

# Penerapan *K-Means Clustering* dengan Metode *Elbow* untuk Mengelompokkan Kabupaten/Kota Berdasarkan Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Provinsi Jawa Barat

Dwi Fitriani<sup>1,\*</sup>, Mujiati Dwi Kartikasari<sup>2</sup>

<sup>1,2</sup> Program Studi Statistika, Universitas Islam Indonesia, Jl. Kaliurang KM 14,5, Kabupaten Sleman Daerah Istimewa Yogyakarta, 55584, Indonesia

\*Corresponding author: [20611071@students.uii.ac.id](mailto:20611071@students.uii.ac.id)



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

**Riwayat Artikel**  
Dikirim: 03 April 2024  
Direvisi: 07 Juni 2024  
Diterima: 20 Juni 2024

## ABSTRAK

Indeks Pembangunan Manusia (IPM) merupakan sebuah ukuran yang digunakan dalam memantau dan mengevaluasi pembangunan manusia. Nilai IPM dipengaruhi oleh 3 komponen, yaitu tingkat pendidikan, tingkat kesehatan, dan standar hidup yang layak. Tujuan penelitian ini adalah untuk mengelompokkan Kabupaten/kota berdasarkan faktor-faktor yang mempengaruhi IPM. Peneliti menggunakan *K-Means clustering* dengan metode *Elbow* dan memperoleh hasil bahwa terdapat 5 *cluster*. *Cluster* pertama terdapat 6 Kabupaten/kota dengan indikator dari faktor-faktor yang berpengaruh pada IPM tingkatan sedang, *cluster* kedua terdapat 5 Kabupaten/kota dengan indikator dari faktor-faktor yang berpengaruh pada IPM tingkatan tinggi, *cluster* ketiga terdapat 4 Kabupaten/kota dengan indikator dari faktor-faktor yang berpengaruh pada IPM tingkatan sangat tinggi, *cluster* keempat terdapat 4 Kabupaten/kota dengan indikator dari faktor-faktor yang berpengaruh pada IPM tingkatan sangat rendah, dan *cluster* kelima terdapat 8 Kabupaten/kota dengan indikator dari faktor-faktor yang berpengaruh pada IPM tingkatan rendah.

**Kata Kunci:** *Cluster, Indeks, IPM, Tingkatan.*

## ABSTRACT

The Human Development Index (HDI) is a measure used to monitor and evaluate human development. The HDI is influenced by three components: education level, health level, and decent standard of living. The purpose of this research is to cluster districts/cities based on factors that influence HDI. Researchers used *K-Means clustering* with the *Elbow* method and obtained the results that there are 5 clusters, in the first cluster there are 6 districts / cities with indicators of factors that influence the HDI at a moderate level, the second cluster there are 5 districts / cities with indicators of factors that influence the HDI at a high level, the third cluster there are 4 districts / cities with indicators of factors that influence the HDI at a very high level, the fourth cluster there are 4 districts / cities with indicators of factors that influence the HDI at a very low level, and the fifth cluster there are 8 districts / cities with indicators of factors that influence the HDI at a low level.

**Keywords:** *Cluster, HDI, Index.*

## 1. Pendahuluan

Pembangunan manusia memiliki tujuan yaitu membuat manusia agar fokus dan menjadi tujuan akhir pada semua aktifitas pembangunan. Tujuan akhir pembangunan manusia yaitu agar setiap masyarakat mampu mengelola dengan baik sumber daya yang ada agar masyarakat bisa hidup lebih bahagia. Indeks pembangunan manusia (IPM) merupakan metrik untuk memantau dan mengevaluasi pembangunan manusia. Apabila IPM suatu wilayah meningkat, maka di wilayah tersebut juga akan meningkat. Hal ini disebabkan oleh fakta bahwa negara kita memiliki kekayaan sumber daya alam yang berlimpah [4].

Apabila pemerintah memanajemen sumber daya alam dengan cermat, maka kekayaan alam dapat meningkatkan pendapatan negara. Namun, faktanya, warga dan tenaga kerja tidak dapat mengelola dan memanfaatkan sumber daya alam secara optimal. Kondisi ini mendorong Indonesia untuk terus membangun, terutama pada bidang sumber daya manusia, namun hasilnya tidak memuaskan [10].

Peneliti melakukan penelitian memberikan masukan kepada Badan Pusat Statistik dalam mengevaluasi IPM dengan menggunakan 5 *cluster*. Hal ini dilakukan untuk melihat apakah setiap *cluster* dapat mewakili salah satu indikator IPM. Penelitian ini berjudul “Penerapan *K-Means Clustering* dengan Metode *Elbow* Untuk Mengelompokkan Kabupaten/Kota Berdasarkan Indeks Pembangunan Manusia di Jawa Barat”.

Berikut merupakan penelitian-penelitian terdahulu yang menjadi acuan dalam penulisan laporan/makalah ini. Pada penelitian (Anita Fitria, Antonito Hornay, dan Gangga Anuraga, 2018) dilakukan pengelompokan pada Kabupaten dan kota di Jawa Timur berdasarkan indikator kemiskinan. Dengan menggunakan *K-Means Clustering* diperoleh sebanyak 4 *cluster*.

Pada penelitian (Rina Yuliana, Hardian Oktavianto, dan Henny Wahyu, 2015) dilakukan pengelompokan wilayah Kabupaten/kota di Jawa Tengah berdasarkan komponen pembentuk IPM. Dengan menggunakan *K-Means Clustering* dan Metode *Elbow* diperoleh sebanyak 2 *cluster*.

## 2. Metodologi Penelitian

### 2.1. Data dan Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder dari publikasi Badan Pusat Statistik (BPS) yaitu faktor yang mempengaruhi indeks pembangunan manusia yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Jawa Barat. Variabel yang digunakan pada penelitian ini yaitu Persentase Penduduk Miskin (persen), Tingkat Pengangguran Terbuka Kabupaten/Kota (persen), dan PDRB per kapita atas dasar harga berlaku menurut Kabupaten/Kota (ribu rupiah) [5].

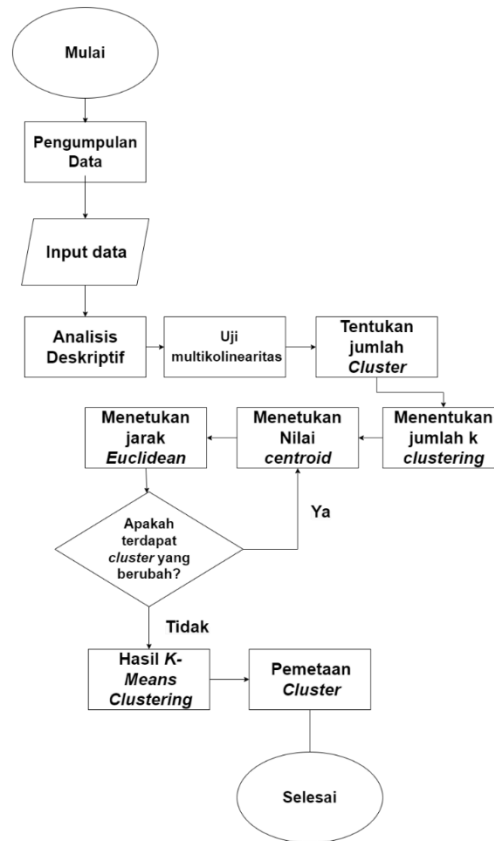
### 2.2. Variabel Penelitian

Variabel yang digunakan pada penelitian ini yaitu Persentase Penduduk Miskin (persen), Tingkat Pengangguran Terbuka Kabupaten/Kota (persen), dan PDRB per kapita atas dasar harga berlaku menurut Kabupaten/Kota (ribu rupiah).

### 2.3. Metode Penelitian

Dalam penelitian ini metode yang dilakukan antara lain analisis statistika deskriptif yang bertujuan untuk melihat gambaran umum dari data IPM dan faktor-faktor yang mempengaruhinya dan analisis *K-Means Clustering* yang bertujuan untuk

mengelompokkan Kabupaten/kota berdasarkan faktor-faktor yang mempengaruhi IPM. Software yang digunakan dalam penelitian ini adalah *Microsoft Excel*, *RStudio*, dan *QGIS*. Tahapan pengolahan data akan ditampilkan pada diagram alir berikut.



**Gambar 1.** Diagram Alur Penelitian

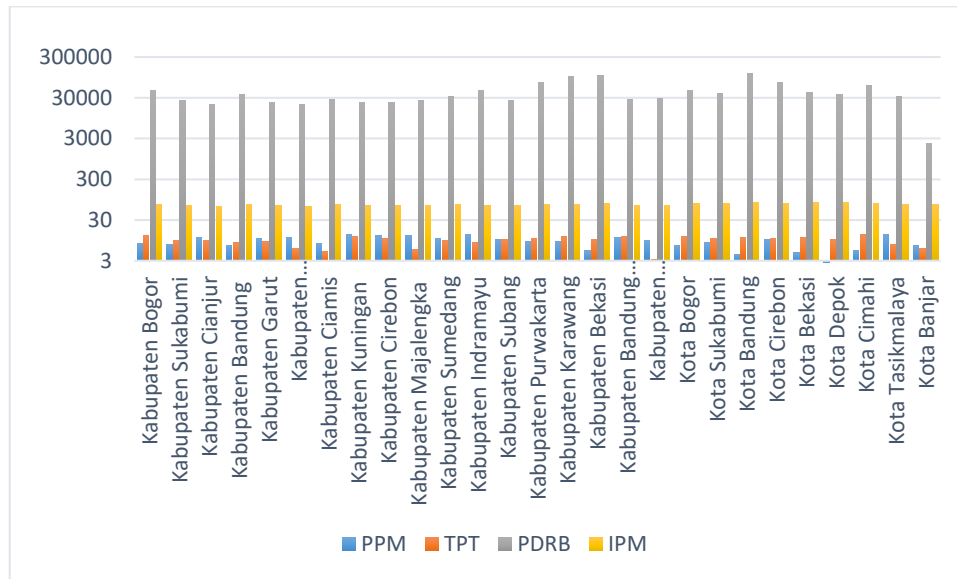
Adapun tahapan penelitian yang dilakukan berdasarkan **Gambar 1** sebagai berikut:

1. Mengidentifikasi dan merumuskan masalah yang diangkat dalam penelitian.
2. Mengumpulkan data mentah faktor yang mempengaruhi Indeks Pembangunan Manusia (IPM) seperti; Persentase Penduduk Miskin (persen), Tingkat Pengangguran Terbuka Kabupaten/kota (persen), dan PDRB per kapita atas dasar harga berlaku menurut Kabupaten/kota (ribu rupiah) Tahun 2021.
3. Melakukan analisis deskriptif untuk melihat gambaran umum data IPM di setiap Kabupaten/Kota di Jawa Barat.
4. Melakukan pengujian asumsi *cluster* meliputi uji kecukupan sampel dengan menggunakan uji multikolinearitas dengan menggunakan nilai VIF, dan deteksi outlier.
5. Menentukan jumlah *cluster* k dengan metode *Elbow*.
6. Menentukan nilai *centroid cluster*.
7. Menghitung jarak objek ke *centroid cluster* dengan menggunakan jarak *Euclidean*.
8. Interpretasi 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 Barat pada tahun 2021.



**Gambar 2.** Barchart Data IPM dan faktor yang mempengaruhinya.

Pada **Gambar 2**, terdapat grafik *barchart* yang menyajikan faktor-faktor yang mempengaruhi indeks pembangunan manusia (IPM) hingga nilai IPM tahun 2021. Diperoleh nilai pada faktor PDRB sangat berpengaruh dalam IPM. Secara garis besar Kabupaten Pangandaran menjadi salah satu yang terendah. Kemudian, Kabupaten Bekasi salah satu daerah yang IPMnya berada di tingkatan sedang dan Kabupaten Bandung menjadi salah satu daerah dengan tingkatan tinggi.

### 3.1. Asumsi Analisis *K-Means Cluster*

Peneliti melakukan analisis *clustering* yang bertujuan untuk mengelompokkan Kabupaten/kota dari faktor-faktor yang mempengaruhi IPM. Penelitian dilakukan dengan menggunakan data faktor yang mempengaruhi indeks pembangunan manusia 2021. Parameter yang digunakan dalam penelitian adalah faktor-faktor yang mempengaruhi IPM berdasarkan Kabupaten/ Kota di Jawa Barat.

**Tabel 1** Perbandingan Nilai ADM dan VIM (Caption)

Kabupaten/Kota	PPM	TPT	PDRB	IPM
Kabupaten Bogor	8.13	12.22	45347	70.6
Kabupaten Sukabumi	7.7	9.51	25615	67.07
Kabupaten Cianjur	11.18	9.32	20002	65.56
Kabupaten Garut	10.65	8.68	23356	66.45
Kabupaten Tasikmalaya	11.15	6.16	20855	65.9
Kabupaten Ciamis	7.97	5.06	27218	70.93
Kabupaten Kuningan	13.1	11.68	22805	69.71
Kabupaten Cirebon	12.3	10.38	22833	69.12
Kabupaten Majalengka	12.33	5.71	25964	67.81
Kabupaten Sumedang	10.71	9.18	32130	71.8
Kabupaten Indramayu	13.04	8.3	44072	67.64
Kabupaten Subang	10.03	9.77	26348	69.13
Kabupaten Purwakarta	8.83	10.7	69985	70.98
Kabupaten Karawang	8.95	11.83	98726	70.94
Kabupaten Bekasi	5.21	10.09	107801	74.45
Kabupaten Bandung Barat	11.3	11.65	26879	68.29
Kabupaten Pangandaran	9.65	3.25	28366	68.28
Kota Bogor	7.24	11.79	45920	76.59
Kota Sukabumi	8.25	10.78	37209	74.6
Kota Bandung	4.37	11.46	121126	81.96
Kota Cirebon	10.03	10.53	72714	75.25
Kota Bekasi	4.74	10.88	39529	81.95
Kota Depok	2.58	9.76	35655	81.37

Kabupaten/Kota	PPM	TPT	PDRB	IPM
Kota Cimahi	5.35	13.07	59929	78.06
Kota Tasikmalaya	13.13	7.66	31558	73.31
Kota Banjar	7.11	6.09	2300	71.92

\*PPM = Presentase Penduduk Miskin (persen); TPT = Tingkat Pengangguran Terbuka (persen); PDRB = Produk Domestik Regional Bruto (ribu rupiah)

### 3.2.1 Uji Multikolinearitas

Sebelum melakukan *clustering* terlebih dahulu dilakukan uji analisis multikolinearitas untuk melihat apakah terdapat hubungan atau kolerasi antar variabel menggunakan *Variance Inflation Factor* (VIF).

**Tabel 2** Nilai VIF IPM

	PPM	TPT	PDRB
IPM	1.230909	1.276521	1.460052

**Tabel 3** Nilai VIF PPM

	TPT	PDRB	IPM
PPM	1.345180	1.479625	1.401771

**Tabel 4** Nilai VIF TPT

	PDRB	IPM	PPM
TPT	1.333863	1.613172	1.418073

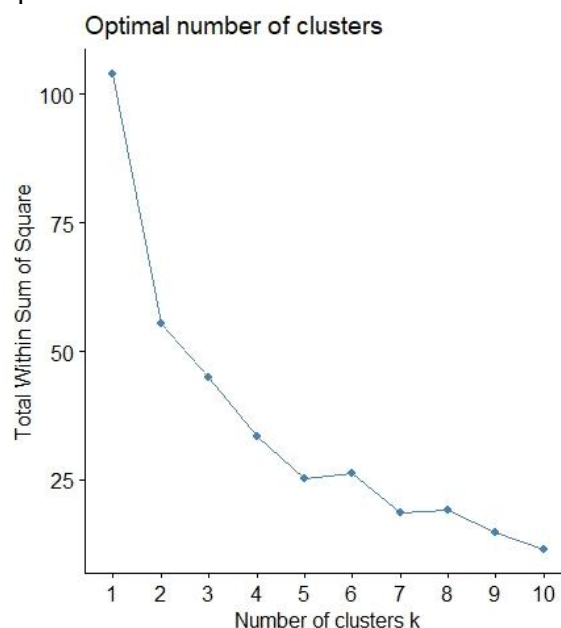
**Tabel 5** Nilai VIF PDRP

	IPM	PPM	TPT
PDRB	2.720821	2.421208	214235

Berdasarkan pada **Tabel 2** hingga **Tabel 5**, diperoleh nilai VIF untuk seluruh variabel  $< 10$ . Sehingga dapat disimpulkan bahwa tidak terdapat multikolinearitas di antara variabel penelitian yang digunakan. Artinya dengan menggunakan hipotesis bahwa  $H_0$  : tidak terjadi multikolinearitas dan  $H_1$  : terjadi multikolinearitas, maka data yang ada gagal tolak  $H_0$  yang berarti asumsi uji multikolinearitas terpenuhi karena nilai-nilai VIF yang dihasilkan kurang dari 10.

### 3.3.1 Penentuan Jumlah Cluster

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



**Gambar 3.** Grafik Optimalisasi *Elbow*.

Berdasarkan **Gambar 3**, jumlah *cluster* yang ditentukan berdasarkan metode *Elbow* adalah  $k = 5$ . Hal tersebut dikarenakan titik yang membentuk siku pada metode *Elbow* adalah pada  $k = 5$  dan juga setelah  $k = 5$  kemudian diikuti oleh nilai SSE yang relatif konstan.

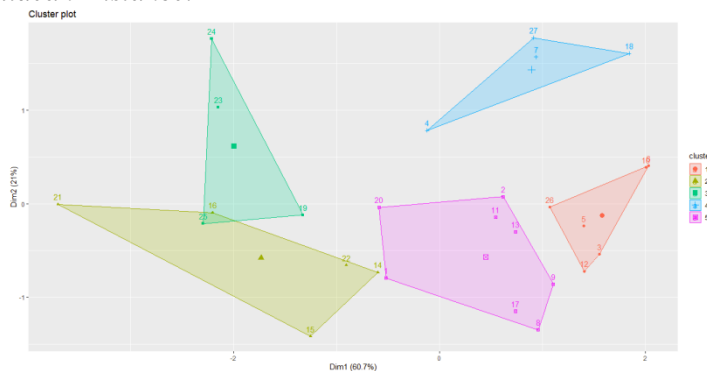
### 3.2.1 Titik Pusat Cluster (Centroid)

Setelah menentukan jumlah *cluster* selanjutnya peneliti menentukan 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 6** Hasil Cluster per-Kabupaten/Kota

IPM	PPM	TPT	PDRB	IPM
10173533	-0.724251	-0.52375071	-0.8706036	10173533
-0.5155119	0.6248691	180.344.550	0.5804695	-0.5155119
-13.796.926	0.8109876	0.09359509	15.795.158	-13.796.926
-0.3454752	-15.288.475	-0.67319402	-0.2040846	-0.3454752
0.4217639	0.511575	-0.44454094	-0.3975564	0.4217639

Nilai *initial cluster center* pada **Tabel 6**, digunakan untuk menghitung jarak antara data dengan *centroid*. Persamaan yang digunakan untuk menghitung jarak pada penelitian ini adalah *Euclidean Distance*.



**Gambar 4.** Cluster Plot.

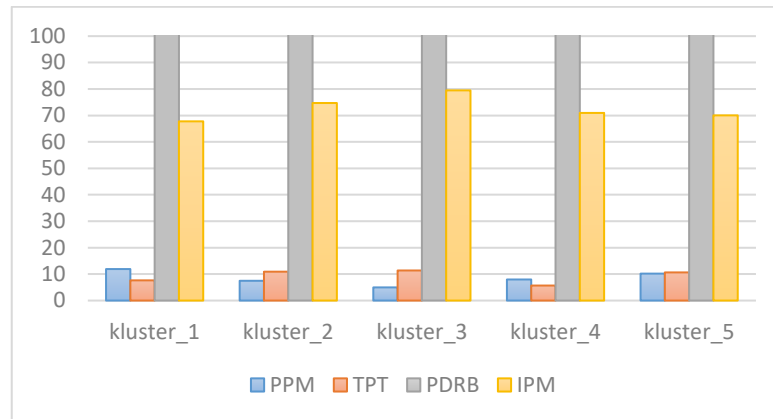
Berdasarkan hasil *cluster* pada **Gambar 4**, menunjukkan bahwa terdapat 3 *cluster* yang terdiri atas *cluster 1*, *cluster 2*, *cluster 3*, *cluster 4*, dan *cluster 5*.

**Tabel 7** Hasil Cluster K-Means

Cluster	Jumlah	Anggota cluster
1	6	Kabupaten Cianjur, Kabupaten Garut, Kabupaten Tasikmalaya, Kabupaten Majalengka, Kabupaten Indramayu, dan Kota Tasikmalaya.
2	5	Kabupaten Purwakarta, Kabupaten Karawang, Kabupaten Bekasi, Kota Bandung, dan Kota Cirebon.
3	4	Kota Bogor, Kota Depok, Kota Bekasi, dan Kota Cimahi.
4	4	Kabupaten Bandung, Kabupaten Ciamis, Kabupaten Pangandaran, dan Kota Banjar.
5	8	Kabupaten Bogor, Kabupaten Sukabumi, Kabupaten Kuningan, Kabupaten Cirebon, Kabupaten Sumedang, Kabupaten Subang, Kabupaten Bandung Barat, Kabupaten Sukabumi. dan Kota Sukabumi.

Berdasarkan hasil *clustering* pada **Tabel 7**, menggunakan *RStudio* diperoleh bahwa terdapat 5 *cluster* pada faktor-faktor yang mempengaruhi IPM per Kabupaten/kota di Provinsi Jawa Barat. Pada *cluster 1* terdapat 6 Kabupaten/Kota, *cluster 2* terdapat 5 Kabupaten/Kota, *cluster 3* terdapat 4 Kabupaten/Kota, *cluster 4* terdapat 4 Kabupaten/Kota,

dan *cluster* 5 terdapat 8 Kabupaten/Kota. Setelah itu, peneliti melakukan profilisasi data untuk menggambarkan karakteristik masing-masing kelompok dengan menggunakan rata-rata nilai pada setiap *cluster*.



**Gambar 5.** Hasil Profilisasi *Cluster K-Means*.

Berdasarkan hasil *clustering* pada **Gambar 5**, dapat dilihat hasil profilisasi terdapat perbedaan yang tidak terlalu signifikan dari kelima *cluster* yang terbentuk. Pada *cluster* 3 jumlah IPM dan TPT (Tingkat Pengangguran Terbuka) memiliki jumlah tertinggi dibandingkan dengan *cluster* lain, serta memiliki jumlah sedang pada PDRB atas dasar harga berlaku, namun perbedaan jumlah tersebut tidak terlalu jauh. Hal ini menunjukkan bahwa pada *cluster* 3 merupakan *cluster* dengan tingkat indikator pada faktor-faktor yang mempengaruhi IPM memiliki tingkat sangat tinggi. Kemudian pada *cluster* 4 jumlah TPT dan PDRB atas dasar harga berlaku memiliki jumlah terendah dibandingkan dengan *cluster* lain. Hal ini menunjukkan bahwa pada *cluster* 4 merupakan *cluster* dengan tingkat indikator pada faktor-faktor yang mempengaruhi IPM memiliki tingkat sangat rendah. Kemudian untuk tingkat tinggi, sedang, dan rendah berturut-turut yaitu *cluster* 3 ditandai dengan memiliki jumlah tertinggi pada PDRB atas dasar harga berlaku serta memiliki jumlah sedang pada TPT dan IPM. Kemudian *cluster* 1 ditandai dengan memiliki jumlah tertinggi pada PPM (Presentase Penduduk Miskin). Selanjutnya *cluster* 5 ditandai dengan memiliki jumlah sedang pada PPM.

### 3.3 Visualisasi dengan *QGIS*



**Gambar 6.** Peta Pesebaran Faktor-Faktor yang Mempengaruhi IPM

Dengan menggunakan *software RStudio* didapatkan 5 *cluster*, pada visualisasi menggunakan *software QGIS* didapatkan *cluster* 1 dengan tingkatan sedang ditandai dengan berwarna jingga, *cluster* 2 dengan tingkatan tinggi ditandai dengan berwarna hijau muda, *cluster* 3 dengan tingkatan sangat tinggi ditandai dengan berwarna hijau tua, *cluster* 4 dengan tingkatan sangat rendah ditandai dengan berwarna biru, dan *cluster* 5 dengan berwarna ungu.

#### 4 Kesimpulan

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

1. Menurut analisis deskriptif nilai pada faktor PDRB sangat berpengaruh dalam IPM. Secara garis besar Kabupaten Pangandaran menjadi salah satu yang sangat rendah. Kemudian, Kabupaten Bekasi salah satu daerah yang IPMnya berada di tingkatan sangat tinggi, Kabupaten Purwakarta berada di tingkatan tinggi dan Kabupaten Bogor menjadi salah satu daerah dengan tingkatan rendah.
2. *Cluster 1* dengan indikator dari faktor-faktor yang berpengaruh pada IPM tingkatan sedang memiliki 6 titik berwarna jingga yaitu Kabupaten Cianjur, Kabupaten Garut, Kabupaten Tasikmalaya, Kabupaten Majalengka, Kabupaten Indramayu, dan Kota Tasikmalaya. *Cluster 2* dengan indikator dari faktor-faktor yang berpengaruh pada IPM tingkatan tinggi memiliki 5 titik berwarna hijau muda yaitu Kabupaten Purwakarta, Kabupaten Karawang, Kabupaten Bekasi, Kota Bandung, dan Kota Cirebon. *Cluster 3* dengan indikator dari faktor-faktor yang berpengaruh pada IPM tingkatan sangat tinggi memiliki 4 titik berwarna hijau tua yaitu Kota Bogor, Kota Depok, Kota Bekasi, dan Kota Cimahi. *Cluster 4* dengan indikator dari faktor-faktor yang berpengaruh pada IPM tingkatan sangat rendah memiliki 4 titik berwarna biru yaitu Kabupaten Bandung, Kabupaten Ciamis, Kabupaten Pangandaran, dan Kota Banjar. *Cluster 5* dengan indikator dari faktor-faktor yang berpengaruh pada IPM tingkatan rendah memiliki 8 titik berwarna ungu yaitu Kabupaten Bogor, Kabupaten Sukabumi, Kabupaten Kuningan, Kabupaten Cirebon, Kabupaten Sumedang, Kabupaten Subang, Kabupaten Bandung Barat, Kabupaten Sukabumi. dan Kota Sukabumi.

#### 5 Daftar Pustaka

- [1] Fatimatu Zahro, N. S. (2022). *Analisis Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Provinsi Jawa Barat* [Skripsi]. Universitas Islam Indonesia.
- [2] Johnson, R. A., & Bhattacharyya, G. K. (2010). *STATISTICS: Principles & Methods* (Sixth Edition). John Wiley & Sons, Inc.
- [3] Hidayat, A. (2014, Maret). *Analisis Cluster Non Hirarki dengan SPSS*. Retrieved from statistikian.com: <https://www.statistikian.com/2014/03/analisis-cluster-nonhirarki-dengan-spss.html>
- [4] [BPS Pusat] Badan Pusat Statistik Pusat. 2019. *Indeks Pembangunan Manusia (IPM) 2019*. Jakarta (ID): Badan Pusat Statistik.
- [5] [BPS Jabar] Badan Pusat Statistik Pusat. 2022. *Indeks Pembangunan Manusia (IPM) Provinsi Jawa Barat 2021*. Bandung (ID): Badan Pusat Statistik.
- [6] Muhajir, S.Si., M.Sc., M. (2021). *Modul Praktikum Statistika Multivariat Terapan*. Yogyakarta: Universitas Islam Indonesia.
- [7] Wu, J., Chang, J., Yang, D. L., & Hung, M. C. (2005). An Efficient k-Means Clustering Algorithm Using Simple Partitioning.
- [8] BBC. (2014, December 30). *AirAsia QZ8501: Does bad weather cause plane crashes?* Diambil kembali dari BBC: <http://www.bbc.com/news/world-30631968>.
- [9] Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2016). *Probability & statistics for engineers & scientists* (Ninth Edition). Pearson Education.
- [10] Anita Fitria, Antonito Hornay, dan Gangga Anuraga. (2018). *K-Means Clustering dengan Metode Elbow untuk Pengelompokan Kabupaten dan Kota di Jawa Timur Berdasarkan Indikator Kemiskinan*.