homogen, sementara objek antar klaster menunjukkan perbedaan karakteristik. Metode ini tergolong dalam pendekatan *unsupervised learning*, karena tidak memerlukan label atau nilai target dalam proses pembentukan kelompok [16]. Klasterisasi telah diterapkan secara luas dalam berbagai bidang, seperti segmentasi pasar, pemetaan wilayah, dan analisis perilaku. Dalam konteks penelitian ini, metode klasterisasi diterapkan untuk mengelompokkan kabupaten dan kota di Indonesia berdasarkan nilai PDRB, guna memberikan dasar dalam penyusunan kebijakan pembangunan yang lebih adil dan terarah [17].

### 2.7. K-Means

K-Means adalah algoritma klasterisasi non-hierarkis yang telah banyak digunakan dalam berbagai penelitian di Indonesia selama lima tahun terakhir, karena kepraktisannya dalam penerapan, efisiensi komputasi, serta kemampuannya dalam mengelola dataset besar secara efektif [18]. Algoritma ini dimulai dengan menentukan jumlah klaster  $k$  dan pemilihan *centroid* awal secara acak. Setiap data kemudian dievaluasi jaraknya terhadap *centroid* tersebut, biasanya menggunakan jarak Euclidean, dan ditempatkan ke klaster yang memiliki jarak terdekat. Setelah pembagian awal selesai, posisi *centroid* diperbarui dengan menghitung rata-rata seluruh data dalam klaster. Proses ini terus diulang hingga posisi *centroid* tidak mengalami perubahan signifikan atau fungsi objektif telah mencapai konvergensi [19]. Secara matematis, fungsi objektif K-Means dapat dituliskan sebagai:

$$J = \sum_{i=1}^k \sum_{x \in C_i} \|x_i - \mu_i\|^2 \quad (2)$$

Dengan  $k$  mewakili jumlah kluster,  $C_i$  sebagai himpunan anggota kluster ke- $i$ ,  $x$  sebagai data ke- $n$ , dan  $\mu_i$  sebagai *centroid* kluster ke- $i$  [18]. Beberapa studi di Indonesia telah menyoroti efektivitas penerapan K-Means. Algoritma ini digunakan untuk mengelompokkan wilayah tertinggal berdasarkan indikator kemiskinan [20], mengklasifikasikan provinsi berdasarkan tingkat risiko COVID-19 [18], serta mengelompokkan wilayah berdasarkan jumlah kasus dan kematian COVID-19 [20].

Berdasarkan temuan tersebut, penelitian ini mengadopsi K-Means untuk mengklasifikasikan kabupaten/kota di Indonesia ke dalam empat kategori ekonomi: sangat rendah, rendah, sedang, dan tinggi, berdasarkan data PDRB tahun 2024. Hasil klusterisasi diharapkan dapat memberikan representasi distribusi ekonomi regional yang objektif sekaligus menjadi acuan bagi perumusan strategi pembangunan nasional yang lebih tepat sasaran [18][20].

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Pengumpulan Data

Penelitian ini menggunakan data PDRB atas dasar harga berlaku tahun 2024 pada tingkat kabupaten dan kota di Indonesia (lihat Tabel 1). Sumber data diperoleh dari BPS dan mencakup total 514 daerah administratif. PDRB dipilih sebagai indikator utama karena dianggap mampu merepresentasikan kapasitas ekonomi suatu wilayah dalam menghasilkan barang dan jasa. Dengan menggunakan data PDRB terkini, analisis ini diharapkan dapat menyajikan potret aktual kondisi ekonomi regional di seluruh Indonesia.

**Tabel 1.** Data Produk Domestik Regional Bruto 2024

| No  | Kabupaten/ Kota    | PDRB     |
|-----|--------------------|----------|
| 0   | Simeulue           | 3150.76  |
| 1   | Aceh Singkil       | 3441.08  |
| 2   | Aceh Selatan       | 7466.23  |
| 2   | Aceh Tenggara      | 6737.89  |
| 4   | Aceh Timur         | 14935.34 |
| ... | ...                | ...      |
| 513 | Pegunungan Bintang | 2527.83  |

#### 3.2. Pra-pemrosesan Data

Pada tahap pra-pemrosesan, data diperiksa dan dibersihkan guna memastikan tidak terdapat nilai kosong atau data yang tidak valid. Setelah itu, dilakukan proses normalisasi untuk menyetarakan skala nilai PDRB, sehingga tidak ada variabel yang mendominasi hasil klusterisasi. Normalisasi ini menggunakan metode *Min-Max scaling*, yang mentransformasikan nilai PDRB ke dalam rentang 0 hingga 1. Prosedur ini sangat penting agar algoritma K-Means dapat melakukan pengelompokan dengan lebih objektif dan tidak terpengaruh oleh perbedaan skala antar wilayah. Data *PDRB pra-prosesing* dapat dilihat pada Tabel 2.

**Tabel 2.** Data PDRB *Pra- Prosesing*

| No | Kabupaten_Kota     | PDRB    | <i>PDRB_Norm</i> | <i>PDRB_Normalized</i> |
|----|--------------------|---------|------------------|------------------------|
| 0  | Simeulue           | 3150.76 | 0.003085         | 0.003085               |
| 1  | Aceh Singkil       | 3441.08 | 0.003399         | 0.003399               |
| 2  | Aceh Selatan       | 7466.23 | 0.007762         | 0.007762               |
| 3  | Aceh Tenggara      | 6737.89 | 0.006973         | 0.006973               |
| 4  | Pegunungan Bintang | 2527.83 | 0.015859         | 0.015859               |

#### 3.3. Penentuan Jumlah Kluster

Jumlah Penelitian ini menetapkan jumlah kluster menjadi empat kelompok, yaitu Sangat Rendah, Rendah, Sedang, dan Tinggi. Penentuan jumlah kluster ini bukan hanya berdasarkan metode eksploratif seperti *Elbow Method* atau *Silhouette Analysis*, tetapi juga disesuaikan dengan kebutuhan interpretasi yang relevan dalam merumuskan kebijakan pembangunan ekonomi di Indonesia.

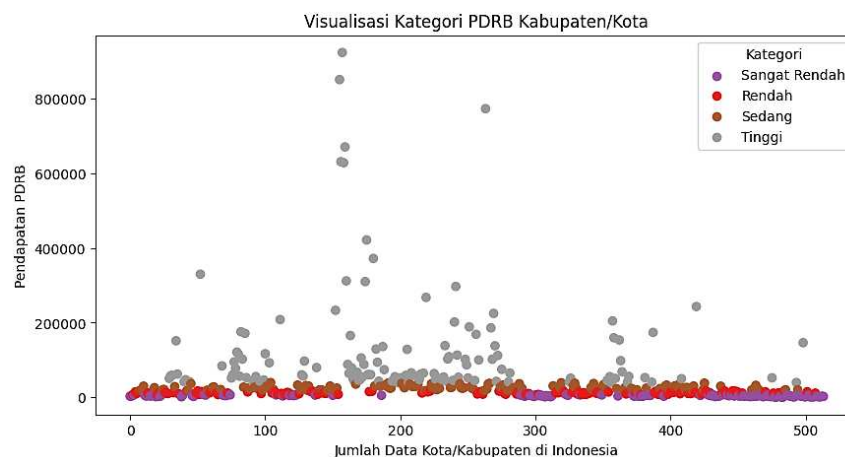
Distribusi data PDRB menunjukkan perbedaan cukup besar antar daerah, sehingga pendekatan berbasis kuartil dianggap sebagai cara yang obyektif dan mewakili. Setiap kuartil mewakili 25% dari seluruh data, sehingga memungkinkan pemetaan wilayah berdasarkan tingkat kemampuan ekonominya secara seimbang. Metode ini sesuai dengan tujuan akhir penelitian, yaitu memberikan klasifikasi wilayah yang mudah dipahami dan bisa digunakan sebagai dasar awal dalam merencanakan pembangunan. Berikut ini penjelasan mengenai klasifikasi wilayah:

1. Sangat Rendah: Kabupaten/ Kota yang berada di kuartil pertama (25% nilai PDRB terendah)
2. Rendah: Kabupaten/ Kota yang berada di kuartil kedua (26–50%)
3. Sedang: Kabupaten/ Kota yang berada di kuartil ketiga (51–75%)
4. Tinggi: Kabupaten/ Kota yang berada di kuartil keempat (25% nilai PDRB tertinggi)

Hasil awal menunjukkan bahwa pendekatan empat kluster ini mampu membagi data secara seimbang dan tidak tidak seimbang, dengan masing-masing kluster terdiri dari jumlah daerah yang cukup merata. Hal ini menegaskan bahwa metode kuartil sebagai dasar penentuan jumlah kluster dalam penelitian ini adalah tepat dan sesuai baik secara statistik maupun secara substansi kebijakan.

### 3.4. Hasil Klasterisasi dengan Algoritma K-Means

Setelah dilakukan pra-pemrosesan, algoritma K-Means diterapkan untuk mengelompokkan kabupaten/kota ke dalam empat kluster yang telah ditentukan. Berdasarkan hasil klasterisasi, diperoleh distribusi yaitu Kategori Sangat Rendah 129 kabupaten/kota, Kategori Rendah 128 kabupaten/kota, Kategori Sedang 128 kabupaten/kota, Kategori Tinggi 129 kabupaten/kota.

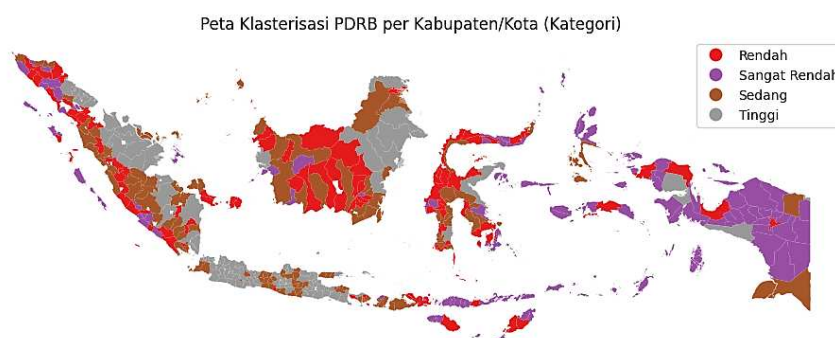


**Gambar 2.** Scatter plot hasil klasterisasi

Berdasarkan Dari Gambar 2 *Scatter Plot* hasil Klasterisasi, dapat terlihat bahwa :

1. Sumbu X Memperlihatkan bahwa *Index* atau jumlah Data Kota/Kabupaten di Indonesia dari urutan 0 hingga 514 daerah administratif yang terdata pada PDRB
2. Sumbu Y Memperlihatkan nilai PDRB per kabupaten/kota.
3. Warna Titik Menunjukkan intensitas nilai PDRB dimulai dari warna ungu (rendah) ke kuning (tinggi).

Dapat diperhatikan bahwa berdasarkan dari hasil penelitian ini terlihat dengan jelas jumlah perbandingan pendapatan di setiap kategorinya berdasarkan dari pendapatan PDRB yang dihasilkan selama 1 tahun penuh, Terlihat ada beberapa daerah yang ketimpangan pendapatannya sangat jauh berbeda dari yang lain. Untuk memahami lebih lanjut dapat kita lihat penyebaran pada setiap daerah dengan memaparkan penyebaran data berdasarkan peta Indonesia dari hasil penelitian ini sehingga terlihat dengan jelas daerah mana saja yang terbagi berdasarkan kluster yang telah terbentuk.



**Gambar 3.** Peta Klasterisasi PDRB dari tingkat klasternya

Dari Hasil Pembagian kluster yang di perlihatkan pada Gambar 3 dari bentuk peta penyebarannya berdasarkan klasterisasinya dapat kita lihat dengan jelas daerah mana saja yang tergolong Kategori Rendah, sangat rendah, sedang dan Tinggi sehingga dapat kita Analisa sebagai berikut :

1. Kesenjangan Ekonomi antar daerah masih tinggi, dengan ditandai dengan dominasi kategori rendah dan sangat rendah masih besar terlihat di luar Pulau Jawa.
2. Pusat-pusat ekonomi Indonesia masih sangat terpusat, terutama di kota-kota besar dan daerah industri.

3. Daerah dengan Kategori sedang memiliki potensi strategis untuk dikembangkan sebagai Kawasan penyangga pertumbuhan ekonomi di Indonesia.

Kemudian di paparkan Tabel 3 untuk jumlah daerah kabupaten/kota yang dihasilkan dari setiap kluster yang terbentuk dari hasil penelitian ini.

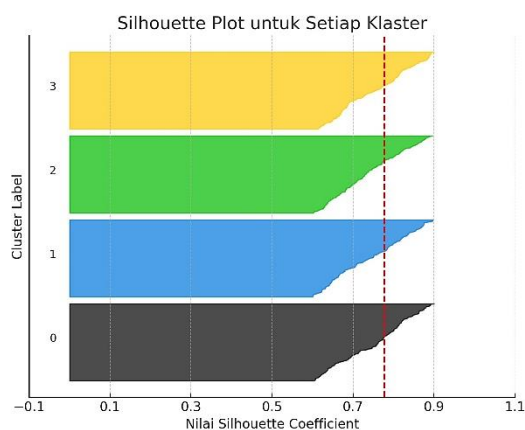
**Tabel 3.** Jumlah setiap tingkatan Klaster

| No | Tingkatan Klaster | Jumlah Kota/ Kabupaten |
|----|-------------------|------------------------|
| 1  | Sangat Rendah     | 129                    |
| 2  | Rendah            | 128                    |
| 3  | Sedang            | 128                    |
| 4  | Tinggi            | 129                    |

Berdasarkan dari jumlah klaster yang terbentuk pada setiap kabupaten dan kota di Indonesia memperlihatkan bahwa masih banyaknya kluster yang berada dalam kategori sangat rendah dan rendah yang tersebar di seluruh Indonesia.

### 3.5. Evaluasi Hasil Klasterisasi

Evaluasi kualitas hasil klasterisasi dilakukan menggunakan *Silhouette Score*, yang pada penelitian ini menghasilkan skor sebesar 0.778 (lihat Gambar 4). Nilai ini menunjukkan bahwa pembagian klaster memiliki pemisahan yang baik dan jarak antar klaster relatif jelas. Semakin mendekati angka 1, semakin baik kualitas klaster yang dihasilkan, sehingga hasil ini dapat dikatakan cukup valid untuk dijadikan dasar analisis lanjutan.



**Gambar 4.** *Silhouette Plot* untuk Setiap Klaster

### 3.6. Interpretasi dan Analisis Klaster

Berdasarkan hasil klasterisasi, kabupaten/kota terbagi menjadi empat kategori, yakni Sangat Rendah, Rendah, Sedang, dan Tinggi. Mayoritas daerah termasuk dalam kategori Rendah dan Sedang, menunjukkan adanya kesenjangan ekonomi antarwilayah. Kabupaten/kota dalam kategori Sangat Rendah umumnya memiliki keterbatasan dalam diversifikasi sektor ekonomi, sementara kategori Tinggi didominasi oleh wilayah dengan sektor industri dan jasa yang berkembang pesat serta menjadi pusat kegiatan ekonomi nasional. Evaluasi menggunakan *Silhouette Score* menghasilkan nilai 0,778, menandakan bahwa pemisahan klaster cukup baik dan jarak antar klaster relatif jelas, sehingga memperkuat validitas hasil sebagai dasar analisis lebih lanjut.

Hasil ini sejalan dengan penelitian sebelumnya yang memanfaatkan K-Means untuk pengelompokan wilayah di Indonesia. Contohnya, penelitian di Pulau Sumatra pada 2023 menemukan bahwa perbedaan karakteristik sosial-ekonomi, seperti tingkat pendidikan, pengeluaran rumah tangga, dan tingkat pengangguran, menjadi faktor utama terbentuknya klaster kemiskinan [21]. Temuan tersebut mendukung hasil penelitian saat ini, di mana variasi kontribusi sektor ekonomi memengaruhi distribusi PDRB antarwilayah. Penelitian lain terkait ketahanan pangan di Indonesia juga menunjukkan efektivitas K-Means dalam mengelompokkan kabupaten/kota berdasarkan indikator ekonomi dan sosial, termasuk PDRB per kapita [22]. Hal ini menegaskan bahwa metode K-Means dapat diterapkan secara luas untuk analisis wilayah berbasis variabel multidimensi.

Selain itu, penelitian di Provinsi Papua menggunakan metode alternatif seperti *Single Linkage* dan K-Medoids untuk pengelompokan kabupaten/kota berdasarkan PDRB sektoral [23]. Hasil menunjukkan bahwa

metode non-hierarkis selain K-Means juga mampu menangani data dengan *outlier* kuat, sehingga dapat menjadi alternatif dalam penelitian mendatang. Jika dibandingkan dengan hasil penelitian ini, K-Means terbukti lebih efisien dan memberikan pemisahan kluster yang jelas, meskipun memiliki keterbatasan dalam menghadapi data ekstrem.

Implikasi dari penelitian ini menunjukkan pentingnya strategi pembangunan yang disesuaikan dengan karakteristik kluster daerah. Kabupaten/kota dalam kategori Sangat Rendah membutuhkan intervensi khusus berupa diversifikasi ekonomi dan peningkatan infrastruktur dasar, sementara daerah dalam kategori Tinggi dapat difokuskan sebagai pusat pertumbuhan baru dengan memperkuat konektivitas ke wilayah sekitarnya. Secara akademik, penelitian ini menegaskan relevansi K-Means dalam analisis spasial-ekonomi di Indonesia, dan secara praktis hasilnya dapat menjadi acuan bagi pemerintah dalam merumuskan kebijakan pembangunan yang lebih merata. Keterbatasan penelitian ini terletak pada penggunaan satu variabel utama, yaitu PDRB, sehingga penelitian selanjutnya disarankan memasukkan indikator sosial-ekonomi tambahan agar hasil klusterisasi lebih komprehensif.

#### 4. KESIMPULAN

Penelitian ini berhasil melakukan pengelompokan kabupaten dan kota di Indonesia berdasarkan data PDRB tahun 2024 menggunakan algoritma K-Means. Hasil klusterisasi menunjukkan terbentuknya empat kategori tingkat ekonomi: Sangat Rendah, Rendah, Sedang, dan Tinggi. Kategori Sangat Rendah mencakup daerah dengan keterbatasan diversifikasi ekonomi dan dominasi sektor primer. Kategori Rendah dan Sedang menggambarkan wilayah yang berada dalam kondisi ekonomi transisi, di mana kontribusi sektor industri dan jasa mulai meningkat. Sementara itu, kategori Tinggi didominasi oleh daerah dengan sektor industri dan jasa yang maju, sekaligus menjadi pusat kegiatan ekonomi nasional.

Evaluasi kualitas klusterisasi menggunakan *Silhouette Score* menunjukkan nilai 0,778. Angka ini menandakan bahwa pemisahan antar kluster cukup jelas dan relatif optimal dibandingkan opsi jumlah kluster lainnya, sehingga kelompok wilayah yang terbentuk dapat dianggap valid dan representatif. Secara praktis, hasil penelitian ini dapat menjadi dasar dalam perumusan strategi pembangunan daerah yang lebih terarah. Pemerintah pusat maupun daerah dapat memanfaatkan pemetaan ini sebagai acuan dalam menentukan prioritas pembangunan, alokasi sumber daya, serta upaya pengurangan kesenjangan antarwilayah.

Kendati demikian, penelitian ini memiliki keterbatasan karena hanya menggunakan satu variabel utama, yaitu PDRB. Penelitian selanjutnya disarankan untuk memasukkan indikator tambahan, seperti tingkat kemiskinan, pendidikan, dan infrastruktur, agar hasil klusterisasi dapat memberikan gambaran yang lebih menyeluruh mengenai kondisi ekonomi daerah.

#### REFERENSI

- [1] Badan Pusat Statistik, "Produk Domestik Regional Bruto Kabupaten/Kota 2019–2023," Jakarta: BPS, 2024.
- [2] Badan Pusat Statistik, "Pertumbuhan Ekonomi Triwulan II 2024," Jakarta: BPS, 2024.
- [3] W. N. Purba dan R. Hartanto, "Perbandingan penerapan algoritma K-Means dan Fuzzy C-Means dalam analisis clustering terhadap pergerakan harga historis saham Bank Rakyat Indonesia," *J. Teknik Informasi dan Komputer (Tekinkom)*, vol. 7, no. 2, pp. 865–872, Des. 2024. DOI: <https://doi.org/10.37600/tekinkom.v7i2.1214>
- [4] A. F. Ningrum dan G. D. Ahadi, "Analisis kluster kabupaten/kota di Provinsi Jawa Timur berdasarkan laju PDRB dengan pendekatan K-Means," *J. Kompetitif*, vol. 8, no. 2, pp. 60–76, Sep. 2022. [Online]. Tersedia: <https://kompetitif.unizar.ac.id/jkomp/article/view/5>
- [5] C. Fitria, I. Amalia, dan M. Sitorus, "Analisis dan pengelompokan kota/kabupaten di Indonesia berdasarkan aspek sosial ekonomi menggunakan algoritma K-Means," *Innotech: Jurnal Ilmu Komputer, Sistem Informasi dan Teknologi Informasi*, vol. 2, no. 1, Jan. 2025. [Online]. Tersedia: <https://ejournal.cyber-univ.ac.id/index.php/innotech/article/view/24>
- [6] R. D. Faturahman dan N. Hidayati, "Implementasi Fuzzy C-Means dalam pengelompokan tingkat kemiskinan pada kabupaten/kota di Provinsi Jawa Tengah," *JUPI*, vol. 10, no. 1, 2025. DOI: <https://doi.org/10.29100/jupi.v10i1.5747>
- [7] D. Sulaiman, R. A. Al-Haritz, F. A. Saputra, dan A. Irawan, "Analisis kluster negara berdasarkan indikator sosial-ekonomi menggunakan Fuzzy C-Means dan K-Means," *Rekursif: J. Informatika*, vol. 12, no. 2, pp. 76–87, 2025. [Online]. Tersedia: <https://ejournal.unib.ac.id/rekursif/article/view/38116>
- [8] E. Widodo, P. Ermayani, L. N. Laila, dan A. T. Madani, "Pengelompokan provinsi di Indonesia berdasarkan tingkat kemiskinan menggunakan analisis hierarchical agglomerative clustering," *Semin. Nas. Off. Stat.*, vol. 2, pp. 557–566, 2021. DOI: <https://doi.org/10.34123/semnasoffstat.v2021i1.971>
- [9] F. D. Astuti, "Implementasi Fuzzy C-Means untuk clustering penduduk miskin (studi kasus: Kecamatan Bantul)," *Teknomatika: J. Informatika dan Komputer*, vol. 9, no. 1, pp. 59–70, 2020. [Online]. Tersedia: <https://ejournal.stkippgiritlungagung.ac.id/index.php/jipi/article/view/5747>
- [10] S. Santosa dan M. A. Sopiya, "Implementasi K-Means clustering untuk segmentasi data ekonomi,"

- Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 6, no. 2, pp. 115–122, 2019.
- [11] D. Rahmalia dan H. Hendrayati, "Segmentasi wilayah berdasarkan indeks pembangunan menggunakan K-Means," *Jurnal Informatika Mulawarman*, vol. 16, no. 1, pp. 45–52, 2021.
  - [12] M. Lestari, "Penerapan K-Means clustering pada data sosial ekonomi kabupaten/kota di Indonesia," *Jurnal Sains dan Teknologi*, vol. 5, no. 2, pp. 67–75, 2020.
  - [13] A. P. Sari dan T. Widianoro, "Penerapan metode K-Means untuk pengelompokan data mahasiswa berdasarkan lama studi," *Jurnal Pilar Nusa Mandiri*, vol. 16, no. 1, pp. 83–88, 2020.
  - [14] R. Wulandari dan H. Nugroho, "Penerapan K-Means untuk klasterisasi kabupaten di Jawa Tengah berdasarkan indikator ekonomi," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 2, pp. 123–132, 2020.
  - [15] S. Pratama dan A. Maulana, "Visualisasi klasterisasi menggunakan PCA pada data sosial ekonomi Indonesia," *Jurnal Sistem Informasi dan Sains Data*, vol. 5, no. 1, pp. 45–54, 2021.
  - [16] D. Lestari, M. Ramadhan, dan N. Fadhilah, "Analisis klaster PDRB dengan metode K-Means untuk perencanaan pembangunan daerah," *Jurnal Ekonomi dan Pembangunan Indonesia*, vol. 22, no. 1, pp. 67–78, 2022.
  - [17] A. A. Putra dan R. Yanuar, "Pengelompokan wilayah di Provinsi Banten berdasarkan indeks pembangunan manusia menggunakan metode K-Means," *Jurnal Teknik Informatika (JUTIF)*, vol. 2, no. 1, pp. 42–48, 2021.
  - [18] D. Abdullah, A. S. Sunge, dan A. T. Zy, "The application of K-means clustering for province classification based on COVID-19 cases in Indonesia," *J. Comput. Networks, Archit. High Perform. Comput.*, vol. 6, no. 3, pp. 1055–1065, Jul. 2024.
  - [19] R. Wahyuni, "K-Means clustering for grouping Indonesia underdeveloped regions in 2020 based on poverty indicators," *Parameter: J. Stat.*, vol. 2, no. 1, pp. 8–15, Des. 2021.
  - [20] H. Mulyani, "Clustering area COVID-19 Indonesia with K-Means clustering method," *J. Inf. Technol. Its Util.*, vol. 5, no. 2, pp. 42–50, Des. 2022.
  - [21] A. S. Rangkuti, "Implementation of K-Means clustering in poverty analysis of regency/city in Sumatera Island in 2023," *Jurnal Ilmiah Sistem Informasi dan Teknologi Terapan Indonesia (SINTESA)*, vol. 2, no. 1, pp. 45–55, 2025.
  - [22] L. A. Pratiwi dan R. Wulandari, "Implementasi algoritma K-Means pada pengelompokan ketahanan pangan di Indonesia menurut kabupaten/kota," *Prosiding Seminar Nasional Data Analytics (SENADA)*, vol. 2, pp. 161–170, 2024.
  - [23] Y. D. Rumabar, "Clustering kabupaten/kota di Provinsi Papua berdasarkan produk domestik regional bruto tokoh lapangan usaha menggunakan single linkage dan K-medoids," *Jurnal Gaussian*, vol. 13, no. 2, pp. 155–166, 2024.