JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.227-239

ANALISIS K-MEANS CLUSTERING PADA KABUPATEN/KOTA DI INDONESIA BERDASARKAN INDEKS PEMBANGUNAN MANUSIA (IPM)

Muhammad Andreanto[*1)], Berlian Chandra Amanullah[2)] , Nur Wakhidah [3)]
1. Sistem Informasi, Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang, Indonesia
2. Sistem Informasi, Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang, Indonesia
3. Sistem Informasi, Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang, Indonesia

Article Info
Kata Kunci:
Clustering; K-Means; Indeks Pembangunan Manusia.

Keywords:
Clustering; Human Development Index; K-Means K-Means.

Article history:
Received 21 January 2025
Revised 7 March 2025
Accepted 15 March 2025
Available online 1 March 2026

DOI :
https://doi.org/10.29100/jipi.v11i1.7455

* Corresponding author.
Corresponding Author
E-mail address:
muhammadandreanto29@gmail.com

ABSTRAK
Indeks Pembangunan Manusia (IPM) ialah indikator komposit yang dipergunakan dalam mengukur kualitas hidup. IPM di Indonesia mengalami peningkatan setiap tahunnya. Walaupun demikian, terdapat kesenjangan signifikan antara Kabupaten/Kota di Indonesia. Oleh karena itu, diperlukan analisis pengelompokan wilayah berdasarkan IPM untuk mendukung perencanaan dan evaluasi kebijakan pemerintah secara tepat sasaran. Metode yang digunakan adalah analisis cluster dengan teknik K-Means. Hasil analisis menunjukkan terbentuknya empat cluster. *Cluster* 0 memiliki nilai IPM tinggi rata-rata 72,47, meliputi Kab/Kota dari Provinsi Jawa Tengah, Jawa Timur, Sulawesi Selatan, Sulawesi Utara, dan Jawa Barat. *Cluster* 1 menunjukkan nilai IPM sedang rata-rata 68,82, mencakup wilayah Kab/Kota dari Provinsi Nusa Tenggara Barat, Sulawesi Tenggara, Papua, Aceh, dan Sumatera Utara. *Cluster* 2 memiliki nilai IPM sangat tinggi, rata-rata 81,03 terdiri dari wilayah Kab/Kota dari Provinsi Jawa Timur, Jawa Tengah, Jawa Barat, dan D.I.Yogyakarta. *Cluster* 3, dengan nilai IPM rendah rata-rata 52,65, mencakup Provinsi Papua dan Papua Barat.

ABSTRACT
The Human Development Index (HDI) is a composite indicator used to measure the quality of life. The HDI in Indonesia has been increasing every year. However, there is a significant disparity between regencies/cities in Indonesia. Therefore, an analysis of regional clustering based on HDI is needed to support targeted government policy planning and evaluation. The method used is cluster analysis with the K-Means technique. The analysis results show the formation of four clusters. Cluster 0 has a high HDI value with an average of 72.47, including regencies/cities from the provinces of Central Java, East Java, South Sulawesi, North Sulawesi, and West Java. Cluster 1 shows a medium HDI value with an average of 68.82, covering regencies/cities from the provinces of West Nusa Tenggara, Southeast Sulawesi, Papua, Aceh, and North Sumatra. Cluster 2 has a very high HDI value, with an average of 81.03, consisting of regencies/cities from the provinces of East Java, Central Java, West Java, and D.I. Yogyakarta. Cluster 3, with a low HDI value averaging 52.65, includes the provinces of Papua and West Papua.

.

I. PENDAHULUAN

KEBERHASILAN pemerintah dalam melaksanakan pembangunan nasional bukan hanya diukur berdasarkan perkembangan ekonomi, tetapi serta upaya meningkatkan tingkat SDM. Salah satu cara dalam menilai keberhasilan ini yaitu melalui Indeks Pembangunan Manusia (IPM) sebagai indikator yang menunjukkan kualitas hidup masyarakat. IPM merupakan ukuran komposit yang dipakai untuk menilai pencapaian dalam peningkatan kualitas hidup[1]. IPM menggambarkan sejauh mana masyarakat dapat menikmati hasil pembangunan melalui akses terhadap layanan kesehatan, pengetahuan, mata pencaharian yang lebih aman, dan keamanan dari tindak kejahatan fisik [2].

Berdasarkan data BPS (2023), IPM Indonesia pada tahun 2023 memperoleh angka 74,39, meningkat sebanyak 0,84 persen dibandingkan tahun lalu.sebesar 73,77[3]. Dalam periode 2020–2023, rata-rata peningkatan IPM Indonesia adalah 0,72 persen per tahun. Akan tetapi, dibalik meningkatnya nilai IPM dari tahun sebelumnya terdapat sebuah kesenjangan antar Kabupaten/Kota di Indonesia. Maka dari itu, diperlukan analisis untuk meng-clusterkan Kabupaten/Kota di Indonesia guna mendukung program serta evaluasi dari pemerintah yang agar lebih efektif dan sesuai tujuan. Salah satu metode yang selaras dengan tujuan ini yaitu analisis cluster menggunakan pendekatan *K-Means.*

K-Means adalah metode data mining dengan pendekatan jarak terpendek untuk mengelompokkan data ke dalam beberapa cluster[4]. Pengelompokan ini dilakukan berdasarkan kriteria numerik yang telah ditentukan [5]. Dalam penelitian ini, metode. K-Means digunakan dalam. meng-clusterkan Kab/Kota. di Indonesia berdasarkan nilai IPM dari setiap Kab/Kota di Indonesia berdasarkan nilai IPM dari setiap Kab/Kota di Indonesia.

Salah satu penelitian terdahulu yang meneliti tentang Indeks Pembangunan Manusia (IPM) pernah diteliti oleh Puspita et al.(2021) yang menganalisis Kab/Kota di Provinsi Banten dengan metode. K-Means berdasarkan indikator IPM menghasilkan 3[6].

Penelitian Anggraini & Rismawati Arum (2022) mengenai Analisis Cluster dengan Algoritma K-Means Pada Provinsi Sumatera Barat Berdasarkan IPM Tahun 2021 mengelompokkan daerah ke dalam tiga cluster. .Cluster 1 mencakup Kab/Kota dengan IPM rendah, Cluster 2 mencakup Kab/Kota dengan nilai IPM paling tinggi, sedangkan Cluster 3 merupakan kelompok Kab/Kota dengan nilai IPM ditengah-tengah[6].

Penelitian terdahulu yang menggunakan *K-Means Clustering* pernah dilakukan oleh Ratna Sari et al. (2020) mengenai pengelompokkan data kemiskinnan di Provinsi Banten dengan metode *K-Means* menggunakan Rapidminer. Dari perhitungan rapidminer menghasilkan tiga *cluster*. Tingkat. kemiskinan sedang, tertinggi, serta rendah[7].

Penelitian terdahulu yang menggunakan *K-Means Clustering* pernah dilakukan oleh Ratna Sari et al. (2020) mengenai pengelompokkan data kemiskinnan di Provinsi Banten dengan metode K-Means menggunakan Rapidminer. Dari perhitungan rapidminer menghasilkan tiga cluster. Tingkat. kemiskinan sedang, tertinggi, serta rendah[8]

Penelitian Maliki et al. (2022) mengenai metode K-Means untuk membandingkan tingkat. pengangguran terbuka Provinsi. yang ada di Indonesia. Terbagi menjadi dua cluster. Berdasarkan analisis data TPT 2016-2019[9].

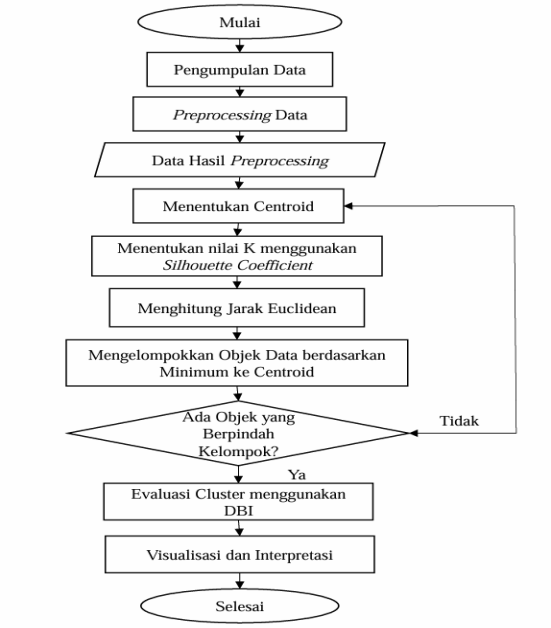
Berdasarkan uraian diatas, metode *K-Means Clustering* digunakan untuk mengelompokan Kab/Kota berdasarkan IPM yang ada di Indonesia pada tahun 2023. Penelitian ini membahas penggunaan metode K-Means untuk meng-clusterkan Kab/Kota di Indonesia menurut IPM melalui empat indikator pembentuk IPM, ialah Umur Harapan Hidup saat Lahir (UHH), Rata-Rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), dan Pengeluaran Riil per Kapita per Tahun.

Penelitian sebelumnya tentang IPM di Indonesia umumnya terbatas pada analisis di tingkat provinsi atau wilayah tertentu, seperti penelitian oleh Puspita et al. (2021) yang menganalisis Kabupaten/Kota di Provinsi Banten, atau penelitian yang dilakukan Anggraini & Rismawati Arum (2022) yang hanya fokus pada Sumatera Barat. Penelitian ini memiliki cakupan lebih luas, yaitu seluruh Kab/Kota di Indonesia, serta penggunaan data terbaru tahun 2023. Selain itu, penelitian ini menggunakan empat indicator utama, yaitu UHH, HLS, RLS, dan PRP, yang memberikan analisis yang lebih komprehensif dibandingkan penelitian sebelumnya yang hanya menggunakan beberapa indikator.

II. METODE PENELITIAN

Metode *K-Means Clustering* digunakan dengan data yang berasal dari publikasi BPS tahun 2023, unit observasi yang digunakan adalah Kab/Kota yang ada di Indonesia. Menggunakan 4 variabel pembentuk IPM secara berurutan ialah X1, X2, X3, X4. Berikut merupakan diagram alir (Flowchart) yang mewakili tahapan-tahapan analisis dalam penelitian ini.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.227-239



Gambar. 1. Diagram Alur Penelitian

Berikut merupakan langkah-langkah pada Algoritma *K-Means*, yaitu[10]: 1)Pengumpulan data kemudian dilanjutkan dengan 2)*Preprocessing* data yang terdiri dari normalisasi data dan mengubah *data frame* mejadi *array,* 3)Menetpkan nilai k dengan Metode Ellbow. Sebelum menentapkan nilai k menjadi jumlah cluster, perlu dilakukan uji kualitas *cluster* dengan *Silhouette Coefficient* menggunakan persamaan (1) [11]

$$S(i) = \frac{b(i) - \alpha(i)}{\max\{\alpha(i), b(i)\}} \quad (1)$$

S(i) adalah nilai *Silhouette Cooffcient*, b(i) merupakan nilai terkecil dari jarak objek i dengan objek lainnya dalam kelompok berbeda C, a(i) yaitu rata-rata jarak objek i dan seluruh objek kelompok yang sama. Setelah melakukan pengukuran kualitas cluster, langkah selanjutnya yaitu 4) Menetapkan nilai k untuk total cluster yang akan terbentuk berdasarkan hasil evaluasi *Silhouette Coofficient* [12], [13], [14], 5)Menetapkan k centroid pada titik pusat cluster pertama secara acak, 6)Menghitung jarak menggunakan rumus Euclidean menggunakan persamaan (2)[15].

$$d_{ik} = \sqrt{\sum_{j}^{m} (CIJ - CKJ)^2} \quad (2)$$

Selanjutnya yaitu mengelompokkan data ke dalam kelompok berdasarkan jarak. terdekat .menggunakan persamaan (3)

$$Min \sum_{k=1}^{k} d_{ik} = \sqrt{\sum_{j}^{m} (Cij - Ckj)^2} \quad (3)$$

Selanjutnya menghitung pusat cluster yang sudah terbentuk menggunakan persamaan (4)

$$C_{kj} = \frac{\sum_{i=1}^{p} x_{ij}}{p} \quad (4)$$

Langkah selanjutnya yaitu mengelompokkan objek dengan menentukan anggota cluster berdasarkan perhitungan jarak minimum objek. Hasil perhitungan ini akan bernilai 0 atau 1, di mana nilai 1 menunjukkan bahwa data tersebut ditempatkan ke *cluster* tertentu, sementara nilai 0 menunjukkan bahwa data tersebut ditempatkan ke cluster yang lain. Proses ini kemudian diulang dari tahap sebelumnya, yaitu melakukan iterasi sampai nilai *centroid* stabil dan anggota kelompok tidak lagi berpindah.

Setelah cluster terbentuk, Langkah selanjutnya mengevalusasi cluster menggunakan Indeks Davies Bouldin (DBI). DBI adalah salah satu alat untuk mengevaluasi kualitas pengelompokan (*clustering*) dalam analisis data. Fungsi utamanya adalah untuk menilai sejauh mana hasil pengelompokan dari suatu algoritma clustering mampu memisahkan kelompok data yang berbeda sekaligus mendekatkan data ke pusat cluster masing-masing. Nilai DBI yang lebih rendah menunjukkan kualitas cluster yang lebih baik[16], [17]. Rumus yang digunakan untuk menghitung nilai DBI (Davies-Bouldin Index). Rumus-rumus tersebut adalah sebagai berikut [18], [19].

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (5)$$

Berdasarkan rumus (5), dapat dijelaskan bahwa m_i mewakili banyaknya data di dalam *cluster* ke – i, kemudian c_i adalah *centroid cluster* ke – i, dan $d(x_j, c_i)$ merupakan jarak antar data ke – i ke centroid *cluster* i, rumus ini berfungsi untuk menghitung matrik kohesi dalam suatu *cluster.* Selanjutnya untuk mengetahui nilai separasi antar Cluster menggunkan rumus (6).

$$SSB_{i,j} = d(c, c_i) \quad (6)$$

Berikutnya melakukan pengukuran rasio dan menentukan nilai perbandingan antara *cluster* k -i dan *cluster* ke-i. Proses ini mrnggunakan rumus (7).

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (7)$$

Selanjutnya melakukan perhitungan untuk menentukan nilai DBI menggunakan persamaan (8).

$$DBI = \frac{1}{K} \sum_{i=1}^{K} max_{j \neq I}(R_{I,J}) \quad (8)$$

Setelah seluruh tahapan selesa, langkah selanjutnya adalah visualisasi dan interpretasi terhadap hasil pembentukan *cluster* oleh metode *K-Means[20]*.

III. HASIL DAN PEMBAHASAN

A. Data Penelitian

Berikut data yang dipakai dalam. mengelompokkan Kab/Kota. di Indonesia berdasarkan IPM Tahun 2023, total data yang digunakan sebanyak 514 data dapat disajikan dalam Tabel I.

TABEL I
DATA PENELITIAN

Kab/Kota	UHH	HLS	RLS	PRP	IPM
Simelue	69,57	14,28	9,81	7686	69,98
Aceh Singkil	69,85	14,35	8,70	9374	71,14
Aceh Selatan	70,33	14,70	8,91	8712	71,14
⋮	⋮	⋮	⋮	⋮	⋮
Intan Jaya	66,49	7,69	3,27	6140	50,39
Deiyai	44,52	9,85	3,27	5104	51,40
Kota Jayapura	71,76	15,26	11,84	15272	81,51

B. *Pra-Procesing Data*

Sebelum data penelitian ini digunakan untuk melakukan proses *Cluster*, maka perlu dilakukan tahap *pra-processing*. Pada tahap ini dilakukan pemilihan Variabel untuk dijadikan acuan dalam proses pengelompokkan data yaitu HLS, UHH, PRP, RLS, dan IPM. Hasil *pra-processing* terdapat dalam Tabel II.

TABEL II
DATA HASIL PRAPROCESING

UHH	HLS	RLS	PRP	IPM	UHH
69,57	14,28	9,81	7686	69,98	69,57
69,85	14,35	8,70	9374	71,14	69,85
70,33	14,70	8,91	8712	71,14	70,33
⋮	⋮	⋮	⋮	⋮	⋮
66,49	7,69	3,27	6140	50,39	66,49
44,52	9,85	3,27	5104	51,40	44,52
71,76	15,26	11,84	15272	81,51	71,76

1) *Normalisasi Data*

Setelah dilakukan pemilihan variabel langkah selanjutnya yaitu menormalisasi data yang digunakan. Dalam tahapan ini bertujuan untuk mempermudah proses analisis dan interpretasi data. Jika data tidak dinormalisasi, fitur dengan skala yang lebih besar dapat mendominasi jarak yang dihitung dan mempengaruhi hasil model.

2) *Mengubah Data Frame Menjadi Array*

Setelah melakukan normalisasi data, variabel awal berupa data frame diubah ke dalam bentuk array. Dalam tahapan ini perlu dilakukan karena dalam algoritma *K-Means Clustering* mengandalkan perhitungan jarak antar titik data dalam ruang fitur yang multidimensi, sehingga lebih mudah jika data dalam bentuk array atau vektor.

KODE PROGRAM I
MENGUBAH DATA FRAME MENJADI ARRAY

```
# Menghapus kolom non-numerik
data_numeric = data.drop(columns=['Kab/Kota'])

# Normalisasi data
scaler = MinMaxScaler()
data_normalized = scaler.fit_transform(data_numeric) # Menghasilkan array NumPy
print(data_normalized)
```

Hasil output kode program I dapat dilihat pada tabel III

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.227-239

TABEL III
DATA ARRAY

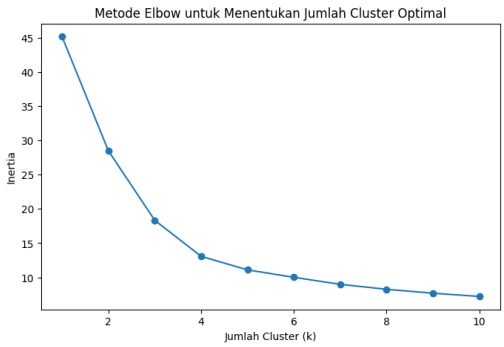
UHH	HLS	RLS	PRP	IPM	UHH
0.40498221	0.7316175	0.71491615	0.16166416	0.63420381	0.40498221
0.42491103	0.73676471	0.61694616	0.24351452	0.65698017	0.42491103
0.45907473	07625	0.63548102	0.21141444	0.65698017	0.45907473
⋮	⋮	⋮	⋮	⋮	⋮
0.19928826	0.24705882	0.13768756	0.08669932	0.24955822	0.19928826
0.18790036	0.40588235	0.13768756	0.03646414	0.26938936	0.18790036
0.56085409	0.80367647	0.13768756	0.52950589	0.86059297	0.56085409

C. *Menentukan Nilai K Menggunakan Metode Ellbow*

Setelah dilakukan normalisasi data dan data array, langkah selanjutnya yaitu menentukan nilai K memakai Metode Ellbow. Metode Ellbow adalah teknik yang dipakai untuk membantu menentukan jumlah optimal cluster dalam algoritma K-Means. Tujuannya adalah untuk memilih jumlah *cluster* yang memberikan pembagian data yang paling baik, tanpa menghasilkan model yang terlalu kompleks atau terlalu sederhana. Apabila nilai kelompok pertama. dibandingkan kelompok kedua. membentuk sudut. pada grafik, maka jumlah *Cluster* yang tepat dapat ditentukan. Berdasarkan perhitungan menggunakan Metode Elbow, nilai K Optimum adalah 4. Oleh karena itu, jumlah *cluster* yang digunakan sebanyak 4 *cluster*.

KODE PROGRAM II
METODE ELLBOW

```
# Menentukan inertia untuk berbagai jumlah cluster#
inertias = []
k_range = range(1, 11)  # Coba jumlah cluster dari 1 hingga 10
for k in k_range:
    kmeanModel  = KMeans(n_clusters=k)
    kmeanModel.fit(data_normalized)
    inertias.append(kmeanModel.inertia_)
# Plot metode elbow
plt.figure(figsize=(8, 5))
plt.plot(k_range, inertias, marker='o')
plt.xlabel('Jumlah Cluster (k)')
plt.ylabel('Inertia')
plt.title('Metode Elbow untuk Menentukan Jumlah Cluster Optimal')
plt.show()
```



Gambar. 2. Hasil Nilai K dengan Metode Ellbow

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.227-239

D. Menentukan Jumlah Cluster

Menentukan jumlah. *cluster*. umumnya menggunakan metode Ellbow. Penelitian. ini akan melakukan perbandingan hasil nilai *Silhouete* dari nilai k= 3 dan nilai k= 4. Apabila . *Silhouete Coefficient* dari i semakin mendekati 1, maka cluster dalam objek. i saling berdekatan dan. objek i terpaut. jauh dengan yang lain.

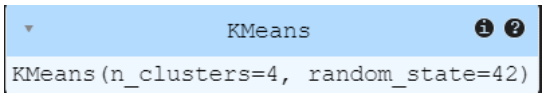
KODE PROGRAM III
HITUNG NILAI SILHOUTTE

```
# Hitung nilai Silhouette.
silhouette_avg = silhouette_score(data_normalized, data['Cluster']).
print(f"Silhouette Score: {silhouette_avg:.3f}").
```

Hasil output nilai *Silhouette Coefficient* dari kode diatas yaitu nilai k=3 yaitu 0,644 dan nilai *Silhouette* dari k=4 yaitu 0.719. Hal ini menunjukkan bahwa kualitas cluster dengan nilai k=4 sangat baik, artinya titik-titik dalam cluster tersebut dekat satu sama lain. Berdasarkan perhitungan menggunakan Metode Elbow dan perbandingan menggunakan nilai *Silhouette*, diperoleh bahwa nilai K Optimum yaitu 4. Maka dari itu, *cluster* yang digunakan untuk penelitian ini sejumlah 4 *cluster*.

KODE PROGRAM IV
MENENTUKAN JUMLAH CLUSTER

```
# Jumlah cluster optimal yang ditemukan adalah 4
optimal_clusters = 4
# Inisialisasi dan jalankan K-means dengan jumlah cluster optimal
kmeanModel = KMeans(n_clusters=optimal_clusters, random_state=42)
                                kmeanModel.fit(data_normalized)
```



Gambar. 3. Jumlah Cluster

E. Menentukan Centroid

Setelah data di-cluster secara random, langkah selanjutnya yaitu menentukan *centroid* dari masing-masing *cluster*. *Centroid* adalah titik yang digunakan untuk mewakili posisi pusat dari sebuah *cluster*, dan berfungsi sebagai penentu lokasi *cluster* tersebut.

KODE PROGRAM V
MENENTUKAN NILAI CENTROID

```
# Mencari nilai centroid untuk setiap cluster
centroids = kmeanModel.cluster_centers_
# Tampilkan centroid
centroids_df = pd.DataFrame(centroids, columns=data_numeric.columns)
print("Nilai Centroid untuk Setiap Cluster:")
```

Hasil Output dari nilai *Centroid* dapat dilihat dalam tabel IV

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.227-239

TABEL IV
NILAI CENTROID

Cluster	UHH	HLS	RLS	PRP	IPM
0	0.713966	0.628037	0.57812	0.324431	0.683144
1	0.487959	0.652766	0.594635	0.216453	0.616861
2	0.788872	0.754442	0.813585	0.510804	0.849600
3	0.191154	0.395203	0.26317	0.067119	0.320664

F. *Menampilkan Hasil Cluster*

Dengan memanfaatkan library sckit-learn, langkah selanjutnya adalah menambahkan kolom "*Cluster*" yang akan menampilkan hasil dari proses *clustering*.

KODE PROGRAM VI
MENAMPILKAN HASIL CLUSTER

```
# Melakukan klustering
kmeans = KMeans(n_clusters=4)  # Jumlah 4 cluster
data['Cluster'] = kmeans.fit_predict(data_normalized)
# Menghapus kolom yang tidak perlu
data.drop(columns=['k-means', 'cluster'], inplace=True, errors='ignore')
# Hasil clustering sudah ada di kolom 'Cluster', tidak ada perubahan yang diperlukan di sini
# Jika Anda ingin mengonfirmasi hasil akhir
print(data)
# Tampilkan hasil clustering
print(data)
```

Output untuk hasil *cluster* dapat dilihat pada tabel V-VIII

TABEL V
CLUSTER 0

Kab/Kota	UHH	HLS	RLS	PRP	IPM
Sampang	73,49	12,54	5,07	9363	66,19
Bangkalan	73,29	11,97	5,99	9438	66,82
Sekadau	73,93	11,92	7,13	7981	66,83
⋮	⋮	⋮	⋮	⋮	⋮
Kota Tegal	74,84	13,18	9,24	14013	77,06
Karangayar	77,72	13,71	9,02	12260	77,31
Kota Cirebon	75,15	10,37	10,37	12506	77,45

Cluster 0 merupakan *cluster* dengan nilai IPM tinggi dengan nilai rata-rata yaitu sebesar 72,47. *Cluster* 0 meliputi beberapa Kab/Kota yang terdapat di berbagai Provinsi di Indonesia. Kab/Kota yang paling banyak terdapat di *cluster* 0 yaitu Kab/Kota dari Provinsi Jawa Tengah, Jawa Timur, Sulawesi Selatan, Sulawesi Utara, dan Jawa Barat.

TABEL VI
CLUSTER 1

Kab/Kota	UHH	HLS	RLS	PRP	IPM
Maybrat	65,82	13,49	7,32	5669	61,29
Sabu Raijua	66,46	13,18	6,98	5899	61,37
Manokwari Selatan	68,09	12,56	6,96	5948	61,55
⋮	⋮	⋮	⋮	⋮	⋮
Kota Banjar	70,75	14,22	10,54	10229	74,45
Aceh Barat Daya	71,89	14,64	9,98	10085	74,62
Jayapura	70,23	14,58	10,32	10671	74,74

Cluster 1 merupakan cluster dengan nilai IPM sedang dengan nilai rata-rata yaitu sebesar 68,82. *Cluster* 1 meliputi beberapa Kab/Kota yang terdapat di berbagai Provinsi di Indonesia. Kab/Kota yang paling banyak terdapat di *cluster* 1 yaitu Kab/Kota dari Provinsi Nusa Tenggara Barat, Sulawesi Tenggara, Papua, Aceh, dan Sumatera Utara.

TABEL VII
CLUSTER 2

Kab/Kota	UHH	HLS	RLS	PRP	IPM
Kota Sibolga	74,02	13,42	10,44	12285	75,3
Toba Samosir	74,22	13,59	10,59	12676	76,38
Deli Serdang	73,65	13,39	10,28	12890	76,52
⋮	⋮	⋮	⋮	⋮	⋮
Kota Jakarta Selatan	76,02	13,66	11,75	24975	86,71
Kota Banda Aceh	75,02	17,93	13,04	17521	88,32
Kota Yogyakarta	75,52	17,62	12,11	19920	88,61

Cluster 2 merupakan *cluster* dengan nilai IPM sangat tinggi dengan nilai rata-rata yaitu sebesar 81,03. *Cluster* 2 meliputi beberapa Kab/Kota yang terdapat di berbagai Provinsi di Indonesia. Kab/Kota yang paling banyak terdapat di cluster 2 yaitu Kab/Kota dari .Provinsi. Jawa Timur, Sumatera Utara, Jawa Tengah, Jawa Barat, dan D.I.Yogyakarta.

TABEL VIII
CLUSTER 3

Kab/Kota	UHH	HLS	RLS	PRP	IPM
Nduga	63,88	4,33	1,71	4352	37,68
Puncak	66,77	5,59	2,19	5975	44,71
Pegunungan Bintang	65,93	6,78	3,37	5849	48,51
⋮	⋮	⋮	⋮	⋮	⋮
Pegunungan Arfak	67,71	11,79	5,51	5438	57,82
Mamberamo Raya	68,75	11,83	6,28	5105	58,49
Paniai	64,49	10,53	4,79	6977	58,83

Cluster 3 merupakan *cluster* dengan nilai IPM rendah dengan nilai rata-rata yaitu sebesar 52,65. Cluster 3 meliputi beberapa Kab/Kota yang terdapat di berbagai Provinsi di Indonesia. Kab/Kota yang paling banyak terdapat di cluster 3 yaitu Kab/Kota dari Provinsi Papua dan Papua Barat.

Berdasarkan hasil *clustering,* terlihat bahwa Kab/Kota di *Cluster* 2 umumnya terletak diwilayah dengan tingkat ekonomi yang lebih maju, seperti Jawa Timur, Jawa Tengan, dan D.I. Yogyakarta. Hal ini didukung oleh kemudahan akses terhadap pendidikan dan layanan kesehatan, serta pengeluaran riil per kapita yang lebih tinggi.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.227-239

Sebaliknya, Kab/Kota di *Cluster* 3 yang didominasi oleh wilayah Papua dan Papua Barat menunjukkan keterbatasan akses layanan dasar, seperti Pendidikan dan Kesehatan, dan tingkat kemiskinan yang lebih tinggi. Faktor geografis seperti lokasi yang terpencil juga memengaruhi rendahnya IPM di wilayah ini.

G. *Evaluasi Cluster menggunakan Davies-Bouldin Index (DBI)*

Analisis sensitivitas dilakukan untuk menguji ketahanan model klustering terhadap perubahan data, menggunakan *Davies-Bouldin Index (DBI)* sebagai metrik evaluasi. DBI mengukur kualitas *clustering* dengan membandingkan jarak intra-cluster dan inter-cluster, di mana nilai DBI yang lebih rendah menunjukkan klustering yang lebih baik. Analisis ini membantu mengidentifikasi variabel kritis, memastikan konsistensi model, dan memvalidasi keandalan *clustering.*

KODE PROGRAM VIII
EVALUASI CLUSTER MENGGUNAKAN DBI

```
    n_clusters = len(np.unique(labels))  # Jumlah cluster
    cluster_k = [X[labels == k] for k in range(n_clusters)]  # Data points per cluster
    centroids = [np.mean(k, axis=0) for k in cluster_k]  # Centroid per cluster
    # Hitung dispersi (intra-cluster distance) untuk setiap cluster
    S = [np.mean(pairwise_distances(k, [centroids[i]])) for i, k in enumerate(cluster_k)]
    # Hitung DBI
    dbi = 0
    for i in range(n_clusters):
        max_ratio = -np.inf
        for j in range(n_clusters):
            if i != j:
                # Hitung rasio (S[i] + S[j]) / d(centroids[i], centroids[j])
                ratio = (S[i] + S[j]) / np.linalg.norm(centroids[i] - centroids[j])
                if ratio > max_ratio:
                    max_ratio = ratio
        dbi += max_ratio
    dbi /= n_clusters
    return dbi
# Generate data contoh
X, _ = make_blobs(n_samples=300, centers=4, cluster_std=0.60, random_state=0)
# Range nilai k (jumlah cluster) yang akan diuji
k_values = range(2, 11)
dbi_scores = []
# Evaluasi DBI untuk setiap nilai k
for k in k_values:
    kmeans = KMeans(n_clusters=k, random_state=0)
    labels = kmeans.fit_predict(X)
    dbi = davies_bouldin_index(X, labels)
    dbi_scores.append(dbi)
    print(f"k = {k}, DBI = {dbi}")
# Tentukan k optimal (nilai k dengan DBI terendah)
optimal_k = k_values[np.argmin(dbi_scores)]
print(f"\nJumlah cluster optimal (k) adalah: {optimal_k}")
```

Hasil Output dari Analisa Sensitivitas dengan DBI dapat dilihat dalam table IX

TABEL IX
EVALUASI MENGGUNAKAN DBI

Cluster	Hasil DBI
K = 2	0.676
K = 3	0.626
K = 4	0.437
K = 5	0.722
K = 6	0.928
K = 7	0.835
K = 8	0.945
K= 9	0.918
K = 10	1.018

Berdasarkan hasil yang terlihat pada tabel IX, cluster terbaik adalah cluster 4 dengan nilai evaluasi DBI = 0.437 hal ini sejalan dengan prinsip dasar DBI , di mana semakin nilai DB mendekati nol dan bukan min (DB >= 0), semakin baik kualitas *cluster* yang optimal

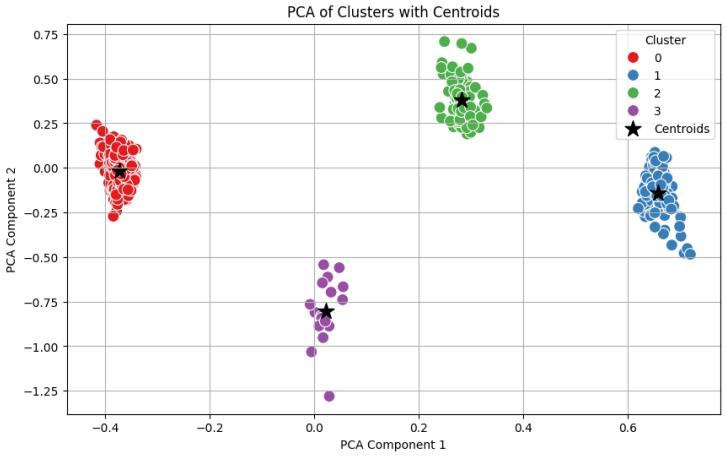
H. Visualisasi Hasil Cluster

Setelah selesai melakukan analisis cluster menggunakan *K-Means Clustering*, langkah selanjutnya yaitu menampilkan visualisasi hasil cluster. Hal tersebut dilakukan agar pembaca lebih mudah dalam memahami hasil analisis yang dilakukan.

KODE PROGRAM VII
VISUALISASI HASIL CLUSTER

```
# Melakukan PCA untuk mengurangi dimensi menjadi 2
pca = PCA(n_components=2).
df_pca = pca.fit_transform(df_normalized).
# Menambahkan hasil PCA ke DataFrame
df['PCA1'] = df_pca[:, 0]
df['PCA2'] = df_pca[:, 1]
# Menjalankan K-means clustering pada data terstandarisasi
kmeans = KMeans(n_clusters=4, random_state=0).
df['Cluster'] = kmeans.fit_predict(df_normalized).
# Menghitung posisi centroid pada komponen PCA
centroids = kmeans.cluster_centers_.
centroids_pca = pca.transform(centroids).
# Visualisasi hasil klustering dan centroid
plt.figure(figsize=(10, 6)).
sns.scatterplot(data=df, x='PCA1', y='PCA2', hue='Cluster', palette='Set1', s=100).
plt.scatter(centroids_pca[:, 0], centroids_pca[:, 1], marker='*', color='black', s=200,
label='Centroids') # Logo centroid
plt.title("PCA of Clusters with Centroids").
plt.xlabel('PCA Component 1').
plt.ylabel('PCA Component 2').
plt.legend(title='Cluster').
plt.grid().
plt.show().
```

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.227-239



Gambar. III. Visualisasi Hasil Cluter

Gambar. III. menunjukkan bahwa Kab/Kota di Indonesia berdasarkan IPM dikelompokkan menjadi 4 cluster berdasarkan masing-masing karakteristiknya. *Cluster* 0 menunjukkan nilai IPM tinggi, *cluster* 1 menunjukkan nilai IPM sedang dan cluster 2 menunjukkan bahwa Kab/Kota tersebut memiliki nilai IPM sangat tinggi, *cluster* 3 menunjukkan bahwa Kab/Kota tersebut memiliki nilai IPM rendah,

Hasil penelitian ini menghasilkan empat *cluster*, berbeda dengan beberapa penelitian sebelumnya seperti Puspita et al. (2021) yang hanya menghasilkan tiga cluster. Perbedaan ini mungkin disebabkan oleh cakupan data dalam lingkup lebih luas. Selain itu, penelitian ini juga menggunakan data tahun 2023, yang mencerminkan perkembangan terkini dalam pembangunan manusia di Indonesia.

Batasan dalam penelitian ini meliputi pemakaian data IPM tahun 2023, yang tidak memungkinkan analisis tren jangka panjang, serta keterbatasan variable karena hanya menggunakan empat indikator utama IPM. Di samping itu metode *K-Means Clustering* memiliki kelemahan ketika menangani *outlier* dan data yang tidak terdistribusi secara normal. Selain itu, Jumlah cluster (k) ditentukan secara subjektif menggunakan metode *Elbow*.

IV. KESIMPULAN

Berdasarkan hasil analisis yang dilakukan menggunakan *K-Means Clustering*, diperoleh 4 *cluster* dengan karakteristik sebagai berikut: *Cluster* 0 merupakan cluster dengan nilai IPM tinggi dengan rata-rata 72,47, mencakup Kab/Kota dari Provinsi Jawa Tengah, Jawa Timur, Sulawesi Selatan, Sulawesi Utara, dan Jawa Barat. *Cluster* 1 memiliki nilai IPM sedang dengan rata-rata 68,82, meliputi Kab/Kota dari Provinsi Nusa Tenggara Barat, Sulawesi Tenggara, Papua, Aceh, dan Sumatera Utara. *Cluster* 2 merupakan cluster dengan nilai IPM sangat tinggi dengan rata-rata 81,03, mencakup Kab/Kota dari Provinsi Jawa Timur, Jawa Tengah, Jawa Barat, dan D.I. Yogyakarta. Sementara itu, *Cluster* 3 memiliki nilai IPM rendah dengan rata-rata 52,65, yang didominasi oleh Kab/Kota dari Provinsi Papua dan Papua Barat. Setelah mengetahui hasil pengelompokan ini, pemerintah Provinsi di Indonesia dapat memberikan arahan dan kebijakan yang sesuai dengan karakteristik masing-masing *cluster*, khususnya kepada Kab/Kota di *cluster* 1 dan 3, supaya dapat meningkatkan IPM di wilayahnya melalui penyusunan program kegiatan yang lebih fokus pada indikator penyusun IPM.

DAFTAR PUSTAKA

[1] BPS, “indeks-pembangunan-manusia-2020,” 2020.
[2] BPS, “indeks-pembangunan-manusia-2021,” 2021.
[3] BPS, “2033,” 2023.
[4] E. A. Saputra and Y. Nataliani, “Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means,” *Journal of Information Systems and Informatics*, vol. 3, no. 3, 2021, [Online]. Available: http://journal-isi.org/index.php/isi
[5] N. Silalahi, “Penentuan Strategi Promosi Universitas Budi Darma Menggunakan Algoritma K-Means Clustering,” *Terapan Informatika Nusantara*, vol. 1, no. 1, pp. 40–46, 2020.
[6] R. N. Puspita, S. A. Pertama, D. Kependudukan, P. Sipil, and K. Tangerang, “ANALISIS K-MEANS CLUSTER PADA KABUPATEN/KOTA DI PROVINSI BANTEN BERDASARKAN INDIKATOR INDEKS PEMBANGUNAN MANUSIA,” vol. 2, no. 3, 2021, doi: 10.46306/lb.v2i3.

[7] L. Anggraini and P. Rismawati Arum, "Analisis Cluster Menggunakan Algoritma K-Means Pada Provinsi Sumatera Barat Berdasarkan Indeks Pembangunan Manusia Tahun 2021 Cluster Analysis Using the K-Means Algorithm in West Sumatra Province Based on the 2021 Human Development Index," 2022.

[8] Y. Ratna Sari, A. Sudewa, D. Ayu Lestari, and T. Ika Jaya, "PENERAPAN ALGORITMA K-MEANS UNTUK CLUSTERING DATA KEMISKINAN PROVINSI BANTEN MENGGUNAKAN RAPIDMINER," 2020.

[9] R. Maliki *et al.*, "Perbandingan Tingkat Pengangguran Terbuka Provinsi di Indonesia Berbasis Metode K-Means Clustering," 2022. [Online]. Available: http://jurnal.bsi.ac.id/index.php/co-science

[10] Yucheng Kao and Szu-Yuan Lee, *2009 IEEE International Conference on Intelligent Computing and Intelligent Systems*. I E E E, 2009.

[11] P. Vania and B. Nurina Sari, "Perbandingan Metode Elbow dan Silhouette untuk Penentuan Jumlah Klaster yang Optimal pada Clustering Produksi Padi menggunakan Algoritma K-Means," *Jurnal Ilmiah Wahana Pendidikan*, vol. 9, no. 21, pp. 547–558, 2023, doi: 10.5281/zenodo.10081332.

[12] S. Paembonan, H. Abduh, and K. Kunci, "Penerapan Metode Silhouette Coeficient Untuk Evaluasi Clutering Obat Clustering; K-means; Silhouette coeficient," 2021. [Online]. Available: https://ojs.unanda.ac.id/index.php/jiit/index

[13] D. Ayu, I. C. Dewi, and K. Pramita, "Analisis Perbandingan Metode Elbow dan Sillhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," 2019.

[14] W. Rohalidyawati, R. Rahmawati, D. Statistika, and F. Sains dan Matematika, "SEGMENTASI PELANGGAN E-MONEY DENGAN MENGGUNAKAN ALGORITMA DBSCAN (DENSITY BASED SPATIAL CLUSTERING APPLICATIONS WITH NOISE) DI PROVINSI DKI JAKARTA," *JURNAL GAUSSIAN*, vol. 9, no. 2, pp. 162–169, 2020, [Online]. Available: https://ejournal3.undip.ac.id/index.php/gaussian/

[15] J. Han, M. Kamber, and J. Pei, "Data Mining: Concepts and Techniques."

[16] I. T. Umagapi, B. Umaternate, S. Komputer, P. Pasca Sarjana Universitas Handayani, B. Kepegawaian Daerah Kabupaten Pulau Morotai, and B. Riset dan Inovasi, "Uji Kinerja K-Means Clustering Menggunakan Davies-Bouldin Index Pada Pengelompokan Data Prestasi Siswa."

[17] A. Anilshi *et al.*, "Universitas Kristen Wira Wacana Sumba Fakultas Sains dan Teknologi SATI: Sustainable Agricultural Technology Innovation Homepage: https://ojs.unkriswina.ac.id/index.php/semnas-FST 3rd Nasional Seminar on Sustainable Agricultural Technology Innovation 2 Agustus 2024/ Pages: 720-743 PENERAPAN ALGORITMA K-MEANS CLUSTERING NILAI KEMAMPUAN SISWA DALAM PELAJARAN MATEMATIKA Application Of The K-Means Clustering Algorithm Of Student Abilities In Mathematics Lessons." [Online]. Available: https://ojs.unkriswina.ac.id/index.php/semnas-FST

[18] A. Z. Wijaya and I. Sembiring, "ANALISIS PERILAKU PENGGUNA INTERNET DENGAN METODE K-MEANS CLUSTERING DAN PENDEKATAN DAVIES BOULDIN INDEX MENGGUNAKAN DATA LOG UNIVERSITAS XYZ," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 2, pp. 878–888, May 2024, doi: 10.29100/jipi.v9i2.4750.

[19] B. Surarso and R. Gernowo, "IMPLEMENTATION OF K-MEDOIDS CLUSTERING FOR HIGH EDUCATION ACCREDITATION DATA," *Jurnal Ilmiah KURSOR*, vol. 10, no. 3, pp. 119–128, 2020.

[20] D. J. Lubis and M. B. Tamam, "Penerapan K-Means Untuk Pengelompokkan Beasiswa Santri di Pondok Pesantren Miftahul Huda Bogor," vol. 12, pp. 7–20, 2022, doi: 10.36350/jbs.v12i1.