

## Klasterisasi Berdasarkan Data Penduduk Miskin di Asia dan Wilayah Indonesia Menggunakan Metode *K-Means*

Muhamad Sabil Bilqisthi<sup>1</sup>, Sa'adah Rahma Ilahi<sup>2</sup>, Arjun Afrianto<sup>3</sup>, Fajar Saputra<sup>4</sup>,  
Rifkillah Muliya Al-Faris<sup>5</sup>, Yeskarwani Gulo<sup>6</sup>

<sup>1,2,3,4,5,6</sup> Universitas Pamulang, Indonesia

E-mail: [muhammadsabil2922@gmail.com](mailto:muhammadsabil2922@gmail.com)<sup>1</sup>, [ilahisaadah@gmail.com](mailto:ilahisaadah@gmail.com)<sup>2</sup>

### Article History:

Received: 01 Juli 2025

Revised: 02 September 2025

Accepted: 23 September 2025

**Keywords:** *Pengelompokan K-Means, Analisis Komponen Utama (PCA), Kemiskinan, Pengelompokan Negara Bagian, Prioritas Kebijakan Sosial.*

**Abstract:** *Penelitian ini bertujuan untuk mengelompokkan negara-negara di Asia dan Indonesia berdasarkan data jumlah penduduk miskin dengan menggunakan metode K-Means Clustering. Data yang digunakan meliputi jumlah penduduk, jumlah penduduk miskin, serta indikator tambahan seperti PDB per kapita dan rasio kemiskinan. Metode Principal Component Analysis (PCA) diterapkan untuk memvisualisasikan hasil pengelompokan dalam ruang dua dimensi. Hasilnya menunjukkan pola pengelompokan negara atau wilayah yang serupa dalam hal kemiskinan, yang dapat digunakan untuk menetapkan prioritas kebijakan penanggulangan kemiskinan di kawasan Asia. Pengelompokan dilakukan dengan menggunakan algoritma K-Means, sedangkan PCA menggunakan visualisasi, menggambarkan distribusi kluster dalam ruang dua dimensi. Tiga kluster utama mewakili wilayah dengan karakteristik sosial ekonomi yang beragam dan spesifik. Kluster pertama ditandai oleh wilayah dengan populasi yang tinggi tetapi tingkat kemiskinan yang rendah dan akses pendidikan yang memadai. Kluster kedua menggambarkan wilayah dengan persentase penduduk miskin yang signifikan. Sementara itu, kluster ketiga menunjukkan akses pendidikan yang terbatas, serta jumlah sekolah dan guru yang relatif sedikit. Visualisasi kluster ini menyoroti perbedaan antarwilayah dan berfungsi sebagai dasar untuk pembuatan kebijakan yang lebih baik di bidang sosial, ekonomi, dan pendidikan.*

### PENDAHULUAN

Kemiskinan tetap menjadi salah satu tantangan terbesar yang dihadapi oleh banyak negara di Asia (Sari dkk., 2021), termasuk Indonesia (Arifin, 2020). Meskipun kemajuan signifikan telah dicapai dalam beberapa sektor, ketimpangan sosial-ekonomi yang ada di berbagai wilayah masih terasa. Banyak daerah ini yang masih menghadapi kesulitan dalam mengakses sumber daya yang cukup memperbaiki kualitas hidup warga. Pemerintah dan organisasi internasional membutuhkan pemetaan yang lebih akurat untuk mengalokasikan sumber daya secara efektif dan merumuskan kebijakan yang tepat guna (Deni, 2023). Penting untuk memiliki data yang mendalam tentang kondisi sosial-ekonomi, termasuk tingkat kemiskinan, yang digunakan sebagai dasar perumusan

kebijakan. Untuk itu, pendekatan berbasis data seperti klasterisasi menjadi relevan.

Pengelompokan wilayah atau negara berdasarkan kesamaan karakteristik sosial-ekonomi dapat memberikan gambaran yang lebih jelas mengenai pola distribusi kemiskinan dan akses terhadap sumber daya. Ketimpangan jumlah penduduk, akses pendidikan, serta tingkat kemiskinan yang bervariasi antar wilayah memerlukan analisis yang mendalam untuk memahami hubungan antar faktor-faktor tersebut. Dalam konteks ini, maka metode klasterisasi seperti algoritma *K-Means* menjadi berguna (Aprilia & Sembiring, 2021). Dengan *K-Means*, dapat mengelompokkan negara atau wilayah berdasarkan indikator-indikator tertentu, seperti jumlah penduduk, jumlah penduduk miskin, rasio partisipasi sekolah, serta ketersediaan fasilitas pendidikan. Teknik ini yang memungkinkan analisis yang lebih objektif dan memberikan wawasan yang lebih baik mengenai distribusi kemiskinan dan akses pendidikan di berbagai daerah (Fajri dkk., 2025).

Selain itu, untuk mempermudah interpretasi dan visualisasi data yang memiliki dimensi yang banyak, metode *Principal Component Analysis* (PCA) ini digunakan (Akbar dkk., 2023). PCA membantu mengurangi dimensi data yang besar dan kompleks sehingga hasil klasterisasi dapat divisualisasikan dalam ruang dua dimensi. Visualisasi ini berguna untuk mampu memudahkan pemahaman terhadap struktur data yang ada dan mengidentifikasi pola-pola tersembunyi yang mungkin tidak terlihat pada analisis biasa. Hasil visualisasi yang lebih jelas akan memperlihatkan dengan lebih gamblang perbedaan antar klaster dan memberikan gambaran yang konkret mengenai karakteristik setiap kelompok wilayah.

Melalui pendekatan data mining seperti *K-Means Clustering* dan visualisasi PCA, penelitian ini bertujuan untuk memberikan kontribusi dalam pemahaman tentang struktur sosial dan ekonomi di Asia serta di wilayah Indonesia. Dengan cara ini, yang diharapkan penelitian dapat mendukung pengambilan keputusan yang lebih berbasis data pada tingkat pemerintah daerah dan organisasi terkait dalam merancang kebijakan yang lebih efektif dalam penanggulangan kemiskinan dan peningkatan kualitas pendidikan. Tujuan dari penelitian ini adalah untuk mengklasterisasi negara-negara Asia dan wilayah Indonesia berdasarkan tingkat kemiskinan, menentukan klaster yang terbentuk dan karakteristik masing-masing wilayah dalam setiap klaster, serta menyajikan hasil visualisasi dua dimensi dari klaster yang terbentuk menggunakan teknik PCA.

## LANDASAN TEORI

*Data mining* merupakan proses penting dalam analisis data yang besar, yang bertujuan untuk mencari pola-pola tersembunyi atau pengetahuan baru yang dapat memberikan wawasan berharga bagi pengambilan keputusan (Hendrastuty, 2024). *Data mining* mencakup berbagai teknik yang dapat diterapkan pada data yang besar dan kompleks, seperti klasifikasi, klasterisasi, asosiasi, dan deteksi anomali. Dalam konteks penelitian ini, *data mining* digunakan mengelompokkan wilayah berdasarkan kemiripan karakteristik demografi dan sosial, dengan harapan dapat menghasilkan pemahaman yang lebih mendalam mengenai pola-pola sosial-ekonomi yang ada. Melalui teknik ini, informasi tersembunyi dalam kumpulan data besar dapat diungkap, memungkinkan peneliti menemukan hubungan yang mungkin tidak terlihat melalui analisis tradisional.

Klasterisasi adalah salah satu teknik utama *data mining* (Muliono & Sembiring, 2019), yang berfokus pada pengelompokan data ke dalam beberapa grup atau klaster berdasarkan kemiripan tertentu di antara data tersebut. Tujuan utama klasterisasi adalah untuk memaksimalkan kesamaan dalam klaster (*intra-cluster*) dan meminimalkan kesamaan antar

klaster (*inter-cluster*). Klasterisasi ini membantu mengidentifikasi kelompok-kelompok data yang memiliki karakteristik serupa, yang dapat digunakan untuk analisis lebih lanjut atau pengambilan keputusan. Salah satu algoritma yang paling sering digunakan dalam klasterisasi adalah *K-Means*, mengelompokkan data dalam  $k$  klaster berdasar kedekatan terhadap *centroid* (titik pusat) dari masing-masing klaster. Algoritma ini yang bekerja dengan iteratif, melakukan penyesuaian pada pembagian klaster hingga pembagian tersebut stabil. *K-Means* sangat efektif digunakan pada *dataset* besar yang terdiri dari variabel numerik, sehingga cocok untuk diterapkan dalam penelitian ini.

Dalam penelitian ini, penggunaan algoritma *K-Means* bertujuan mengelompokkan wilayah-wilayah yang memiliki kemiripan karakteristik sosial-ekonomi berdasarkan data yang tersedia, seperti tingkat kemiskinan, kependudukan, dan pendidikan. Proses klasterisasi ini memerlukan iterasi berulang untuk memastikan bahwa data dikelompokkan dengan benar dan konsisten. *K-Means* memanfaatkan kedekatan antara data dengan *centroid* klaster untuk menentukan kelompok yang paling tepat bagi setiap titik data. Hasil dari klasterisasi ini diharapkan dapat memberikan pemahaman yang lebih baik mengenai distribusi kemiskinan dan pendidikan di berbagai wilayah, serta membantu dalam merumuskan kebijakan yang lebih tepat sasaran untuk setiap kelompok wilayah yang teridentifikasi. Meskipun *K-Means* efektif mengelompokkan data, dalam banyak kasus, data yang terlibat memiliki dimensi yang tinggi (Handoko dkk., 2025), sehingga diperlukan teknik lain untuk menyederhanakan data tanpa mengorbankan informasi penting.

Untuk menyederhanakan kompleksitas data memiliki banyak dimensi, *Principal Component Analysis* (PCA) digunakan dalam penelitian ini. PCA adalah teknik reduksi dimensi yang bertujuan untuk mengubah data yang memiliki banyak variabel menjadi beberapa komponen utama yang tetap mampu menjelaskan sebagian besar variabilitas dalam data tersebut (Carudin dkk., 2024). Teknik ini berguna ketika data yang ada mengandung variabel yang saling berkorelasi, yang dapat membingungkan analisis. Dengan PCA, hasil klasterisasi yang diperoleh dari *K-Means* dapat divisualisasikan dalam bentuk dua dimensi, sehingga memudahkan interpretasi dan analisis. Hal ini tidak hanya membantu peneliti untuk melihat seberapa baik klaster-klaster terbentuk, tetapi juga memberikan gambaran yang lebih jelas mengenai hubungan antar klaster yang ada. Dengan visualisasi yang dihasilkan oleh PCA, diharapkan diperoleh wawasan yang lebih mudah dipahami, sehingga dapat diambil keputusan yang lebih tepat dalam upaya penanggulangan kemiskinan dan perbaikan kualitas pendidikan di berbagai wilayah.

Penelitian ini juga melibatkan beberapa indikator sosial yang relevan dalam menggambarkan kondisi sosial-ekonomi di berbagai wilayah. Indikator kependudukan, yang mencakup jumlah penduduk laki-laki dan perempuan, menjadi penting untuk menggambarkan distribusi populasi di setiap wilayah. Selain itu, untuk jumlah penduduk miskin, rasio kemiskinan, dan PDB per kapita digunakan untuk bisa menggambarkan keadaan sosial-ekonomi negara atau wilayah. Pendidikan, sebagai salah satu faktor penting dalam peningkatan kualitas hidup, juga menjadi indikator utama dalam penelitian ini, yang diukur melalui jumlah sekolah, tenaga pengajar, serta angka partisipasi kasar (APK). Indikator-indikator ini saling terkait dan memberi gambaran menyeluruh mengenai tantangan sosial-ekonomi yang dihadapi setiap wilayah. Pengelompokan wilayah ini berdasarkan indikator-indikator ini diharapkan membantu merumuskan kebijakan yang tepat.

## **METODE PENELITIAN**

Penelitian ini yang menggunakan pendekatan secara kuantitatif dengan tujuan eksplorasi

untuk menemukan pola atau kelompok dalam data yang berkaitan dengan karakteristik wilayah, khususnya dalam aspek demografis dan pendidikan. Teknik yang digunakan dalam penelitian ini adalah *data mining*, dengan fokus algoritma *unsupervised learning*, yakni *K-Means Clustering*. Metode ini memungkinkan untuk mengelompokkan data wilayah berdasarkan kemiripan atribut-atribut tertentu, seperti jumlah penduduk, tingkat kemiskinan, dan fasilitas pendidikan. Hasil dari klusterisasi kemudian divisualisasikan dengan *Principal Component Analysis (PCA)* mengurangi dimensi data, sehingga memudahkan dalam interpretasi dan analisis.

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari instansi pemerintah serta sumber terpercaya lain. Jenis data yang digunakan mencakup beberapa variabel penting, antara lain jumlah total penduduk, jumlah penduduk miskin, jumlah sekolah, jumlah tenaga pengajar, serta data persebaran wilayah administratif. Data ini untuk memberikan gambaran yang lebih jelas mengenai karakteristik sosial ekonomi dan pendidikan di berbagai wilayah, yang menjadi fokus penelitian ini. Data sekunder ini diambil dari berbagai sumber yang sudah teruji kredibilitasnya, seperti data dari BPS dan instansi terkait lainnya.

Pengumpulan data ini dilakukan dengan menggunakan metode dokumentasi, yaitu dengan pencatatan dan pengumpulan data dari berbagai sumber, seperti buku yang tersedia di Perpustakaan Nasional serta sumber-sumber lainnya yang relevan. Data yang diperoleh akan diproses lebih lanjut dengan tahapan *pre-processing*. Pada tahapan ini, data akan dibersihkan dari nilai yang kosong atau *outlier* yang bisa memengaruhi akurasi hasil analisis. Selanjutnya, data akan dinormalisasi untuk memastikan skala variabel yang digunakan dalam analisis berada pada level yang sebanding, sehingga tidak ada variabel yang mendominasi hasil analisis.

Setelah data siap, langkah selanjutnya adalah melakukan klusterisasi menggunakan algoritma *K-Means*. Proses ini diawali dengan menentukan nilai *k*, yang menunjukkan jumlah kluster yang akan dibuat. Kemudian, algoritma *K-Means* mengelompokkan data berdasarkan kedekatannya dengan *centroid* masing-masing kluster. Berikut contoh kode implementasi:

```
kmeans = KMeans(n_clusters=3, random_state=42)
df["Cluster"] = kmeans.fit_predict(fitur_scaled)
centroids_scaled = kmeans.cluster_centers_
centroids_original = scaler.inverse_transform(centroids_scaled)
```

- n\_clusters=3** : algoritma akan mencoba mencari 3 kelompok (*cluster*) dalam data.
- random\_state=42** : memastikan hasil pengelompokan akan konsisten setiap kali menjalankan kode (untuk tujuan re-produktivitas).
- kmeans\_2d.cluster\_centers\_** : berisi koordinat pusat (*centroid*) dari setiap kluster dalam skala yang telah disesuaikan.
- scaler.inverse\_transform()** : digunakan untuk mengembalikan koordinat *centroid* ke skala aslinya (sebelum dilakukan pen-skala-an).

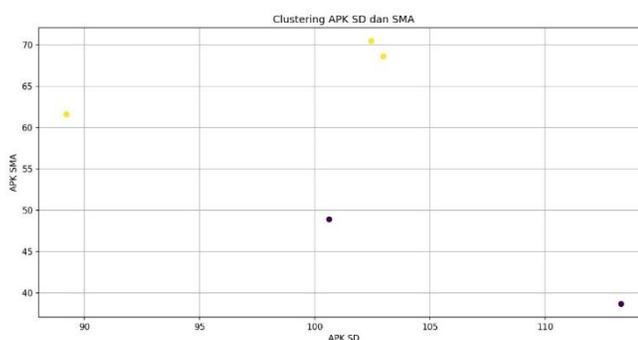
Setelah klusterisasi selesai, langkah berikutnya adalah melakukan visualisasi dengan *PCA*. *PCA* akan digunakan untuk mengurangi dimensi data dan memvisualisasikan hasil klusterisasi dalam bentuk *scatter plot* dua dimensi. Tahapan terakhir menganalisis dan menginterpretasikan hasil klusterisasi, untuk mengetahui karakteristik sosial-ekonomi dari setiap kluster wilayah yang terbentuk. Berdasarkan analisis ini, peneliti memberikan deskripsi lanjut wilayah yang termasuk dalam setiap kluster dan faktor-faktor yang mempengaruhi perbedaan antar kluster.

## HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini terdiri atas beberapa variabel numerik, yaitu jumlah total penduduk, PDB per kapita, jumlah penduduk miskin, jumlah sekolah, dan jumlah tenaga pengajar. Sumber data diperoleh dari instansi resmi seperti BPS dan dinas terkait, yang memastikan kredibilitasnya. Setelah pengumpulan data, tahap selanjutnya adalah pembersihan data dari nilai kosong (*missing values*) terdapat pada beberapa variabel. Kemudian, dilakukan proses normalisasi untuk memastikan bahwa seluruh variabel berada dalam skala yang sebanding, agar hasil analisis tidak dipengaruhi perbedaan skala antar variabel. Setelah data diproses, *file* tersebut disimpan dalam format CSV mempermudah pengolahan dan analisis lebih lanjut peneliti.

### 1. Hasil Klasterisasi dengan *K-Means*

Proses klasterisasi dilakukan dengan menggunakan algoritma *K-Means*. Berdasarkan proses eksplorasi, jumlah klaster optimal ditentukan sebanyak tiga ( $k = 3$ ) menggunakan metode visual, seperti *elbow method* atau *silhouette score*. Setiap wilayah administratif, seperti kabupaten atau kecamatan, berhasil dikelompokkan ke dalam salah satu dari tiga klaster berdasarkan kemiripan indikator demografi, kemiskinan, dan pendidikan. Karakteristik tiap klaster adalah sebagai berikut: Klaster 1 merupakan wilayah jumlah penduduk besar dan tingkat kemiskinan rendah, Klaster 2 terdiri wilayah dengan jumlah penduduk miskin relatif tinggi, sementara Klaster 3 adalah wilayah dengan jumlah sekolah dan guru rendah, menunjukkan akses pendidikan terbatas.



**Gambar 1. *K-Means* Klastering Penduduk**

Hasil klasterisasi menunjukkan perbedaan signifikan dalam karakteristik sosial dan ekonomi antar wilayah. Klaster 1 mencerminkan wilayah yang memiliki potensi ekonomi yang lebih baik dan akses ke sumber daya yang lebih baik, termasuk pendidikan. Klaster 2 ini mengindikasikan wilayah dengan tantangan dalam pengentasan kemiskinan, yang dapat mempengaruhi berbagai aspek pembangunan. Sementara itu, Klaster 3 ini menggambarkan wilayah mungkin memerlukan perhatian lebih dalam hal peningkatan infrastruktur pendidikan untuk bisa meningkatkan akses dan kualitas pendidikan bagi penduduknya.

### 2. Visualisasi Klaster dengan *Principal Component Analysis (PCA)*

Untuk memudahkan interpretasi dan memahami distribusi klaster, hasil klasterisasi yang divisualisasikan menggunakan *Principal Component Analysis (PCA)*. PCA berfungsi mereduksi dimensi data yang awalnya terdiri dari banyak variabel menjadi dua komponen utama, sehingga memudahkan representasi dalam bentuk *scatter plot*. Dalam visualisasi tersebut, hasil klasterisasi terlihat dengan pemisahan yang jelas antar klaster. Klaster 1 cenderung terkonsentrasi di satu area plot, merepresentasikan wilayah memiliki indikator demografi yang kuat dan tingkat kemiskinan yang terkendali. Klaster 2 tersebar di area lain pada plot, yang mencerminkan keragaman wilayah dengan tingkat kemiskinan yang relatif tinggi. Sementara itu, Klaster 3 yang terlihat terpisah dan menunjukkan wilayah yang memiliki keterbatasan dalam hal akses pendidikan, dengan jumlah

sekolah dan tenaga pengajar yang rendah.

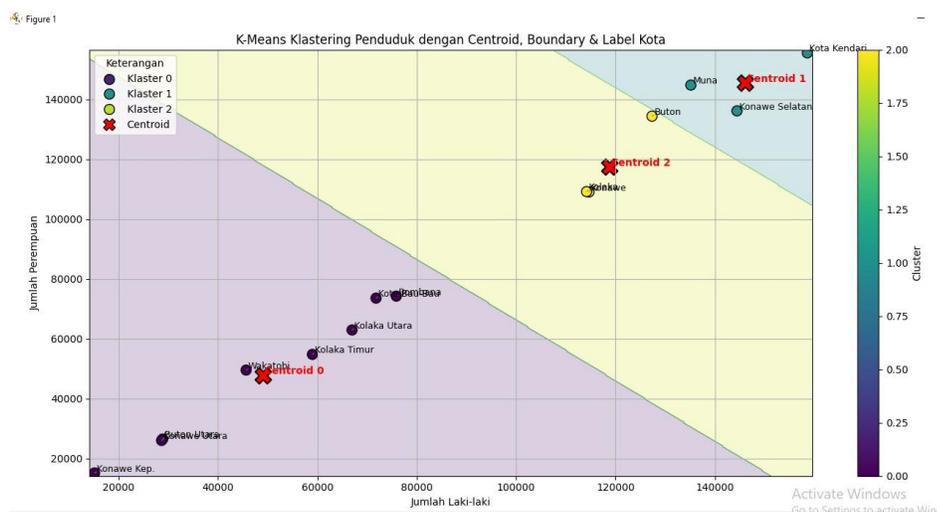


**Gambar 2. Klastering Wilayah DIY Berdasarkan Tingkat Pendidikan**

Hasil visualisasi *scatter plot* ini mengungkapkan ketiga klaster memiliki karakteristik sosial-ekonomi berbeda dan saling berhubungan dengan faktor kemiskinan, pendidikan, dan demografi. Klaster 1, yang terkonsentrasi di satu area, menggambarkan wilayah yang relatif lebih maju dengan kondisi demografi yang baik dan tingkat kemiskinan yang rendah. Wilayah dalam klaster ini memiliki potensi untuk mengembangkan berbagai sektor lainnya, termasuk sektor pendidikan dan infrastruktur. Sebaliknya, Klaster 2 yang tersebar lebih luas menunjukkan adanya wilayah dengan keragaman sosial yang lebih tinggi, khususnya dengan tingkat kemiskinan lebih tinggi. Wilayah dalam klaster ini mungkin memerlukan perhatian besar dalam hal perencanaan kebijakan yang berfokus pengentasan kemiskinan dan peningkatan kesejahteraan sosial. Sementara itu, Klaster 3 yang terpisah menggambarkan wilayah yang lebih tertinggal, terutama dalam hal akses pendidikan. Kurangnya sekolah dan tenaga pengajar di wilayah ini menjadi tantangan utama yang harus diatasi agar wilayah tersebut berkembang secara optimal. Hasil klasterisasi ini memberikan gambaran yang jelas dan dapat digunakan sebagai dasar dalam merancang kebijakan pembangunan yang lebih terfokus dan sesuai dengan kondisi masing-masing klaster.

### 3. Visualisasi Klaster Jumlah Penduduk dengan Metode *K-Means*

Untuk memahami pola sebaran jumlah penduduk berdasarkan jenis kelamin di wilayah Sulawesi Tenggara, dilakukan klasterisasi menggunakan algoritma *K-Means*. Visualisasi hasil klaster kemudian ditampilkan dalam bentuk *scatter plot* 2 dimensi dengan penambahan *centroid*, *boundary* antar klaster, serta label nama kota. Hasil visualisasi menunjukkan bahwa Klaster 0 (berwarna ungu) terpusat di bagian bawah kiri, mencakup kota/kabupaten dengan jumlah penduduk laki-laki dan perempuan yang rendah, seperti Konawe Kepulauan, Buton Utara, dan Kolaka Timur. Klaster 1 (berwarna biru kehijauan) berada di kanan atas, merepresentasikan daerah dengan jumlah penduduk yang besar, seperti Kota Kendari, Konawe Selatan, dan Muna. Sedangkan Klaster 2 (berwarna kuning kehijauan) yang tersebar di tengah, mencerminkan wilayah dengan penduduk menengah, contohnya Kolaka, Buton, dan Konawe.

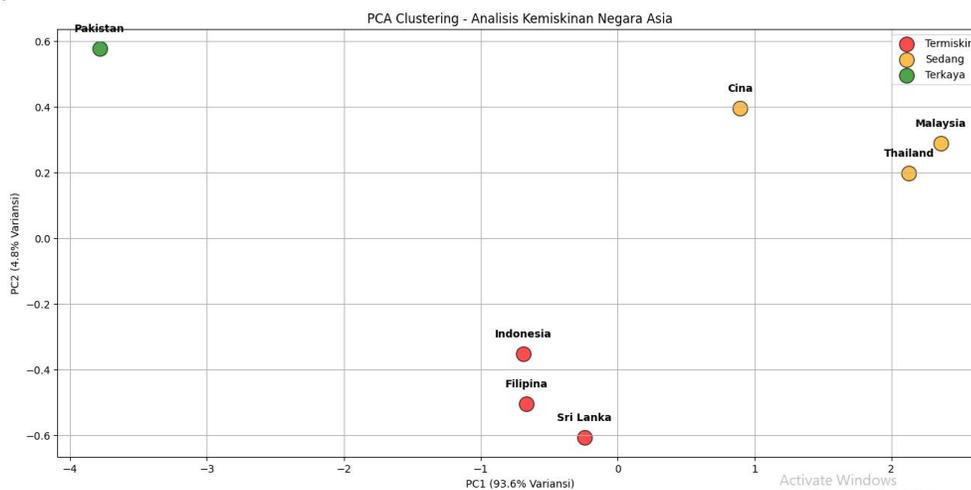


**Gambar 3. K-Means Klustering Penduduk**

Hasil visualisasi ini menunjukkan perbedaan mencolok dalam pola sebaran jumlah penduduk di wilayah Sulawesi Tenggara. Klaster 0, mencakup wilayah dengan jumlah penduduk rendah, menunjukkan bahwa daerah-daerah ini mungkin memiliki tingkat urbanisasi yang rendah dan akses ke fasilitas dasar yang terbatas. Klaster 1, yang menunjukkan wilayah dengan jumlah penduduk besar, mencerminkan daerah yang lebih padat dan mungkin memiliki infrastruktur serta layanan yang lebih baik, seperti di Kota Kendari dan sekitarnya. Klaster 2, tersebar di tengah, menunjukkan wilayah dengan populasi menengah, yang kemungkinan berada dalam fase perkembangan dengan potensi pertumbuhan lebih lanjut. Hasil ini memberikan wawasan yang berguna bagi perencanaan kebijakan dan pengembangan wilayah, khususnya meningkatkan pemerataan distribusi penduduk dan pengembangan infrastruktur.

#### 4. Analisis Kemiskinan Negara Asia

Penelitian ini mengelompokkan negara-negara Asia ke dalam tiga klaster utama berdasarkan indikator sosial-ekonomi yang relevan menggunakan algoritma K-Means, dan memvisualisasikan hasilnya dengan *Principal Component Analysis* (PCA). Temuan ini selaras tujuan penelitian untuk memberikan pemahaman yang lebih baik tentang struktur sosial-ekonomi negara-negara tersebut. Implikasi hasil klasterisasi menunjukkan bahwa Klaster Merah (Termiskin) mencakup negara-negara seperti Sri Lanka, Indonesia, dan Filipina, yang cenderung memiliki GNP per kapita yang lebih rendah dan harapan hidup yang moderat. Negara-negara dalam klaster ini yang mungkin menghadapi tantangan dalam hal kemiskinan atau akses terhadap sumber daya dasar, meskipun tidak seberat klaster hijau. Klaster Kuning (Sedang) mencakup Cina, Malaysia, dan Thailand, yang menunjukkan kondisi ekonomi yang lebih maju dengan GNP per kapita yang lebih tinggi dan harapan hidup yang baik. Sementara itu, Klaster Hijau (Terkaya) diwakili oleh Pakistan, yang meskipun dilabeli 'Terkaya' berdasarkan visualisasi, sebenarnya memiliki GNP per kapita terendah, tingkat kematian anak tertinggi, dan masalah gizi yang paling signifikan. Hal ini menunjukkan adanya ketidakselarasan antara label visual dan data aktual. Hasil visualisasi PCA membantu dalam mengidentifikasi wilayah-wilayah yang memiliki kesamaan pola berdasarkan data ekonomi dan demografi utama, sehingga memberi dasar yang kuat bagi pemerintah daerah merancang kebijakan yang spesifik dan efektif, disesuaikan kebutuhan unik masing-masing klaster wilayah.



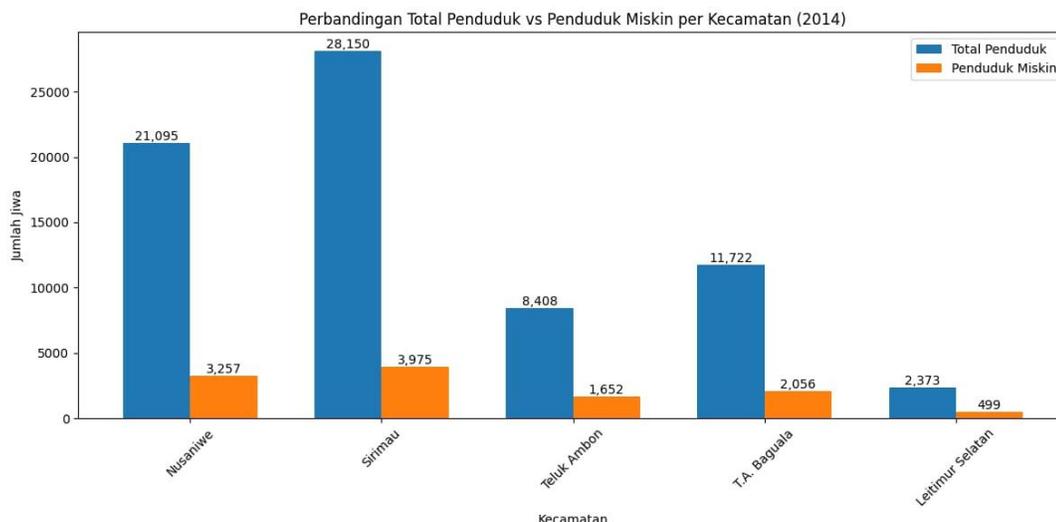
**Gambar 4. Klastering Analisis Kemiskinan Negara Asia**

Hasil klasterisasi menunjukkan terdapat perbedaan signifikan dalam perkembangan sosial-ekonomi antar negara-negara Asia yang dikelompokkan. Klaster Merah menggambarkan negara-negara dengan tantangan besar dalam hal pengentasan kemiskinan dan peningkatan kualitas hidup, sehingga membutuhkan perhatian lebih dalam pengelolaan sumber daya dan distribusi kekayaan. Klaster Kuning, yang terdiri negara-negara dengan ekonomi yang stabil dan maju, menunjukkan bahwa mereka berada pada posisi yang lebih baik dalam meningkatkan kualitas hidup warganya, meskipun masih memiliki ruang untuk perbaikan. Sementara itu, Klaster Hijau yang diwakili oleh Pakistan menunjukkan bahwa meskipun negara tersebut masuk dalam kategori kaya menurut label visual, data aktualnya mengungkapkan adanya ketidakselarasan dalam indikator sosial-ekonomi, yang memerlukan kebijakan yang holistik dan berbasis data untuk bisa mengatasi masalah-masalah mendasar seperti kemiskinan, kesehatan, dan gizi. Keseluruhan hasil klasterisasi ini memberikan gambaran yang lebih jelas tentang bagaimana negara-negara Asia mengidentifikasi kekuatan dan tantangan mereka, serta merancang kebijakan yang lebih terfokus pada perbaikan kondisi sosial-ekonomi yang spesifik bagi masing-masing klaster.

## 5. Pembahasan

Penelitian ini berhasil mengelompokkan wilayah di Indonesia ke dalam tiga klaster utama berdasarkan indikator demografi, kemiskinan, dan pendidikan menggunakan algoritma K-Means, dengan visualisasi hasil menggunakan PCA. Temuan ini selaras dengan tujuan penelitian untuk memberikan pemahaman yang lebih baik tentang struktur sosial-ekonomi wilayah di Indonesia. Implikasi hasil klasterisasi menunjukkan bahwa Klaster 1 (Wilayah Populasi Tinggi, Kemiskinan Rendah) dapat menjadi fokus pengembangan lebih lanjut dalam sektor ekonomi dan investasi, mengingat potensi sumber daya manusia yang besar dan tingkat kesejahteraan yang relatif baik. Kebijakan di klaster ini yang dapat diarahkan pada peningkatan kualitas hidup dan keberlanjutan pembangunan. Klaster 2 (Wilayah Kemiskinan Tinggi) mencakup wilayah-wilayah memerlukan perhatian khusus dalam program pengentasan kemiskinan, di mana kebijakan yang tepat sasaran, seperti program bantuan sosial, peningkatan akses lapangan kerja, dan pelatihan keterampilan, sangat diperlukan meningkatkan taraf hidup penduduk. Klaster 3 (Wilayah Akses Pendidikan Terbatas) menyoroti daerah-daerah yang membutuhkan intervensi di bidang pendidikan, dengan prioritas pada peningkatan jumlah sekolah dan tenaga pengajar, serta program peningkatan angka partisipasi sekolah guna pemerataan akses pendidikan. Hasil visualisasi PCA ini juga membantu dalam mengidentifikasi wilayah yang tumpang tindih atau mempunyai karakteristik campuran,

meskipun secara umum ketiga klaster menunjukkan pemisahan yang jelas. Ini memberikan dasar yang kuat bagi pemerintah daerah untuk merancang kebijakan yang spesifik dan efektif, bukan hanya secara umum, tetapi juga disesuaikan kebutuhan masing-masing klaster wilayah.



**Gambar 5. Perbandingan Total Penduduk vs Penduduk Miskin Per Kecamatan**

Interpretasi hasil klasterisasi ini menunjukkan bahwa terdapat perbedaan signifikan dalam kondisi sosial-ekonomi antar wilayah di Indonesia, yang membutuhkan pendekatan kebijakan yang disesuaikan dengan karakteristik masing-masing klaster. Klaster 1, yang memiliki populasi tinggi dan kemiskinan rendah, menunjukkan potensi besar untuk pengembangan ekonomi dan investasi, sehingga kebijakan dapat difokuskan pada peningkatan kualitas hidup dan pengelolaan sumber daya manusia secara berkelanjutan. Di sisi lain, Klaster 2, yang memiliki tingkat kemiskinan tinggi, membutuhkan perhatian lebih dalam upaya pengentasan kemiskinan dengan kebijakan yang dapat meningkatkan akses terhadap pekerjaan dan peluang ekonomi. Sementara Klaster 3, menunjukkan wilayah dengan keterbatasan akses pendidikan, memerlukan program khusus untuk memperbaiki infrastruktur pendidikan dan meningkatkan kualitas serta kuantitas tenaga pengajar. Temuan ini memberikan arahan yang jelas bagi pemerintah daerah dalam merancang kebijakan yang lebih tepat sasaran dan efektif, dengan mempertimbangkan kondisi unik dari setiap klaster wilayah.

## KESIMPULAN

Penelitian ini yang mengelompokkan wilayah di Indonesia berdasarkan data kependudukan, kemiskinan, dan pendidikan menggunakan metode K-Means yang divisualisasikan dengan *Principal Component Analysis* (PCA). Tiga klaster utama yang diidentifikasi, merepresentasikan wilayah dengan karakteristik sosial-ekonomi yang berbeda: Klaster pertama dicirikan oleh wilayah dengan populasi tinggi namun tingkat kemiskinan rendah, Klaster kedua menggambarkan wilayah dengan persentase penduduk miskin yang signifikan, dan Klaster ketiga menunjukkan wilayah dengan akses pendidikan yang terbatas. Visualisasi klaster menggunakan PCA efektif dalam menyoroti perbedaan antar wilayah, yang dapat dijadikan dasar untuk perumusan kebijakan yang lebih baik di bidang sosial dan pendidikan. Berdasarkan hasil penelitian ini, maka beberapa saran diajukan, antara lain untuk pemerintah, hasil klasterisasi ini dapat menjadi panduan awal untuk merumuskan kebijakan yang lebih terarah dan spesifik sesuai

dengan kebutuhan masing-masing klaster wilayah, baik dalam program pengentasan kemiskinan maupun peningkatan akses dan mutu pendidikan. Saran untuk penelitian selanjutnya meliputi mengintegrasikan indikator lain seperti kesehatan, infrastruktur, atau pertumbuhan ekonomi untuk mendapatkan klasterisasi yang lebih komprehensif, mencoba algoritma klasterisasi lain seperti *Hierarchical Clustering* atau DBSCAN untuk membandingkan hasilnya dan mencari klasterisasi yang lebih optimal, menerapkan analisis temporal untuk melihat perubahan karakteristik klaster dari waktu ke waktu, serta memperluas cakupan data agar mencakup seluruh provinsi di Indonesia untuk mendapatkan gambaran yang lebih representatif.

### DAFTAR REFERENSI

- Akbar, M. N., Salsabila, A., Asri, A. P., & Syawir, M. (2023). Analisis clustering untuk segmentasi pengguna kartu kredit dengan menggunakan algoritma K-Means dan Principal Component Analysis. *AGENTS: Journal of Artificial Intelligence and Data Science*, 3(1), 16-24. <https://doi.org/10.24252/jagti.v3i1.56>
- Aprilia, K., & Sembiring, F. (2021). Analisis Garis Kemiskinan Makanan Menggunakan Metode Algoritma K-Means Clustering. Dalam *Prosiding Seminar Nasional Sistem Informasi Dan Manajemen Informatika Universitas Nusa Putra* (Vol. 1, hlm. 1-10). <https://sismatik.nusaputra.ac.id/index.php/sismatik/article/view/1>
- Arifin, J. (2020). Budaya kemiskinan dalam penanggulangan kemiskinan di Indonesia. *Sosio Informa*, 6(2), 114-132. <https://doi.org/10.33007/inf.v6i2.2372>
- Carudin, C., Marisa, M., Murnawan, M., Reba, F., Koibur, M. E., Thantawi, A. M., Halim, A., & Wattimena, F. Y. (2024). *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia.
- Deni, A. (2023). *Kebijakan Publik (Formulasi, Implementasi, Evaluasi, dan Analisis)*. Cendikia Mulia Mandiri.
- Fajri, A., Fathudin, M. A., & Zein, R. A. (2025). Analisis Klasterisasi Dinamika Sosial Ekonomi 2022. Dalam *Gunung Djati Conference Series* (Vol. 50, hlm. 207-216). <https://conferences.uinsgd.ac.id/index.php/gdcs/article/view/2662>
- Handoko, K., Astiti, S., Novrian, W., Putra, T. N., & Darmansah, D. (2025). *K-Means Clustering Menggunakan Aplikasi Rapidminer*. PT. Sonpedia Publishing Indonesia.
- Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika dan Ilmu Komputer (JIMA-ILKOM)*, 3(1), 46-56. <https://doi.org/10.58602/jima-ilkom.v3i1.26>
- Muliono, R., & Sembiring, Z. (2019). Data Mining Clustering Menggunakan Algoritma K-Means Untuk Klasterisasi Tingkat Tridarma Pengajaran Dosen. *CESS (Journal of Computer Engineering, System and Science)*, 4(2), 2502-2714. <https://doi.org/10.24114/cess.v4i2.13620>
- Sari, W. I., Nasution, L. N., & Novalina, A. (2021). Analisis leading indicator kebijakan moneter dalam mengatasi kemiskinan di 5 negara Asia Tenggara. *Jurnal Kajian Ekonomi dan Kebijakan Publik (JEpa)*, 6(2), 610-618. <https://jurnal.pancabudi.ac.id/index.php/jepa/article/view/3862>