



Klasterisasi Pembangunan Manusia di Indonesia Menggunakan Algoritma *K-Means* Berdasarkan IPM dan Literasi Pendidikan

Dona Oktaviani Habiby^{1*}

Universitas Singaperbangsa Karawang, 2310631050075@student.unsika.ac.id

Karunia Eka Lestari²

Universitas Singaperbangsa Karawang, karunia@fkip.unsika.ac.id

ABSTRAK

Penelitian ini bertujuan untuk mengelompokkan 38 provinsi di Indonesia berdasarkan Indeks Pembangunan Manusia (IPM) yang dikombinasikan dengan variabel pendukung literasi pendidikan yaitu tingkat kegemaran membaca masyarakat (TKM), tingkat penyelesaian pendidikan SMA/Sederajat (TPP), dan rata-rata lama sekolah (RLS). Metode penelitian yang digunakan adalah kuantitatif deskriptif dengan algoritma *k-means clustering*. Validasi hasil pengelompokan diuji menggunakan *silhouette method*, yaitu teknik pengukuran kualitas kluster berdasarkan kedekatan jarak antar objek di dalam satu kluster dan jarak antar kluster yang berbeda. Hasil analisis dengan $k = 3$ membagi wilayah ke dalam kategori tinggi, sedang, dan rendah. Meskipun nilai rata-rata *silhouette* tertinggi diperoleh pada $k = 2$ (0,74), pemilihan $k = 3$ (0,33) dilakukan karena memberikan hasil yang lebih representatif secara substantif dalam menggambarkan sebaran pembangunan manusia di Indonesia. Hasil ini menjadi acuan penting bagi pemangku kebijakan untuk melakukan intervensi yang lebih spesifik pada daerah-daerah dengan literasi rendah.

Kata kunci: IPM, *K-means clustering*, Literasi pendidikan, *Silhouette method*

PENDAHULUAN

Pembangunan manusia merupakan aspek utama dalam upaya meningkatkan kesejahteraan masyarakat dan kualitas sumber daya manusia. Indeks Pembangunan Manusia (IPM) yang diinisiasi oleh UNDP pada 1990 menjadi ukuran pencapaian pembangunan yang berfokus pada tiga aspek utama: usia panjang dan kesehatan, wawasan, serta taraf hidup memadai (Nasution, 2019; Nurlina *et al.*, 2023). Menurut Badan Pusat Statistik (BPS, 2023), IPM digunakan untuk melihat sejauh mana masyarakat mampu menikmati dampak pembangunan dalam hal ekonomi, kesehatan, dan pembelajaran. Pernyataan tersebut konsisten dengan pandangan Sari *et al.* (2022) yang menyatakan bahwa IPM tidak hanya mencerminkan kemajuan ekonomi, tetapi juga menjadi ukuran keberhasilan pemerintah untuk mendorong taraf hidup warga. Di Indonesia, IPM digunakan secara luas sebagai indikator kinerja pembangunan daerah, serta menjadi dasar dalam perencanaan dan alokasi anggaran, termasuk dalam penentuan Dana Insentif Daerah (Puspita, 2021). Meskipun IPM Indonesia menunjukkan tren peningkatan setiap tahun, kesenjangan antarwilayah masih cukup tinggi, di mana daerah perkotaan cenderung memiliki nilai IPM lebih tinggi dibandingkan dengan daerah pedesaan (Rahmati, 2021).

Dalam konteks pembangunan manusia, pendidikan memiliki peran besar dalam peningkatan IPM. Sari *et al.* (2022) menegaskan bahwa tingkat penyelesaian pendidikan dan rata-rata lama sekolah merupakan faktor dominan yang memengaruhi capaian IPM di Indonesia. Selain itu, tingkat literasi masyarakat seperti tingkat kegemaran membaca (TKM) juga memberikan dampak pada mutu manusia yang berimplikasi pada peningkatan IPM.

Meskipun kajian mengenai IPM di Indonesia telah banyak dilakukan, namun sejauh ini belum banyak penelitian yang secara komprehensif mengintegrasikan IPM langsung menggabungkan dengan variabel literasi Pendidikan seperti (TKM) serta indikator ketuntasan pendidikan (TPP dan RLS) dalam satu model klasterisasi. Sebagian besar penelitian terdahulu cenderung membahas indikator tersebut secara terpisah. Padahal,

penggabungan variabel ini sangat krusial untuk memetakan kualitas sumber daya manusia secara lebih mendalam dan melihat sejauh mana kualitas budaya membaca berpengaruh pada capaian pendidikan di setiap provinsi.

Untuk menganalisis kesamaan karakteristik tersebut, pendekatan *data mining* digunakan karena mampu mengekstraksi pola tersembunyi dari data besar melalui tahapan *data selection*, *pre-processing/cleaning*, *transformation*, *data mining*, dan *interpretation/evaluation* (Luthfi dan Wijayanto, 2021). Salah satu teknik penambangan data yang sering dimanfaatkan dalam klasifikasi wilayah berdasarkan indikator sosial ekonomi adalah metode *k-means clustering*. Menurut Yudhanegara dan Lestari (2024), algoritma *k-means* membagi data ke dalam beberapa kelompok (*cluster*) dengan karakteristik serupa. Analisis kluster menjadi metode yang tepat untuk mengelompokkan wilayah berdasarkan kesamaan karakteristik indikator pembangunan manusia, sehingga memudahkan identifikasi kelompok wilayah dengan tingkat pembangunan yang seragam maupun yang membutuhkan perhatian khusus (Sakhi *et al.*, 2025). Metode ini efektif dalam mengidentifikasi pola data multivariat serta memberikan gambaran wilayah dengan tingkat pembangunan manusia yang relatif homogen (Sikana dan Wijayanto, 2021; Rahmati, 2021). Selain itu, temuan studi menyatakan bahwa teknik *k-means* dapat menghasilkan pemetaan wilayah yang lebih jelas dibandingkan metode kluster lainnya (Puspita, 2021), sehingga relevan digunakan untuk mendukung analisis kebijakan berbasis data.

Dengan demikian, studi ini memanfaatkan pendekatan klasterisasi *k-means* untuk mengkategorisasi 38 provinsi di Indonesia berdasarkan Indeks Pembangunan Manusia (IPM), tingkat kegemaran membaca (TKM), tingkat penyelesaian pendidikan SMA/ sederajat (TPP), dan rata-rata lama sekolah (RLS). Hasil analisis ditujukan untuk menyajikan ilustrasi yang lebih menyeluruh tentang derajat pembangunan manusia di Indonesia serta menjadi landasan dalam membangun strategi pembangunan yang lebih merata dan berkelanjutan.

METODE

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan metode analisis kluster (*cluster analysis*). Data yang digunakan adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS), mencakup 38 provinsi di Indonesia. Variabel yang digunakan meliputi Indeks Pembangunan Manusia (IPM), Tingkat Kegemaran Membaca (TKM), Tingkat Penyelesaian Pendidikan (TPP) SMA/ Sederajat, dan Rata-rata Lama Sekolah (RLS).

Dalam pelaksanaannya, proses pengelompokan data dilakukan melalui tahapan *Data Mining* sebagaimana dijelaskan oleh Luthfi dan Wijayanto (2021), yang terdiri dari lima tahap utama, yaitu *data selection*, *pre-processing/cleaning*, *transformation*, *data mining*, dan *interpretation/evaluation*. Tahapan-tahapan tersebut dijabarkan sebagai berikut:

1. *Data Selection*

Data dipilih dari laporan resmi BPS untuk seluruh provinsi di Indonesia. Informasi yang dikumpulkan mencakup empat indikator utama (IPM, TKM, TPP, dan RLS). Dalam kode program, data dimuat menggunakan *library (readxl)* untuk memastikan pembacaan format data yang tepat.

2. *Pre-processing/Cleaning*

Fase ini bertujuan untuk memastikan kualitas data sebelum diolah. Dilakukan verifikasi untuk memastikan tidak terdapat nilai yang hilang (*missing value*) menggunakan fungsi *na.omit()*. Hal ini menjamin bahwa setiap dari 38 provinsi memiliki data yang lengkap untuk dianalisis guna menghindari bias pada hasil klasterisasi.

3. *Transformation*

Mengingat setiap variabel memiliki rentang nilai yang berbeda, dilakukan proses

standarisasi atau normalisasi menggunakan fungsi *scale()*. Tahapan ini sangat krusial karena algoritma *K-Means* berbasis pada perhitungan jarak *Euclidean*, sehingga seluruh variabel harus berada pada skala yang sebanding agar variabel dengan nilai besar tidak mendominasi hasil akhir.

4. *Data Mining* (Proses Pengelompokan dengan *K-Means*)

Tahap ini merupakan inti analisis untuk mengelompokkan provinsi berdasarkan kemiripan karakteristik. Berdasarkan kode program, diterapkan parameter teknis berikut untuk menjamin validitas:

- Menentukan jumlah *cluster* (*k*) optimal: Menggunakan bantuan metode *Elbow* dan *Silhouette* melalui fungsi *fviz_nbclust()*. Peneliti menetapkan $k = 3$ untuk mengklasifikasikan wilayah ke dalam kategori Tinggi, Sedang, dan Rendah.
- Inisialisasi dan Stabilitas: Menetapkan *set.seed(123)* untuk mengunci titik pusat (*centroid*) awal yang dipilih secara acak agar hasil pengelompokan tetap konsisten saat simulasi diulang. Selain itu, digunakan parameter *nstart = 25* untuk menjalankan algoritma dengan 25 konfigurasi awal yang berbeda guna menemukan hasil yang paling stabil.
- Menaksir kedekatan tiap poin data terhadap setiap *centroid* dengan menggunakan rumus *Euclidean distance* menurut Kassambara, (2017) sebagai berikut.

$$d_{euc(x,y)} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_i - y_i)^2}$$

Atau dapat disederhanakan sebagai berikut.

$$d_{euc(x,y)} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}; i = 1, 2, 3, \dots, n$$

Keterangan:

$d_{Euclidan}$ merupakan jarak data ke pusat cluster (*centroid*).

x_i merupakan data pertama dari indikator pertama.

y_i merupakan data pertama dari indikator kedua

n merupakan jumlah pusat cluster (*centroid*)

- Alokasi Data: Menempatkan setiap provinsi ke dalam *cluster* yang memiliki jarak *Euclidean* terdekat dengan *centroid*-nya.
- Iterasi dan Konvergensi: Memperbarui posisi *centroid* secara berulang (*iterative*) berdasarkan rata-rata anggota *cluster* yang baru hingga posisi *centroid* stabil atau mencapai iterasi maksimum sebanyak 100 kali.

5. *Interpretation/Evaluation*

Tahap terakhir adalah mengevaluasi kekuatan klaster menggunakan *Silhouette Method* melalui fungsi (*fviz_silhouette*). Skor ini menjelaskan seberapa baik suatu objek berada di dalam klasternya. Selanjutnya, dilakukan pengurutan ulang (*reordering*) klaster berdasarkan rata-rata IPM tertinggi ke terendah, sehingga klaster 1 didefinisikan sebagai kategori Tinggi, klaster 2 sebagai Sedang, dan klaster 3 sebagai Rendah. Hasil ini kemudian divisualisasikan dalam bentuk grafik sebaran klaster untuk mempermudah interpretasi kebijakan.

HASIL DAN PEMBAHASAN

Setelah melalui rangkaian tahapan proses *Data Mining*, yang mencakup *data selection*, *pre-processing/cleaning*, *transformation*, *data mining*, dan *interpretation/evaluation*, hasil pengolahan data kemudian disajikan dan dibahas pada bagian berikut.

1. *Data Selection*

Informasi yang dipakai dalam riset ini adalah data sekunder dari BPS, dengan IPM sebagai variabel pokok pada 38 provinsi di Indonesia. Untuk memperkuat analisis dan mendapatkan hasil pengelompokan yang lebih komprehensif, digunakan pula tiga variabel pendukung, yaitu tingkat kegemaran membaca (TKM), tingkat penyelesaian pendidikan SMA/ sederajat (TPP), dan rata-rata lama sekolah (RLS).

Keempat variabel ini dipilih karena secara konseptual merepresentasikan aspek pendidikan, ekonomi, dan kesejahteraan, yang menjadi komponen utama dalam pengukuran pembangunan manusia. Adapun atribut data yang digunakan dalam penelitian ini dapat dilihat pada Tabel 1:

Tabel 1. Atribut Dataset

Variabel	Keterangan	Kode
IPM	Indeks Pembangunan Manusia	X ₁
TKM	Tingkat Kegemaran Membaca	X ₂
TPP	Tingkat Penyelesaian Pendidikan SMA/ Sederajat	X ₃
RLS	Rata-rata Lama Sekolah	X ₄

2. *Pre-processing/Cleaning*

Tahap *preprocessing* dilaksanakan guna menjamin keandalan data sebelum dianalisis. Seluruh variabel diperiksa menggunakan fungsi (*summary*) pada RStudio untuk mendeteksi adanya *missing value* atau data yang tidak valid. Hasil pemeriksaan menunjukkan bahwa seluruh data provinsi telah lengkap dan tidak terdapat nilai yang hilang.

Selanjutnya, dilakukan normalisasi data menggunakan fungsi *scale()* pada RStudio agar setiap variabel memiliki skala yang setara. Hal ini penting mengingat setiap indikator memiliki satuan dan rentang nilai yang berbeda, seperti IPM (indeks), RLS (tahun), dan TKM serta TPP (persentase). Proses normalisasi bertujuan untuk menghindari dominasi satu variabel terhadap hasil klasterisasi.

3. *Transformation*

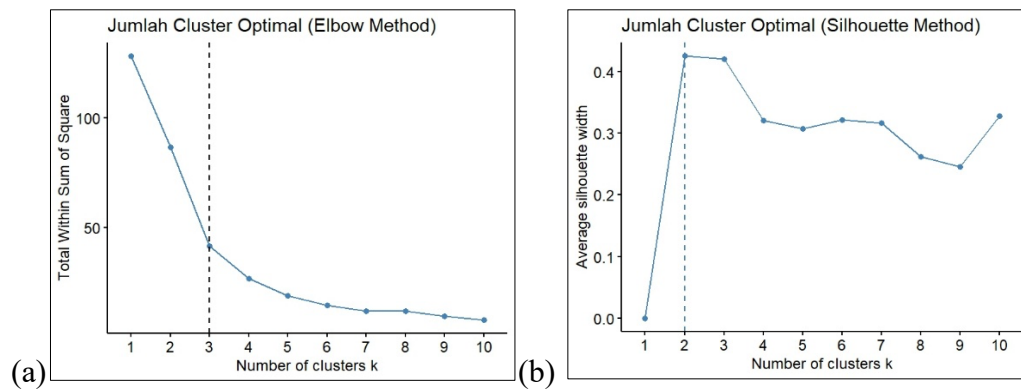
Pada tahap *transformation*, data hasil normalisasi diubah ke dalam bentuk matriks numerik agar dapat diolah oleh algoritma *k-means clustering*. Setiap observasi merepresentasikan satu provinsi, sedangkan setiap kolom merepresentasikan satu variabel (IPM, TKM, TPP, RLS).

4. *Data Mining*

Pada tahap ini, algoritma *k-means clustering* digunakan untuk mengelompokkan wilayah-wilayah administratif di Indonesia menurut kesamaan nilai indeks pembangunan manusia (IPM), tingkat kegemaran membaca (TKM), tingkat penyelesaian pendidikan SMA/ sederajat (TPP), dan rata-rata lama sekolah (RLS).

a. Penentuan Jumlah Kluster Optimal

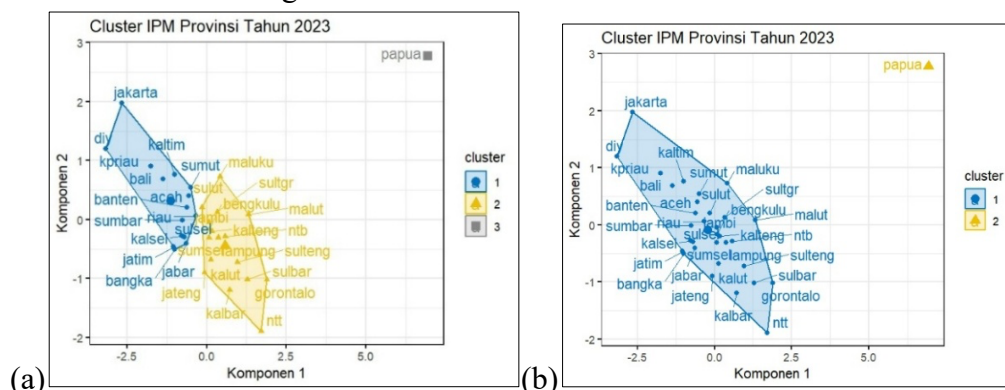
Sebelum melakukan proses pengelompokan (*clustering*), langkah awal yang perlu tahap awal adalah menetapkan banyaknya kelompok yang paling optimal. Dalam penelitian ini, estimasi jumlah kluster dikerjakan dengan dua pendekatan, yaitu metode *elbow* dan metode *silhouette* sebagai berikut.



Gambar 1. Penentuan Jumlah Cluster Optimal dengan Metode (a) *Elbow* dan (b) *Silhouette*

Berdasarkan hasil analisis penentuan jumlah kluster optimal pada Gambar 1(a), hasil metode *Elbow* menunjukkan *elbow point* (titik tekuk) berada pada $k = 3$, di mana penurunan nilai *total within sum of squares (WSS)* mulai melandai setelah tiga kluster yang menghasilkan pengelompokan data yang lebih rinci dan segmentasi yang lebih detail. Hal ini mengindikasikan bahwa penggunaan tiga kluster merupakan jumlah optimal karena setelah titik tersebut penambahan kluster tidak lagi memberikan pengurangan variasi yang signifikan di dalam kluster. Sementara itu, pada Gambar 1(b) menunjukkan hasil metode *Silhouette* dengan nilai rata-rata *silhouette width* tertinggi pada $k = 2$, yang berarti secara statistik dua kluster memberikan pemisahan paling baik antar kelompok dan memberikan pengelompokan yang lebih global dan minim segmentasi.

b. Proses Klasterisasi dengan *K-Means*



Gambar 2. Visual Hasil Klasterisasi dengan (a) $k = 3$ dan (b) $k = 2$

Proses *data mining* dilakukan dengan menerapkan algoritma *k-means* berdasarkan hasil penentuan jumlah kluster optimal. Klasterisasi pertama dilakukan dengan $k = 3$, yang mengelompokkan provinsi ke dalam kategori tinggi, sedang, dan rendah. Visualisasi hasil klasterisasi pada Gambar 2(a) menunjukkan bahwa provinsi dengan nilai IPM tinggi seperti DKI Jakarta, DI Yogyakarta, Bali, Kalimantan Timur, dan Kepulauan Riau cenderung berkelompok dalam satu wilayah yang berdekatan, menandakan kemiripan karakteristik pembangunan manusia yang relatif lebih baik. Sebaliknya, provinsi-provinsi di kawasan timur seperti Nusa Tenggara Timur, Maluku, dan Gorontalo tergolong dalam kluster Sedang, sedangkan Papua terpisah jauh dari kluster lainnya karena memiliki capaian IPM paling rendah.

Sebagai pembandingan, klasterisasi juga dilakukan dengan jumlah kluster $k = 2$, sesuai hasil optimal dari metode *Silhouette*. Hasil klasterisasi dengan $k = 2$ pada

gambar 2(b) menunjukkan bahwa provinsi-provinsi di Indonesia terbagi ke dalam dua kategori, yaitu tinggi dan rendah. Visualisasi hasil klasterisasi pada Gambar 2 memperlihatkan bahwa hampir seluruh provinsi, kecuali Papua, bergabung dalam satu kluster besar dengan karakteristik pembangunan manusia yang relatif serupa. Kondisi ini menunjukkan adanya kesenjangan yang cukup signifikan antara Papua dan provinsi-provinsi lainnya dalam hal capaian IPM, TKM, TPP, dan RLS.

5. Interpretation/Evaluation

Cluster	Kategori	Rata_rata_IPM	Jumlah_Provinsi
1	Tinggi	76.55733	15
2	Sedang	72.31471	17
3	Rendah	63.01000	1

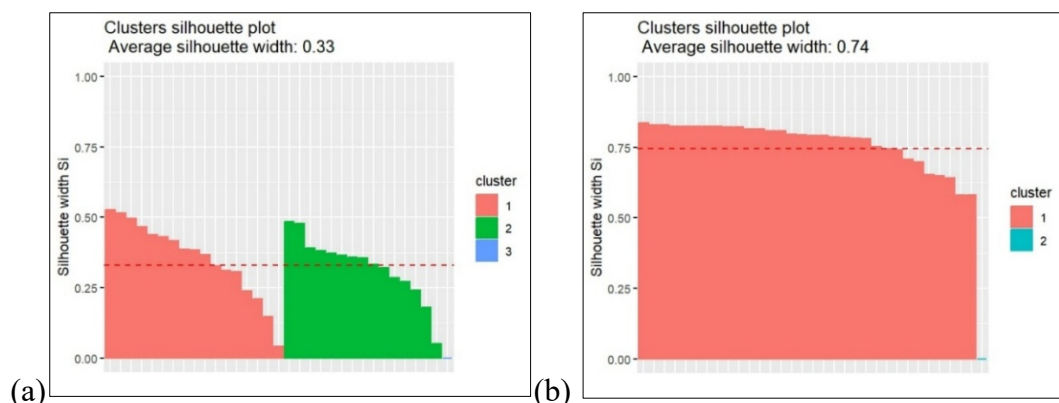
Gambar 3. Ringkasan Hasil Klasterisasi *K-Means* dengan $k = 3$

Berdasarkan hasil klasterisasi dengan $k = 3$, provinsi-provinsi di Indonesia terbagi menjadi tiga kategori, yaitu tinggi, sedang, dan rendah. Hasil ini menunjukkan bahwa terdapat perbedaan yang cukup mencolok antara provinsi-provinsi di Indonesia. Provinsi dengan *cluster* 1 (tinggi) dengan rata-rata IPM sebesar 76,56. Sementara itu, *cluster* 2 (sedang) terdiri atas 17 provinsi yang memiliki rata-rata IPM sebesar 72,31, mencerminkan provinsi dengan capaian pembangunan manusia yang cukup baik namun masih menghadapi tantangan. Adapun *cluster* 3 (rendah) dengan rata-rata IPM sebesar 63,01, yang menunjukkan kesenjangan pembangunan manusia yang signifikan dibandingkan cluster lainnya.

Cluster	Kategori	Rata_rata_IPM	Jumlah_Provinsi
1	Tinggi	74.30344	32
2	Rendah	63.01000	1

Gambar 4. Ringkasan Hasil Klasterisasi *K-Means* dengan $k = 2$

Untuk memperkuat validasi hasil, analisis juga dilakukan dengan $k = 2$ sesuai hasil optimal dari *Silhouette Method*. Hasil klasterisasi ini menunjukkan bahwa sebagian besar provinsi tergabung dalam *cluster* 1 (tinggi) dengan rata-rata IPM sebesar 74,30, sedangkan *cluster* 2 (rendah) dengan rata-rata IPM sebesar 63,01. Kondisi ini menegaskan adanya kesenjangan yang tajam antara Papua dan provinsi-provinsi lain di Indonesia dalam hal capaian IPM, tingkat kegemaran membaca (TKM), tingkat penyelesaian pendidikan (TPP), dan rata-rata lama sekolah (RLS).



Gambar 5. Hasil Evaluasi dengan *Silhouette* dengan (a) $k = 3$ dan (b) $k = 2$

Hasil evaluasi nilai *Silhouette* menunjukkan bahwa *average silhouette width* untuk $k = 3$ sebesar 0,33, sedangkan untuk $k = 2$ meningkat secara signifikan menjadi 0,74. Nilai ini mengindikasikan bahwa model dengan dua klaster memiliki pemisahan antar kelompok yang lebih jelas dan struktur data yang lebih kompak, sehingga secara statistik hasil pengelompokan tersebut lebih optimal. Namun, pembagian tiga klaster lebih disarankan untuk kepentingan kebijakan publik karena pembagian ini mampu memisahkan wilayah-wilayah berdasarkan tingkatannya, sehingga intervensi pemerintah dalam meningkatkan minat baca dan rata-rata lama sekolah dapat dilakukan dengan strategi yang lebih spesifik dan tepat sasaran bagi masing-masing tingkatan.

Hasil ini memperlihatkan jarak yang cukup jauh antara klaster rendah dengan klaster lainnya. Hasil penelitian ini yang menempatkan wilayah timur Indonesia pada klaster rendah sejalan dengan temuan (Rahmati, 2021) yang menyatakan adanya disparitas pembangunan antara wilayah barat dan timur Indonesia. Kondisi ini menegaskan bahwa intervensi kebijakan tidak bisa disamaratakan untuk seluruh wilayah di Indonesia.

SIMPULAN

Hasil analisis memperlihatkan bahwa metode komputasi *k-means clustering* berhasil mengelompokkan provinsi di Indonesia ke dalam tiga kategori pembangunan manusia berdasarkan IPM dan literasi pendidikan. Klaster tinggi mencakup 15 provinsi dengan rata-rata IPM 76,56, klaster sedang terdiri dari 17 provinsi dengan rata-rata IPM dan literasi pendidikan 72,31, dan klaster rendah hanya terdiri dari Papua dengan IPM dan literasi pendidikan 63,01. Nilai *average silhouette* sebesar 0,33 untuk $k = 3$ dan 0,74 untuk $k = 2$ menunjukkan bahwa dua klaster lebih optimal secara statistik. Namun, pembagian tiga klaster lebih disarankan karena memberikan gambaran yang lebih jelas mengenai variasi pembangunan manusia antarprovinsi, yang masih menunjukkan kesenjangan cukup besar dan intervensi kebijakan tidak bisa disamaratakan untuk seluruh wilayah di Indonesia. Secara praktis, hasil klasterisasi berpotensi menjadi rujukan bagi pemerintah nasional maupun lokal dalam menyusun strategi pembangunan manusia yang lebih terarah, berbasis data, dan sesuai dengan karakteristik setiap kelompok wilayah. Penggunaan metode *data mining* dengan algoritma *k-means* juga terbukti efektif dalam memberikan pemetaan yang komprehensif terhadap kondisi pembangunan manusia di Indonesia, serta mendukung perencanaan pembangunan yang lebih inklusif dan berkeadilan antarprovinsi.

DAFTAR PUSTAKA

- Badan Pusat Statistik Indonesia. (18 Februari 2024). *Indeks Pembangunan Manusia Menurut Provinsi, 2023*. <https://www.bps.go.id/id/statistics-table/3/V25GaFNHaExaMnhITm1sWmRrUIJZelJzYUc1SGR6MDkjMyMwMDAw/indeks-pembangunan-manusia-menurut-provinsi.html?year=2023>
- Badan Pusat Statistik Indonesia. (22 April 2024). *Tingkat Penyelesaian Pendidikan Menurut Jenjang Pendidikan dan Provinsi, 2023*. Diakses pada 24 September 2025, dari <https://www.bps.go.id/id/statistics-table/2/MTk4MCMY/tingkat-penyelesaian-pendidikan-menurut-jenjang-pendidikan-dan-provinsi.html>
- Badan Pusat Statistik Indonesia. (12 Desember 2025). *Rata-Rata Lama Sekolah Penduduk Umur 15 Tahun ke Atas Menurut Provinsi, 2023*. Diakses pada 10 Januari 2026, dari <https://www.bps.go.id/id/statistics-table/2/MTQyOSMy/rata-rata-lama-sekolah-penduduk-umur-15-tahun-ke-atas-menurut-provinsi.html>
- Badan Pusat Statistik Indonesia. (2024). *Tingkat Kegemaran Membaca Masyarakat dan Unsur Penyusunnya Menurut Provinsi, 2024*. Diakses pada 24 September 2025, dari <https://www.bps.go.id/id/statistics->

<table/3/TIROMldrTjVjVzEwVWtkbmRUSk5abkk0T0U5Q1FUMDkjMw==/tingkat-kegemaran-membaca-masyarakat-dan-unsur-penyusunnya-menurut-provinsi--2024.html?year=2024>

- Kassambara, A. (2017). *Multivariate analysis I: Practical guide to cluster analysis in R – Unsupervised machine learning*. Taylor & Francis Group.
- Luthfi, E., & Wijayanto, A. W. (2021). Analisis perbandingan metode hierarchical, K-Means, dan K-Medoids clustering dalam pengelompokan indeks pembangunan manusia Indonesia. *INOVASI*, 17(4), 761–773. <https://doi.org/10.30872/jinv.v17i4.10106>
- Nasution, M. (2019). Faktor determinan indeks pembangunan manusia di Indonesia. *Jurnal Budget: Isu dan Masalah Keuangan Negara*, 4(1), 121–143.
- Nurlina, N., Ridha, A., & Asnidar, A. (2023). Analisis determinan indeks pembangunan manusia di Indonesia tahun 1990–2021. *Jurnal Samudra Ekonomi dan Bisnis*, 14(2), 239–250. <https://doi.org/10.33059/jseb.v14i2.7287>
- Puspita, R. N. (2021). Analisis K-Means cluster pada kabupaten/kota di Provinsi Banten berdasarkan indikator indeks pembangunan manusia. *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, 2(3), 267–281.
- Rahmati, R. (2021). Analisis cluster dengan algoritma K-Means, Fuzzy C-Means dan Hierarchical clustering (Studi Kasus: Indeks Pembangunan Manusia Tahun 2019). *JIKO (Jurnal Informatika dan Komputer)*, 5(2), 73–80.
- Sakhi, D. A., Fatmala, F. N., Auralia, K., Pratama, A. R., & Damaliana, A. T. (2025). Klasterisasi kabupaten/kota di Provinsi Jawa Tengah berdasarkan komponen indeks pembangunan manusia menggunakan metode UMAP dan K-Means. *MATHunesa: Jurnal Ilmiah Matematika*, 13(2), 607–618.
- Sari, R. Y., Oktavianto, H., & Sulisty, H. W. (2022). Algoritma K-Means dengan metode Elbow untuk mengelompokkan kabupaten/kota di Jawa Tengah berdasarkan komponen pembentuk indeks pembangunan manusia. *Jurnal Smart Teknologi*, 3(2), 104–108.
- Sikana, A. M., & Wijayanto, A. W. (2021). Analisis perbandingan pengelompokan indeks pembangunan manusia Indonesia tahun 2019 dengan metode partitioning dan hierarchical clustering. *Jurnal Ilmu Komputer*, 14(2), 66–78.
- United Nations Development Programme. (1990). *Human development report 1990: Concept and measurement of human development*. Oxford University Press.
- Yudhanegara, M. R., & Lestari, K. E. (2024). Pelatihan penggunaan algoritma K-Means clustering untuk mengidentifikasi karakteristik siswa. *JALIYE: Jurnal Abdimas, Loyalitas, dan Edukasi*, 3(1), 19–25. <https://doi.org/10.47662/jaliye.v3i1.710>