

Penerapan K-Means Clustering dalam Mengelompokkan Kabupaten/Kota Berdasarkan Indikator IPM di Jawa Tengah

Muhammad Reza Fadillah^{1,*}, Muhammad Rifqy Ardiansyah², Muhammad Rifki Junaidi³, Fatkhurokhman Fauzi⁴

^{1,2,3,4} Program Studi Statistika, Universitas Muhammadiyah Semarang, Jl. Kedungmundu No.18, Kedungmundu, Kota Semarang Jawa Tengah 50273, Indonesia

*Corresponding author: reza04667@gmail.com



P-ISSN: 2986-4178
E-ISSN: 2988-4004

Riwayat Artikel

Dikirim: 08 Januari 2025
Direvisi: 21 Agustus 2025
Diterima: 21 Agustus 2025

ABSTRAK

IPM merupakan ukuran keberhasilan pembangunan manusia yang didasarkan pada tiga dimensi utama, yaitu kesehatan, pendidikan, dan standar hidup layak. Jawa Tengah dipilih sebagai objek penelitian karena menunjukkan tren peningkatan IPM yang konsisten, namun masih terdapat disparitas antar daerah. Penelitian ini bertujuan untuk mengelompokkan Kabupaten/Kota di Jawa Tengah berdasarkan indikator pembentuk IPM tahun 2023 dengan menggunakan metode analisis cluster non-hirarki K-Means. Metode ini dipilih karena efisien, sederhana, dan mampu menghasilkan pengelompokan yang jelas. Cluster pertama terdapat 16 Kabupaten/kota dengan indikator dari faktor-faktor yang berpengaruh pada IPM tingkatan rendah, cluster kedua terdapat 4 Kabupaten/kota dengan indikator dari faktor-faktor yang berpengaruh pada IPM tingkatan tinggi, cluster ketiga terdapat 15 Kabupaten/kota dengan indikator dari faktor-faktor yang berpengaruh pada IPM tingkatan sedang.

Kata Kunci: Clustering, K-Means, IPM

1. Pendahuluan

Pembangunan manusia merupakan inti dari upaya pembangunan yang dilakukan oleh suatu negara atau daerah [1]. Tujuan utama pembangunan manusia adalah untuk meningkatkan kualitas hidup manusia dan memperluas pilihan-pilihan yang tersedia bagi mereka [2]. Hal ini sejalan dengan pandangan bahwa manusia adalah aset dan tujuan utama dari pembangunan itu sendiri, bukan semata-mata sebagai alat untuk mencapai pertumbuhan ekonomi [3]. Jawa Tengah dipilih sebagai objek penelitian karena menunjukkan tren peningkatan Indeks Pembangunan Manusia (IPM) yang konsisten, yaitu 71,87 pada 2020 menjadi 73,39 pada 2023. Kenaikan tersebut mencerminkan adanya perbaikan kualitas hidup masyarakat, sehingga Jawa Tengah relevan untuk dikaji dalam konteks pembangunan manusia.

Analisis cluster non-hirarki K-means mengkategorikan objek ke dalam kelompok cluster berdasarkan jarak mereka ke cluster pusat terdekat. K-Means sendiri sangat efisien dalam menangani kumpulan data yang sangat besar. Metode clustering K-Means digunakan dalam penelitian ini untuk mengelompokkan Kabupaten/Kota di Jawa Tengah

berdasarkan indikator IPM. Pengelompokan ini akan memberikan informasi mengenai karakteristik dan perbedaan antar Kabupaten/Kota dalam hal capaian pembangunan manusia. Penelitian ini dharapkan dapat membantu pengambilan kebijakan serta membuat program pembangunan yang lebih sesuai dengan kebutuhan masing-masing wilayah.

Berikut merupakan beberapa penelitian terdahulu yang menjadi acuan dalam penulisan artikel ini. Pada penelitian (Dwi Fitrianingsih dan Mujiati Dwi Kartikasari, 2024) dilakukan pengelompokan pada Kabupaten dan kota di Jawa Barat berdasarkan faktor-faktor yang mempengaruhi IPM [4]. Dengan menggunakan K-Means Clustering diperoleh sebanyak 5 cluster.

Pada penelitian (S N Mayasari dan J Nugraha, 2022) dilakukan pengelompokan wilayah Kabupaten/kota di Jawa Tengah berdasarkan kemiskinan [5]. Dengan menggunakan K-Means Clustering dan Metode Elbow diperoleh sebanyak 3 cluster.

2. Metodologi Penelitian

2.1 Data dan sumber data

Data yang digunakan dalam penelitian ini menggunakan data sekunder yang bersumber dari publikasi Badan Pusat Statistik (BPS) [6]. Faktor data yang digunakan adalah data terkait faktor-faktor yang memengaruhi indeks pembangunan manusia di Provinsi Jawa Tengah. Variabel yang digunakan dalam penelitian ini meliputi Angka Harapan Hidup, Harapan Lama Sekolah, Rata-rata lama sekolah, Pengeluaran per kapita, Indeks Pembangunan Manusia.

2.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah Angka Harapan Hidup, Harapan Lama Sekolah, Rata-rata Lama Sekolah, Pengeluaran per Kapita, dan Indeks Pembangunan Manusia.

2.3 Metode Penelitian

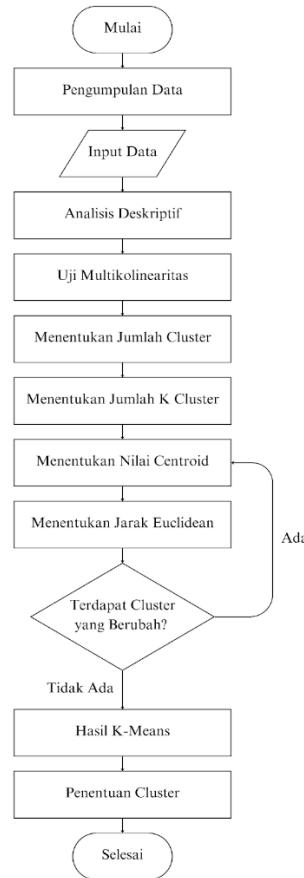
Dalam melakukan penelitian, metode yang digunakan yaitu analisis statistika deskriptif dan Analisis K-Means Clustering. Analisis deskriptif sendiri digunakan untuk melihat gambaran umum dari data IPM dan komponen yang mempengaruhinya.

Metode data clustering non hirarki K-Means membuat bentuk satu atau lebih cluster/kelompok. kelompok-kelompok dalam metode ini memiliki karakteristik yang sama dengan masing-masing kelompok. Analisis K-Means Clustering digunakan untuk mengelompokkan Kabupaten/Kota berdasarkan komponen yang mempengaruhi IPM [7]. Dalam algoritma K-Means, jumlah cluster dan pusat cluster yang dipilih secara acak harus ditetukan terlebih dahulu. selanjutnya, objek yang dimasukkan ke dalam cluster berdasarkan dari jarak terdekat objek ke titik pusat cluster. Pada iterasi selanjutnya, pusat cluster dapat dihitung dengan menggunakan persamaan sebagai berikut

$$v_{lj} = \left(\frac{1}{n_l} \right) \sum_{k=1}^{n_l} X_{kj}$$

dengan v_{lj} berfungsi sebagai pusat cluster ke-l pada variabel ke-j, x_{kj} berfungsi sebagai objek data ke-k pada variabel ke-j, dan banyak objek pada cluster ke-l. Ketika objek tidak lagi berpindah antar cluster, maka proses iterasi akan dihentikan.

software yang digunakan dalam penelitian ini meliputi Microsoft Excel, Google Collab, tableau, dan datawrapper.



Gambar 1. Diagram Alur Penelitian

Adapun tahapan penelitian yang dilakukan berdasarkan Gambar sebagai berikut:

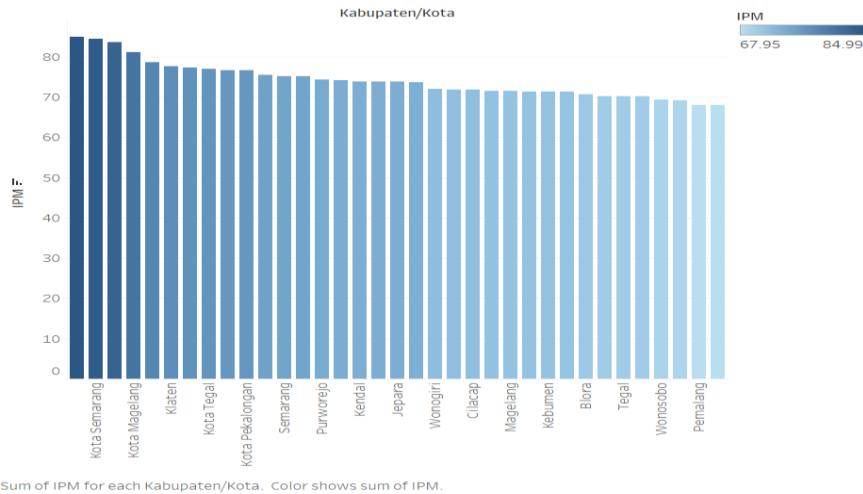
- 1) Mengidentifikasi dan merumuskan masalah yang diangkat dalam penelitian.
- 2) Mengumpulkan data mentah tentang faktor-faktor yang mempengaruhi Indeks Pembangunan Manusia (IPM) seperti; Angka Harapan Hidup (persen), Harapan Lama Sekolah (persen), Rata-rata Lama Sekolah (persen), dan Pengeluaran per Kapita Disesuaikan (ribu rupiah).
- 3) Melakukan analisis deskriptif untuk melihat gambaran umum data IPM di setiap Kabupaten/Kota di Jawa Tengah.
- 4) Melakukan pengujian asumsi cluster yang mencakup uji kecukupan sampel dengan menggunakan uji multikolinieritas dengan nilai VIF dan deteksi outlier.
- 5) Menggunakan metode Elbow untuk menentukan jumlah cluster K.
- 6) Menentukan nilai centroid cluster.
- 7) Menghitung jarak antara objek ke centroid cluster dengan menggunakan jarak Euclidean.
- 8) Menginterpretasikan hasil cluster, pemetaan cluster dan penarikan kesimpulan.

3. Hasil dan Pembahasan

3.1 Analisis Deskriptif

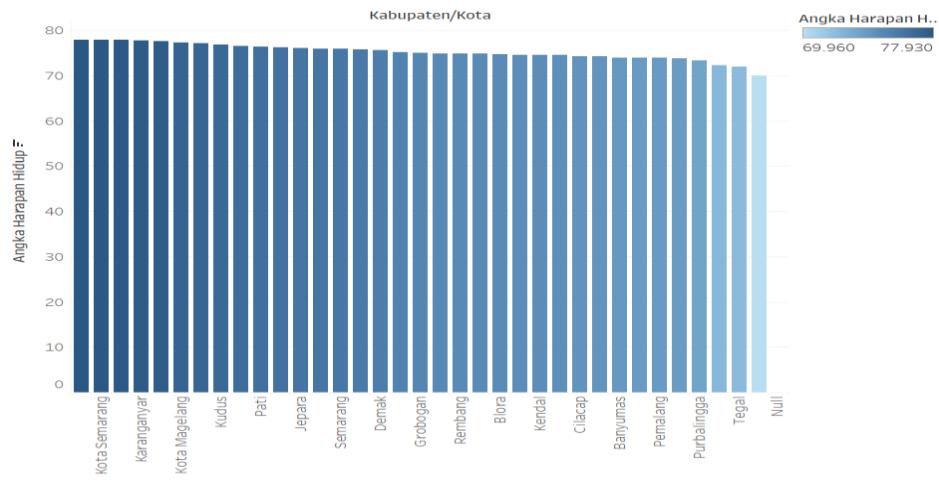
Analisis statistik deskriptif merupakan bagian dari statistik yang mempelajari cara pengumpulan data dan penyajian data sehingga mudah dipahami. Sebelum melakukan

analisis cluster perlu dilihat untuk karakteristik dari data faktor-faktor yang mempengaruhi Indeks Pembangunan Manusia di Provinsi Jawa Tengah pada tahun 2023.



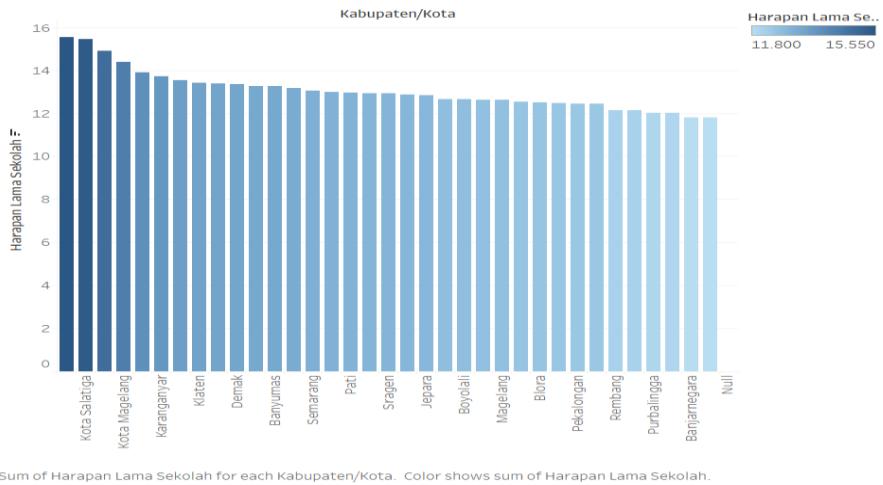
Gambar 2.

Berdasarkan grafik tingkat IPM tahun 2023, dapat dilihat bahwa data tertinggi ada pada Kota Salatiga dengan nilai 84,99. sedangkan data terendah adalah Kabupaten Brebes yang bernilai 67,95.



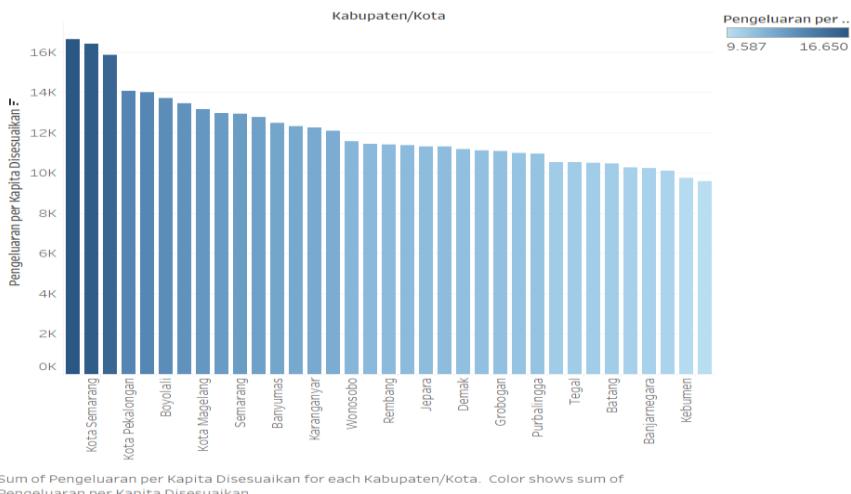
Gambar 3.

Berdasarkan grafik tingkat Angka Harapan Hidup tahun 2023, dapat dilihat bahwa data tertinggi ada pada Kota Salatiga dengan nilai 77,93. sedangkan data terendah adalah Kabupaten Brebes yang bernilai 69,96.



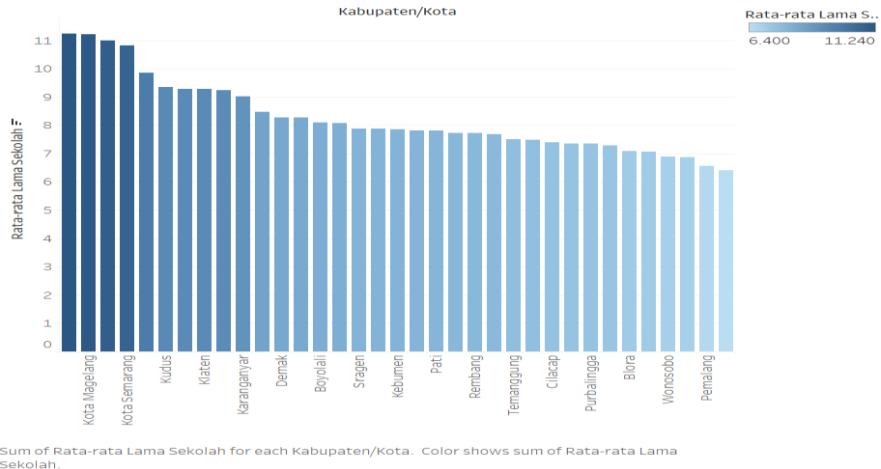
Gambar 4.

Berdasarkan grafik tingkat Harapan Lama Sekolah tahun 2023, dapat dilihat bahwa data tertinggi ada pada Kota Semarang dengan nilai 15,55. sedangkan data terendah adalah Kota Wonosobo yang bernilai 11,80.



Gambar 5.

Berdasarkan grafik tingkat Pengeluaran per Kapita Disesuaikan tahun 2023, dapat dilihat bahwa data tertinggi ada pada Kota Salatiga dengan nilai 16.650. sedangkan data terendah adalah Kota Pemalang yang bernilai 9.587.



Gambar 6.

Berdasarkan grafik tingkat Rata-rata Lama Sekolah tahun 2023, dapat dilihat bahwa data tertinggi ada pada Kota Salatiga dengan nilai 11,24. sedangkan data terendah adalah Kabupaten Brebes yang bernilai 6,40.

Sehingga berdasarkan grafik barchart yang menyajikan faktor-faktor yang mempengaruhi nilai Indeks Pembangunan Manusia (IPM) tahun 2023. Diperoleh nilai pada faktor Pengeluaran per Kapita Disesuaikan (PPKD) sangat berpengaruh terhadap IPM. Secara garis besar Kabupaten Brebes memiliki nilai IPM yang terendah. Kemudian, Kabupaten Purworejo salah satu daerah yang IPMnya berada di tingkatan sedang dan Kota Salatiga menjadi daerah yang memiliki nilai IPM tertinggi. Hasil klastering dapat membantu pemerintah dalam merumuskan kebijakan yang lebih tepat sasaran, karena setiap kelompok daerah dapat memperoleh intervensi pembangunan yang sesuai dengan kondisi dan kebutuhannya.

3.2 Asumsi analisis K-Means Cluster

Tabel 1.

Kabupaten/Kota	AHH	HLS	RRLS	PKD	IPM
Cilacap	74.25	12.67	7.39	11432	71.83
Banyumas	73.98	13.26	7.87	12492	73.86
Purbalingga	73.37	12.02	7.34	10964	70.24
Banjarnegara	74.47	11.82	6.86	10226	69.14
Kebumen	73.83	13.37	7.86	9734	71.37
Purworejo	75.21	13.53	8.46	11110	74.28
Wonosobo	72.17	11.8	6.89	11577	69.37
Magelang	74.2	12.61	7.82	10493	71.45
Boyolali	76.23	12.66	8.09	13716	75.41
Klaten	77.07	13.41	9.27	12968	77.59
Sukoharjo	77.86	13.91	9.84	12319	78.65
Wonogiri	76.56	12.52	7.67	10283	71.97
Karanganyar	77.72	13.71	9.02	12260	77.31
Sragen	75.97	12.92	7.87	13439	75.1

Grobogan	75.04	12.46	7.28	11083	71.49
Blora	74.71	12.51	7.08	10541	70.63
Rembang	74.77	12.15	7.72	11399	71.89
Pati	76.39	12.96	7.8	11385	73.59
Kudus	76.86	13.26	9.34	12088	76.71
Jepara	76.04	12.85	8.26	11306	73.85
Demak	75.6	13.34	8.27	11166	74.07
Semarang	75.95	13.05	8.07	12943	75.13
Temanggung	75.77	12.61	7.5	10108	71.33
Kendal	74.58	12.99	7.73	12755	73.86
Batang	74.85	12.15	7.07	10470	70.2
Pekalongan	73.87	12.44	7.47	11297	71.4
Pemalang	73.85	12.01	6.55	9587	68.03
Tegal	72	12.92	7.34	10537	70.23
Brebes	69.96	12.44	6.4	10993	67.95
Kota Magelang	77.22	14.4	11.2	13175	81.17
Kota Surakarta	77.63	14.9	11	15870	83.54
Kota Salatiga	77.93	15.44	11.24	16650	84.99
Kota Semarang	77.9	15.55	10.81	16420	84.43
Kota Pekalongan	74.6	12.87	9.29	14056	76.71
Kota Tegal	74.77	13.18	9.24	14013	77.02

3.2.1 Uji Multikolinearitas

Tabel. 2 Hasil Uji Multikolinearitas

Indikator	VIF	Tolerance
AHH	0,421	2,373
HLS	0,174	5,733
RLS	0,113	8,833
PPKD	0,330	3,034

1. Hipotesis:

H0: Tidak terjadi multikolinieritas antara variabel Angka Harapan Hidup/AHH(X1), Harapan Lama Sekolah/HLS(X2), Rata-rata Lama Sekolah/RLS(X3), dan Pengeluaran per Kapita Disesuaikan/PPKD(X4), dengan Indeks Pembangunan Manusia(Y).

H1: Terjadi multikolinieritas antara variabel Angka Harapan Hidup/AHH(X1), Harapan Lama Sekolah/HLS(X2), Rata-rata Lama Sekolah/RLS(X3), dan Pengeluaran per Kapita Disesuaikan/PPKD(X4), dengan Indeks Pembangunan Manusia(Y).

2. Statistik Uji:

VIF AHH = 2,373

VIF HLS = 5,733

VIF RLS = 8,833

VIF PPKD = 3,034

3. Kriteria Pengujian:

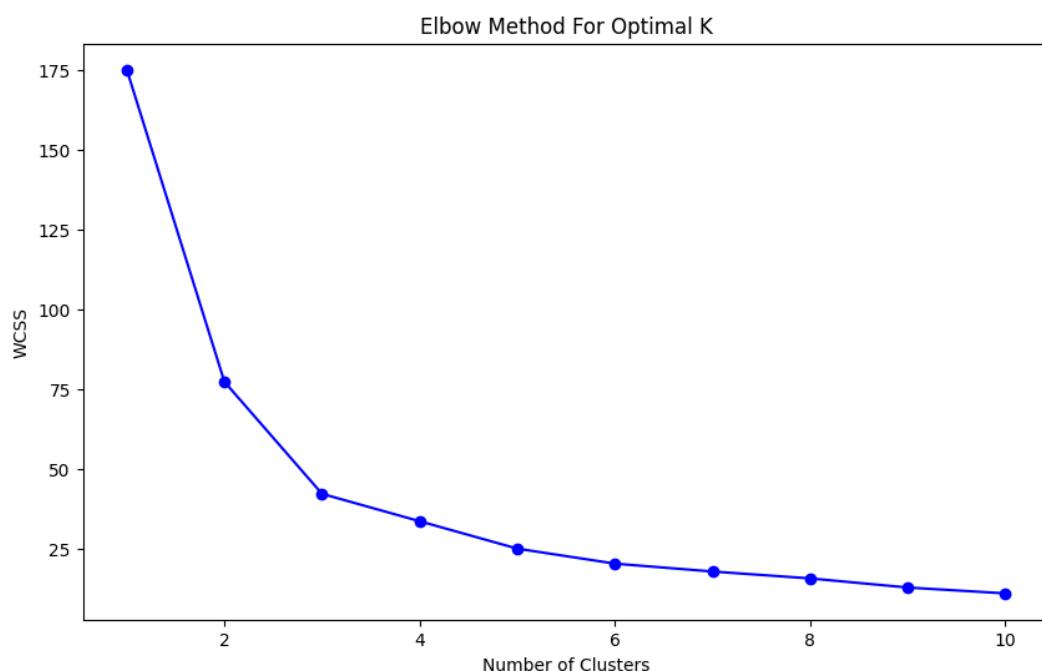
H0 ditolak jika nilai VIF > 10.

4. Keputusan & Kesimpulan:

Berdasarkan Tabel 2 dapat diketahui bahwa VIF < 10, maka H0 diterima. Jadi, dapat disimpulkan bahwa tidak terjadi multikolinieritas antara variabel Angka Harapan Hidup/AHH(X1), Harapan Lama Sekolah/HLS(X2), Rata-rata Lama Sekolah/RLS(X3), dan Pengeluaran per Kapita Disesuaikan/PPKD(X4), dengan Indeks Pembangunan Manusia(Y).

3.2.2 Penentuan Jumlah Cluster

Menentukan jumlah cluster melalui pendekatan menggunakan metode elbow untuk menentukan cluster optimal.



Gambar 7.

Berdasarkan output, jumlah cluster yang ditentukan berdasarkan metode Elbow adalah $k = 3$. Hal tersebut dikarenakan titik yang membentuk siku pada metode Elbow adalah pada $k = 3$ dan juga setelah $k = 3$ kemudian diikuti oleh nilai SSE yang relatif konstan.

3.2.3 Titik Pusat Cluster (Centroid)

Setelah menentukan jumlah cluster selanjutnya peneliti mencari nilai centroid untuk masing-masing cluster. Nilai centroid diperoleh dari perhitungan nilai rata-rata pada

setiap cluster dengan iterasi pertama diberikan secara acak sampai dengan nilai initial cluster center. Proses pengelompokan berhenti jika nilai initial cluster center yang baru sama dengan initial cluster yang lama. Misalkan diperoleh nilai initial cluster center;

Tabel 3.

AHH	HLS	RRLS	PPKS	IPM
-0.70069223	-0.71352282	-0.76640355	-0.76472382	-0.83443741
1.36094476	2.24959929	2.17468528	1.97893995	2.15149696
0.38448645	0.16119786	0.23758104	0.28798809	0.31633404

Nilai initial cluster center pada Tabel 2 diatas, digunakan untuk menghitung jarak antara data dengan centroid. Persamaan yang digunakan untuk menghitung jarak pada penelitian ini adalah Euclidean Distance.

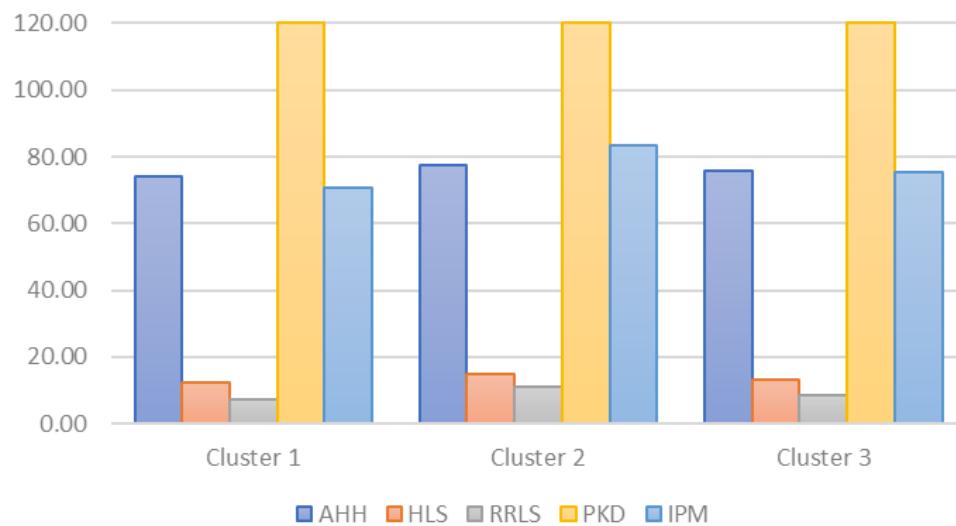
Hasil Cluster K-Means

Tabel 4.

Cluster	Jumlah	Anggota Cluster
1	16	Cilacap, Purbalingga, Banjarnegara, Kebumen, Wonosobo, Magelang, Wonogiri, Grobogan, Blora, Rembang, Temanggung, Batang, Pekalongan, Pemalang, Tegal, Brebes.
2	4	Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang.
3	15	Banyumas, Purworejo, Boyolali, Klaten, Sukoharjo, Karanganyar, Sragen, Pati, Kudus, Jepara, Demak, Semarang, Kendal, Kota Pekalongan, Kota Tegal

Berdasarkan hasil clustering pada Tabel 3 diatas, menggunakan Google Collab diperoleh bahwa terdapat 3 cluster pada faktor-faktor yang mempengaruhi IPM per Kabupaten/kota di Provinsi Jawa Tengah. Pada cluster 1 terdapat 16 Kabupaten/Kota, cluster 2 terdapat 4 Kabupaten/Kota,dan cluster 3 terdapat 15 Kabupaten/Kota. Setelah

itu, peneliti melakukan profilisasi data untuk menggambarkan karakteristik masing-masing kelompok dengan menggunakan rata-rata nilai pada setiap cluster.



Gambar 8.

Berdasarkan hasil clustering pada Gambar diatas, dapat dilihat hasil profilisasi terdapat perbedaan yang tidak terlalu signifikan dari ketiga cluster yang terbentuk. Pada cluster 2 Jumlah IPM dan seluruh Indikator (Angka Harapan Hidup, Harapan Lama Sekolah, Rata Rata Lama Sekolah dan Pengeluaran per Kapita Disesuaikan) memiliki jumlah tertinggi dibandingkan dua cluster lainnya. Pada cluster 3 Jumlah IPM dan jumlah IPM dan seluruh Indikator (Angka Harapan Hidup, Harapan Lama Sekolah, Rata Rata Lama Sekolah dan Pengeluaran per Kapita Disesuaikan) memiliki jumlah sedang. Dan pada cluster 1 Jumlah IPM dan seluruh Indikator (Angka Harapan Hidup, Harapan Lama Sekolah, Rata Rata Lama Sekolah dan Pengeluaran per Kapita Disesuaikan) memiliki jumlah rendah.

Sekolah, Rata Rata Lama Sekolah dan Pengeluaran per Kapita Disesuaikan) memiliki jumlah yang rendah.

3.3 Visualisasi dengan Datawrapper



Gambar 9.

Dengan menggunakan software Google Collab didapatkan 3 cluster, pada visualisasi menggunakan software Datawrapper didapatkan cluster 1 dengan tingkatan rendah ditandai dengan berwarna ungu, cluster 2 dengan tingkatan sedang ditandai dengan berwarna hijau,

4. Kesimpulan

Berdasarkan Hasil analisis Clustering yang dilakukan dalam penelitian ini menghasilkan kesimpulan sebagai berikut :

1. Menurut analisis deskriptif nilai pada faktor PKD (Pengeluaran per Kapita Disesuaikan) sangat berpengaruh dalam IPM. Secara garis besar Kabupaten Cilacap menjadi salah satu yang rendah. Kemudian, Kota Semarang salah satu daerah yang IPMnya berada di tingkatan tinggi dan Kabupaten Banyumas berada di tingkatan sedang.
 2. Cluster 1 dengan indikator dari faktor-faktor yang berpengaruh pada IPM tingkatan rendah memiliki 16 titik yaitu Kabupaten Cilacap, Kabupaten Purbalingga, Kabupaten Banjarnegara, Kabupaten Kebumen, Kabupaten Wonosobo, Kabupaten Magelang, Kabupaten Wonogiri, Kabupaten Grobogan, Kabupaten Blora, Kabupaten Rembang, Kabupaten Temanggung, Kabupaten Batang, Kabupaten Pekalongan, Kabupaten Pemalang, Kabupaten Tegal, Kabupaten Brebes. Cluster 2 dengan indikator dari faktor-faktor yang berpengaruh pada IPM tingkatan tinggi memiliki 4 titik yaitu Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang. Cluster 3 dengan indikator dari faktor-faktor yang berpengaruh pada IPM tingkatan sedang memiliki 15 titik yaitu Kabupaten Banyumas, Kabupaten Purworejo, Kabupaten Boyolali, Kabupaten Klaten, Kabupaten Sukoharjo, Kabupaten Karanganyar, Kabupaten Kabupaten Sragen, Kabupaten Pati, Kabupaten Kudus, Kabupaten Jepara, Kabupaten Demak, Kabupaten Semarang, Kabupaten Kendal, Kota Pekalongan, Kota Tegal. Daerah

dengan IPM rendah umumnya masih menghadapi keterbatasan dalam pendidikan, kesehatan, dan daya beli, sehingga diperlukan kebijakan yang berfokus pada peningkatan akses dan pemerataan pembangunan. Sebaliknya, daerah dengan IPM tinggi menunjukkan kondisi pembangunan manusia yang relatif lebih baik, sehingga perlu menjaga kualitas layanan publik serta menjadi contoh bagi daerah lain.

5. Daftar Pustaka

- [1] N. Ramadhanisa and N. Triwahyuningtyas, “Analisis Faktor Yang Mempengaruhi Indeks Pembangunan Manusia Di Provinsi Lampung,” *SIBATIK J. J. Ilm. Bid. Sos. Ekon. Budaya, Teknol. dan Pendidik.*, vol. 1, no. 7, pp. 1049–1061, 2022, doi: 10.54443/sibatik.v1i7.121.
- [2] S. L. Hasibuan, Rujiman, and Sukardi, “Analisis Determinan Indeks Pembangunan Manusia (IPM) di Indonesia,” *J. Penelit. Pendidik. Sos. Hum.*, vol. 5, no. 2, pp. 139–141, 2020.
- [3] UNPD, *Human Development Report 2019: beyond income, beyond averages, beyond today*. 2019.
- [4] D. Fitrianingsih and M. D. Kartikasari, “Penerapan K-Means Clustering dengan Metode Elbow untuk Mengelompokkan Kabupaten / Kota Berdasarkan Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Provinsi Jawa Barat,” vol. 2, no. 2, pp. 250–257, 2024.
- [5] S. N. Mayasari and J. Nugraha, “Implementasi K-Means Cluster Analysis untuk Mengelompokkan Kabupaten/Kota Berdasarkan Data Kemiskinan di Provinsi Jawa Tengah Tahun 2022,” *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 3, no. 2, pp. 317–329, 2023, doi: 10.24002/konstelasi.v3i2.7200.
- [6] BADAN PUSAT STATISTIK KABUPATEN SEMARANG, “No Title.” <https://semarangkab.bps.go.id/indicator/26/84/1/indeks-pembangunan-manusia-ipm-di-provinsi-jawa-tengah.html>.
- [7] Belia Mailien, A. Salma, Syafriandi, and D. Fitria, “Comparison K-Means and Fuzzy C-Means Methods to Grouping Human Development Index Indicators in Indonesia,” *UNP J. Stat. Data Sci.*, vol. 1, no. 1, pp. 23–30, 2023, doi: 10.24036/ujsds/vol1-iss1/4.