

PENGGEROMBOLAN KABUPATEN/KOTA DI INDONESIA BERDASARKAN INDIKATOR INDEKS PEMBANGUNAN MANUSIA MENGUNAKAN METODE *K-MEANS* DAN *FUZZY C-MEANS**

Hanniva¹, Anang Kurnia^{2‡}, Septian Rahardiantoro³, Ahmad Ansori
Mattjik⁴

^{1,2,3,4}Department of Statistics, IPB University, Indonesia

[‡]corresponding author: anangk@apps.ipb.ac.id

Copyright © 2022 Hanniva, Anang Kurnia, Septian Rahardiantoro, and Ahmad Ansori Mattjik. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

The achievement of the human development index in Indonesia differs between regions with striking gaps occurring in the western and eastern parts of Indonesia. This difference in achievement can be seen more clearly by grouping regencies/municipalities in Indonesia based on the four indicators of the human development index. With this aim, this study uses the k-means and fuzzy c-means methods to determine the optimal cluster size with two distance approaches, namely the Euclidean and Manhattan distances on the human development index indicators data in 2020. In addition, this study also seeks to identify the distribution of regencies/municipalities based on the characteristics of the human development index indicators in the clustering result. The result is that the best distance measure is Euclidean distance with optimal cluster size is four for k-means and six for fuzzy c-means. In addition, the clustering results obtained by the k-means method are more optimal than the fuzzy c-means because the evaluation value is better. In general, the four clusters formed were in accordance with the grouping carried out by BPS with the percentage of conformity reaching 66,54%. In summary, most regencies/municipalities on the Island of Sumatera, Java, Borneo and Sulawesi have higher life expectancy and percapita expenditure than many regencies/municipalities in the Nusa Tenggara Islands (besides Bali), Moluccas and Papua. Very high achievement for each HDI indicators is dominated by the capital city of each province with unfavorable conditions occurring in most regencies/municipalities in Papua Province.

Keywords: cluster analysis, fuzzy c-means, human development index, k-means.

* Received: Sep 2021; Reviewed: Oct 2021; Published: Jan 2022

1. Pendahuluan

1.1 Latar Belakang

Sumber daya manusia merupakan aset berharga yang dimiliki oleh suatu negara. Peningkatan kualitas sumber daya manusia perlu diperhatikan untuk mencapai kemajuan suatu bangsa. Pemerintah telah melakukan berbagai strategi untuk meningkatkan kualitas sumber daya manusia melalui pembangunan manusia. Capaian pembangunan manusia diukur dengan indeks pembangunan manusia atau IPM yang dibentuk dari tiga dimensi dasar yaitu umur panjang dan hidup sehat, pengetahuan serta standar hidup layak. Indikator yang digunakan dalam mencerminkan dimensi-dimensi tersebut pada perhitungan IPM adalah angka harapan hidup saat lahir, rata-rata lama sekolah, angka harapan lama sekolah dan pengeluaran perkapita disesuaikan. Badan Pusat Statistik mengelompokkan wilayah ke dalam empat kategori berdasarkan capaian IPM. Capaian yang diperoleh menunjukkan terjadi kesenjangan yang mencolok antara wilayah bagian barat dan timur Indonesia (BPS 2015). Dengan demikian, perlu dilakukan identifikasi terhadap persebaran kabupaten/kota di Indonesia berdasarkan capaian dari setiap indikator IPM sehingga dapat diketahui lebih jelas indikator yang menjadi permasalahan pada suatu wilayah. Penelitian ini melakukan penggerombolan kabupaten/kota di Indonesia berdasarkan kedekatan lokasi menggunakan konsep jarak untuk kemudian dapat mengonfirmasi pengelompokan yang dilakukan oleh BPS serta dapat mengidentifikasi lebih jelas mengenai karakteristik dan kesenjangan yang terjadi antar kabupaten/kota di Indonesia pada setiap indikator IPM tersebut.

Analisis gerombol secara umum terdiri dari tiga metode yaitu metode grafik, metode hierarki dan non-hierarki (Mattjik dan Sumertajaya 2011), sedangkan penentuan gerombol terbagi dua yaitu secara tegas dan secara samar (Ross 2010). Metode non-hierarki melakukan penggerombolan dengan menentukan jumlah dan pusat gerombol terlebih dahulu (Mattjik dan Sumertajaya 2011). Penentuan gerombol secara tegas diakomodasi oleh metode *k-means* sedangkan secara samar diakomodasi oleh *fuzzy c-means*. Kedua metode tersebut sering digunakan karena cocok pada data yang berukuran besar yang memiliki peubah kontinu (Lathifaturrahmah 2014). Syafrina (2015) dalam tesisnya juga melakukan penggerombolan daerah di Indonesia berdasarkan peubah IPM tahun 2012 menggunakan *k-medoid* dan *fuzzy c-means*. Penentuan ukuran gerombol didasarkan pada jumlah kategori IPM oleh Badan Pusat Statistik dan memperoleh hasil dengan *fuzzy c-means* dapat membentuk gerombol dengan homogenitas yang tinggi pada setiap gerombol yang terbentuk. Penggerombolan kabupaten/kota di Indonesia berdasarkan indikator IPM yang dilakukan pada penelitian ini adalah menggunakan metode *k-means* dan *fuzzy c-means*. Pada praktiknya proses penggerombolan diterapkan dengan menggunakan dua ukuran jarak yaitu jarak Euclidean dan jarak Manhattan untuk kemudian dapat memilih metode terbaik berdasarkan ukuran jarak terbaik yang digunakan serta ukuran gerombol optimal yang terbentuk. Berdasarkan hasil yang diperoleh kemudian dilakukan penjabaran mengenai karakteristik gerombol yang terbentuk sebagai identifikasi terhadap karakteristik kabupaten/kota di Indonesia berdasarkan capaian indikator IPM pada tahun 2020.

2. Metodologi

2.1 Data

Penelitian ini menggunakan data indikator indeks pembangunan manusia dengan perhitungan metode baru yang terdiri dari 514 kabupaten/kota di Indonesia. Data merupakan data sekunder yang berasal dari publikasi situs web Badan Pusat Statistik pada tahun 2020. Peubah yang digunakan dirincikan pada Tabel 1.

Tabel 1 Indikator indeks pembangunan manusia.

Kode	Peubah	Satuan
AHH	Angka Harapan Hidup saat Lahir	Tahun
RLS	Rata-rata Lama Sekolah	Tahun
HLS	Angka Harapan Lama Sekolah	Tahun
PPP	Pengeluaran per Kapita Disesuaikan	Ribu rupiah/orang/tahun

2.2 Prosedur Analisis Data

Analisis data dilakukan menggunakan perangkat lunak R Studio dengan pemrograman R versi 4.0.2. Prosedur analisis data pada penelitian ini yaitu:

1. Eksplorasi data untuk mengetahui karakteristik data melalui *boxplot*.
2. Praproses Data
 - a. Melakukan standardisasi peubah untuk memastikan peubah dengan nilai yang besar tidak mendominasi peubah lainnya yang bernilai kecil (Han *et al.* 2012).
 - b. Melakukan pemeriksaan korelasi antar peubah menggunakan korelasi Pearson dikarenakan penggunaan jarak Euclidean dan jarak Manhattan digunakan jika antar peubah tidak saling berkorelasi. Apabila terdapat korelasi yang cukup kuat antar peubah maka dapat ditangani dengan mentransformasi peubah asal menggunakan analisis komponen utama (Khairunnisa 2017). Penanganan dengan analisis komponen utama diharapkan dapat memperoleh hasil gerombol yang lebih baik.
3. Melakukan penggerombolan menggunakan metode *k-means* dan *fuzzy c-means* dengan ukuran jarak Euclidean dan jarak Manhattan. Penggerombolan dilakukan dengan menetapkan nilai bibit (*seed*) 1-100 pada perangkat lunak R sebagai set acak terhadap penentuan titik pusat gerombol awal pada metode *k-means* dan penentuan derajat keanggotaan untuk matriks partisi awal pada metode *fuzzy c-means*. Pengulangan sebanyak 100 kali yang dilakukan dapat digunakan untuk memeriksa kestabilan dari pembentukan gerombol (Mattjik dan Sumertajaya 2011). Ukuran jarak terbaik dan ukuran gerombol optimal ditentukan menggunakan dua ukuran validitas yaitu rasio simpangan baku dan indeks Davies-Bouldin.
4. Melakukan visualisasi hasil gerombol dengan diagram pencar komponen utama.
5. Mengevaluasi hasil penggerombolan untuk memilih hasil terbaik dari kedua metode yang digunakan. Evaluasi dilakukan dengan membandingkan nilai dua validitas yang digunakan sebelumnya serta menambahkan validitas lainnya sebagai pertimbangan yaitu koefisien siluet.
6. Melakukan pemetaan persebaran kabupaten/kota berdasarkan hasil penggerombolan yang diperoleh.

3. Hasil dan Pembahasan

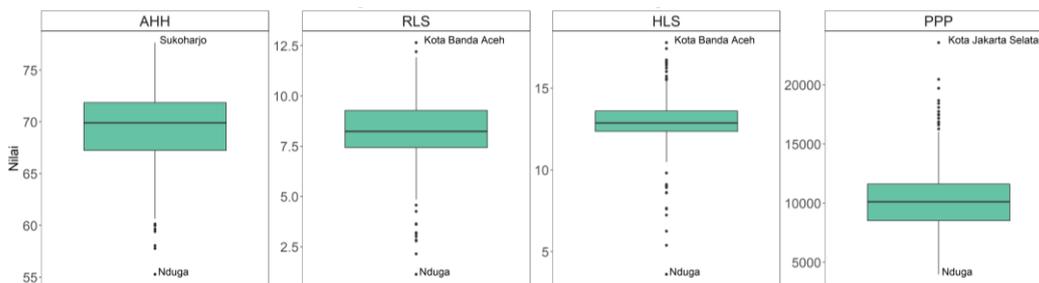
3.1 Eksplorasi Data

Mayoritas kabupaten/kota di Indonesia memiliki capaian IPM berkategori sedang. Beberapa wilayah mengalami kesenjangan pada capaian IPM. Rataan capaian IPM pada kategori rendah memiliki perbedaan nilai yang cukup besar dengan rata-ran IPM pada kategori tinggi yaitu sebesar 55,06.

Tabel 2 Karakteristik wilayah berdasarkan pengelompokan nilai IPM.

Kategori	Rataan nilai IPM	Jumlah Kabupaten/Kota
Rendah	51,10	22
Sedang	66,76	267
Tinggi	73,52	189
Sangat tinggi	81,85	36

Kesenjangan IPM terjadi karena perbedaan capaian indikator IPM. Gambar 1 menunjukkan terdapat pencilan bawah pada setiap indikator selain pengeluaran perkapita dan pencilan atas pada setiap indikator selain angka harapan hidup. Pencilan bawah beberapa indikator berasal dari Indonesia bagian timur. Lebih dari 40% kabupaten/kota dari Provinsi Papua memiliki capaian sangat rendah pada dimensi pengetahuan. Kabupaten/kota dengan capaian setiap indikator terendah adalah Kabupaten Nduga yang berasal dari Provinsi Papua.



Gambar 1 Diagram kotak garis indikator IPM tahun 2020.

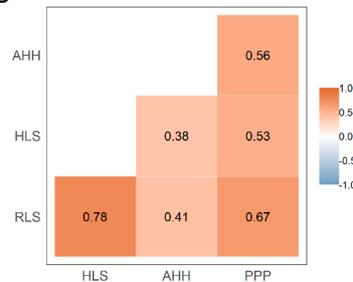
3.2 Praproses Data

Tahapan praproses data melakukan standardisasi peubah dan pemeriksaan korelasi antar peubah. Tabel 3 menunjukkan indikator AHH dan PPP memiliki nilai dan keragaman yang lebih besar. Setelah dilakukan standardisasi peubah telah memiliki rentang dan keragaman yang lebih homogen.

Tabel 3 Statistika deskriptif indikator indeks pembangunan manusia.

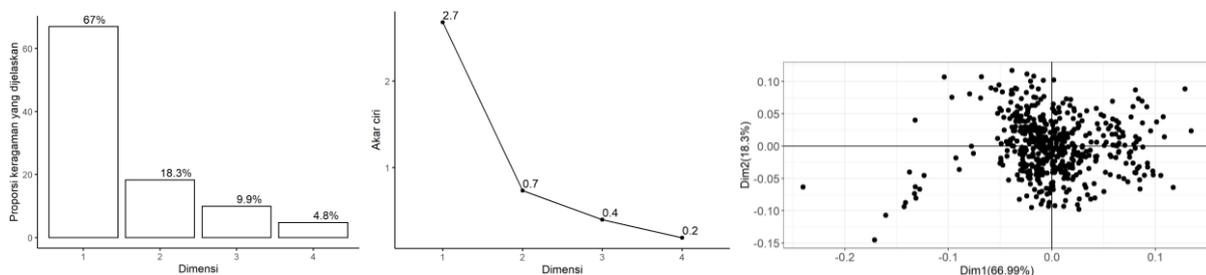
Peubah	Sebelum distandardisasi			Setelah distandardisasi		
	Minimum	Maksimum	Simpangan Baku	Minimum	Maksimum	Simpangan Baku
AHH	55,27	77,65	3,46	-4,13	2,34	1,00
RLS	1,13	12,65	1,63	-4,42	2,64	1,00
HLS	3,61	17,79	1,33	-7,02	3,63	1,00
PPP	3.975,00	23.575,00	2.694,15	-2,33	4,94	1,00

Pemeriksaan nilai korelasi dilakukan untuk melihat hubungan linier antar peubah. Gambar 2 menunjukkan hubungan paling kuat terjadi antara indikator rata-rata lama sekolah dan harapan lama sekolah dengan nilai yang cukup besar yaitu sebesar 0,78. Untuk memperoleh hasil penggerombolan yang lebih baik, maka dilakukan analisis komponen utama (AKU) sehingga diperoleh peubah baru yang tidak saling berkorelasi. AKU dilakukan pada data yang sudah distandardisasi sebelumnya.



Gambar 2 Koefisien korelasi pearson indikator IPM.

Berdasarkan Gambar 3, dua komponen utama menggambarkan peubah asal dengan keragaman kumulatif sebesar 85,3%. Selain itu, mengacu pada pendapat Jolliffe (2002) yang mengatakan bahwa batas nilai akar ciri yang baik adalah sebesar 0,7. Penelitian ini akan dilakukan menggunakan dua komponen utama pada tahapan analisis gerombol selanjutnya. Diagram pencar dua komponen utama 3(c) memperlihatkan kabupaten/kota memencar secara tidak beraturan, mayoritas berada pada sisi tengah namun terdapat beberapa yang memiliki nilai cukup berbeda dibanding mayoritas yang lain yaitu kabupaten/kota yang berada pada sisi kiri bawah diagram.



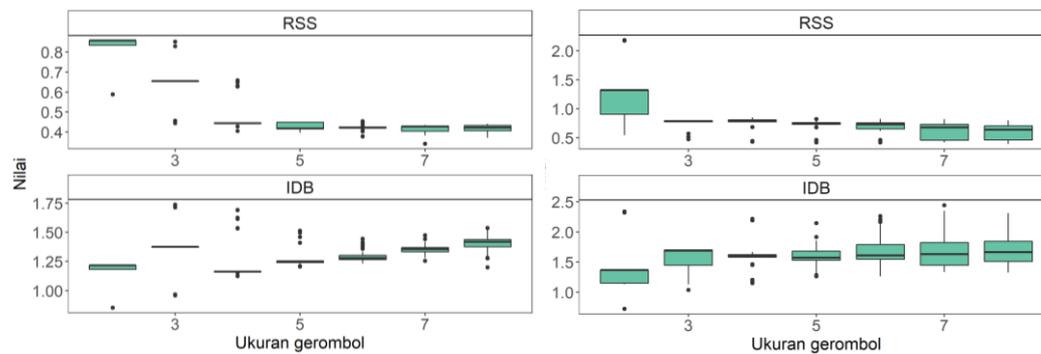
Gambar 3 (a) Proporsi keragaman, (b) Akar ciri, (c) Diagram komponen utama.

3.3 Analisis Gerombol

3.3.1 Penggerombolan Metode *K-Means*

Penggerombolan dilakukan menggunakan dua komponen utama dan dinilai kebaikannya menggunakan dua validitas yaitu indeks Davies-Bouldin dan rasio simpangan baku. Indeks Davies-Bouldin (IDB) mengukur kualitas gerombol yang didasarkan pada nilai kohesi dan separasi (Muhammad 2015), sedangkan rasio simpangan baku (RSS) melalui perbandingan keragaman di dalam dan antar gerombol (Johnson dan Wichern 2007). Gambar 4 menampilkan diagram kotak kedua validitas dari 100 kali penggerombolan menggunakan jarak Euclidean dan jarak Manhattan

untuk ukuran gerombol dua hingga delapan. Penggerombolan terbaik diperoleh saat nilai kedua validitas tersebut bernilai kecil. Secara umum penggunaan jarak Euclidean menghasilkan rentang nilai rasio simpangan baku dan indeks Davies-Bouldin yang lebih rendah sehingga penggunaan jarak Euclidean lebih baik dibanding jarak Manhattan. Pada penggunaan jarak Euclidean tersebut, ukuran gerombol optimal yang direkomendasikan adalah sebesar empat karena memiliki nilai kedua validitas yang cukup rendah.



Gambar 4 Sebaran indeks validitas metode *k-means*: (a) menggunakan jarak Euclidean, (b) menggunakan jarak Manhattan.

Hasil yang dipilih dari 100 kali penggerombolan pada ukuran gerombol empat adalah yang memiliki nilai kedua validitas yang rendah. Karakteristik penggerombolan yang dipilih ditampilkan pada Tabel 4. Berdasarkan rata-rata nilai setiap indikator, secara berurutan gerombol 4, 3, 1, dan 2 dikategorikan sebagai kabupaten/kota dengan capaian indikator IPM rendah, sedang, tinggi dan sangat tinggi. Gerombol dengan kategori sedang dan tinggi lebih dicirikan oleh perbedaan pada angka harapan hidup dan pengeluaran perkapita karena kedua gerombol tersebut memiliki nilai yang tidak berbeda jauh untuk indikator rata-rata lama sekolah dan harapan lama sekolah.

Tabel 4: Rataan gerombol metode *k-means* ($c = 4$).

Gerombol	AHH	RLS	HLS	PPP	Kategori
G1	71,07	8,16	12,76	10.490,00	Tinggi
G2	72,84	10,79	14,63	14.184,00	Sangat tinggi
G3	66,24	7,82	12,82	8.453,00	Sedang
G4	64,17	3,36	8,07	5.042,00	Rendah

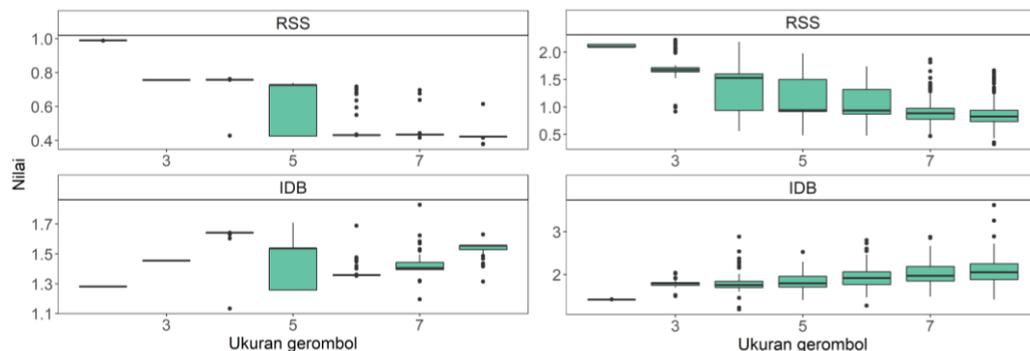
Tabel 5 menunjukkan anggota terbanyak adalah gerombol yang memiliki capaian indikator IPM yang tinggi lalu diikuti gerombol dengan capaian indikator IPM sedang. Secara umum pencapaian kabupaten/kota berdasarkan indikator IPM cukup baik karena hanya sedikit yang tergolong pada kategori indikator IPM yang rendah. Gerombol yang dihasilkan berisikan mayoritas kabupaten/kota yang sesuai dengan pengelompokan yang dilakukan BPS dengan total kesesuaian sebesar 66,54%.

Tabel 5: Distribusi gerombol berdasarkan pengelompokan IPM.

Kategori indikator IPM	Kategori IPM				Total	Kesesuaian gerombol	Persentase kesesuaian
	Rendah	Sedang	Tinggi	Sangat tinggi			
Rendah	14	0	0	0	14	14/22	63,64%
Sedang	8	158	7	0	173	158/267	59,18%
Tinggi	0	109	134	0	243	134/189	70,89%
Sangat tinggi	0	0	48	36	84	36/36	100,00%
Total	22	267	189	36	514	342/514	66,54%

3.3.2 Penggerombolan Metode *Fuzzy C-Means*

Penggerombolan pada metode *fuzzy c-means* juga dilakukan menggunakan dua komponen utama. Gambar 5 menampilkan diagram kotak validitas indeks Davies-Bouldin dan rasio simpangan baku yang diperoleh dari 100 kali penggerombolan menggunakan jarak Euclidean dan Manhattan untuk ukuran gerombol dua hingga delapan. Secara umum penggunaan jarak Euclidean menghasilkan rentang nilai kedua validitas yang lebih rendah sehingga penggunaan jarak Euclidean lebih baik dibanding jarak Manhattan. Pada penggunaan jarak Euclidean tersebut, ukuran gerombol optimal yang dipilih adalah sebesar enam karena memiliki nilai kedua validitas yang cukup rendah.



Gambar 5 Sebaran indeks validitas metode *fuzzy c-means*: (a) menggunakan jarak Euclidean, (b) menggunakan jarak Manhattan.

Tabel 6 Rataan gerombol metode *fuzzy c-means* ($c = 6$).

Gerombol	AHH	RLS	HLS	PPP	Kategori
G1	68,33	8,35	13,10	9.729,00	RLS HLS sedang
G2	71,78	8,97	13,24	11.420,00	Tinggi
G3	72,92	10,90	14,74	14.621,00	Sangat tinggi
G4	64,17	3,36	8,07	5.042,00	Rendah
G5	71,09	7,30	12,20	9.619,00	AHH sedang
G6	65,06	7,67	12,74	7.885,00	Sedang

Hasil penggerombolan yang dipilih dari ukuran gerombol enam adalah yang memiliki nilai kedua validitas yang rendah dari 100 kali penggerombolan.

Penggerombolan yang dipilih memiliki karakteristik yang ditampilkan pada Tabel 6. Gerombol 3 memiliki rataan setiap indikator IPM tertinggi sedangkan gerombol 4 dengan rataan setiap indikator IPM terendah. Gerombol 1 dicirikan dengan rataan indikator rata-rata lama sekolah dan harapan lama sekolah ketiga tertinggi sedangkan gerombol 5 dengan rataan angka harapan hidup ketiga tertinggi.

Tabel 7 menunjukkan banyak anggota pada setiap gerombol hampir tersebar merata kecuali pada gerombol dengan kategori indikator IPM rendah dan sangat tinggi. Hasil penggerombolan yang diperoleh memperjelas karakteristik pada kategori sedang dan tinggi diantaranya dengan capaian indikator IPM sedang, angka harapan hidup sedang, capaian dimensi pengetahuan (rata-rata lama sekolah dan harapan lama sekolah) sedang dan capaian indikator IPM tinggi. Mayoritas gerombol yang dihasilkan sesuai dengan pengelompokan yang dilakukan oleh BPS dengan total kesesuaian sebesar 81,91%.

Tabel 7 Distribusi gerombol berdasarkan pengelompokan IPM.

Kategori indikator IPM	Kategori IPM				Total	Kesesuaian gerombol	Persentase kesesuaian
	Rendah	Sedang	Tinggi	Sangat tinggi			
Rendah	14	0	0	0	14	14/22	63,64%
Sedang	8	90	0	0	98	90/267	33,71%
AHH sedang	0	93	19	0	112	93/267	34,83%
RLS HLS sedang	0	83	29	0	112	83/267	31,09%
Tinggi	0	1	105	0	106	105/189	55,56%
Sangat tinggi	0	0	36	36	72	36/36	100,00%
Total	22	267	189	36	514	421/514	81,91%

Tabel 8 Rataan gerombol metode *fuzzy c-means* ($c = 4$).

Gerombol	AHH	RLS	HLS	PPP	Kategori
G1	71,16	8,13	12,74	10.481,00	Tinggi
G2	72,76	10,72	14,58	14.120,00	Sangat tinggi
G3	64,17	3,36	8,07	5.042,00	Rendah
G4	66,35	7,84	12,82	8.504,00	Sedang

BPS mengelompokkan kabupaten/kota di Indonesia ke dalam empat kategori berdasarkan selang IPM. Selain itu, penggerombolan metode *k-means* memberikan rekomendasi ukuran gerombol yang optimal adalah sebesar empat. Dengan demikian, pada *fuzzy c-means* juga dilakukan penggerombolan dengan ukuran gerombol empat untuk dapat dijadikan sebagai perbandingan. Hasil yang dipilih dari 100 kali penggerombolan pada ukuran gerombol empat adalah yang memiliki nilai kedua validitas yang rendah dan hasil yang terpilih tersebut memiliki karakteristik yang ditampilkan pada Tabel 8. Secara berurutan gerombol 3, 4, 1 dan 2 memiliki capaian indikator IPM dari terendah hingga tertinggi. Gerombol dengan indikator IPM sedang dan tinggi juga lebih dicirikan oleh perbedaan pada angka harapan hidup dan

pengeluaran perkapita karena kedua gerombol tersebut memiliki capaian untuk rata-rata lama sekolah dan harapan lama sekolah yang tidak berbeda jauh.

Tabel 9 menunjukkan anggota terbanyak adalah gerombol dengan capaian indikator IPM tinggi sedangkan yang paling sedikit memiliki capaian indikator rendah. Penggerombolan tersebut memberikan distribusi wilayah yang tidak berbeda jauh dengan hasil penggerombolan metode *k-means*. Hasil gerombol yang diperoleh mayoritas juga sesuai dengan pengelompokan yang dilakukan oleh BPS dengan total kesesuaian sebesar 66,54%.

Tabel 9 Distribusi gerombol berdasarkan pengelompokan IPM.

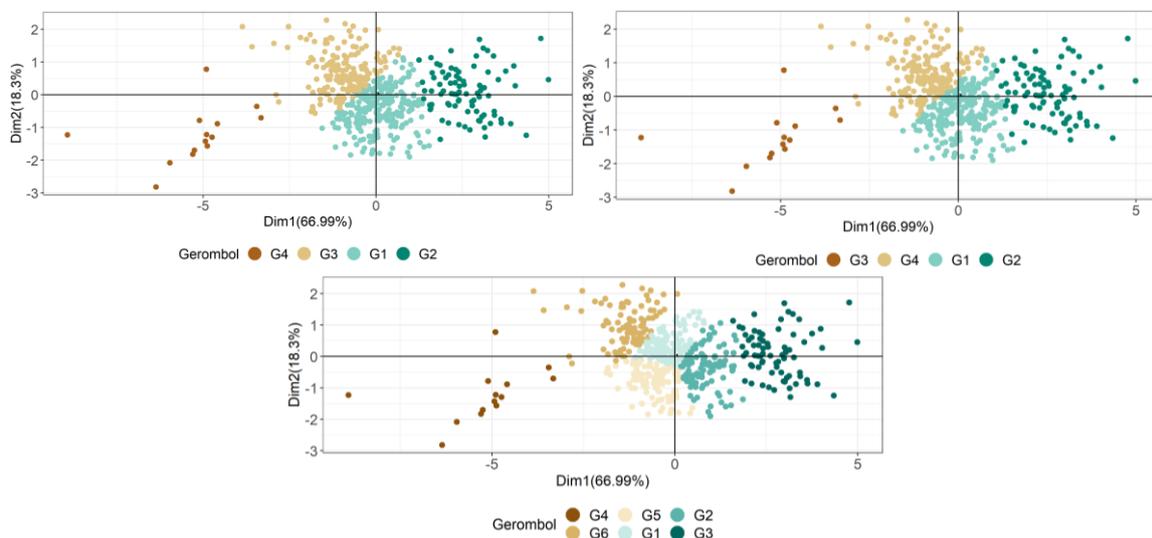
Kategori indikator IPM	Kategori IPM				Total	Kesesuaian gerombol	Persentase kesesuaian
	Rendah	Sedang	Tinggi	Sangat tinggi			
Rendah	14	0	0	0	14	14/22	63,64%
Sedang	8	164	9	0	181	164/267	61,42%
Tinggi	0	103	128	0	231	128/189	67,73%
Sangat tinggi	0	0	52	36	88	36/36	100,00%
Total	22	267	189	36	514	342/514	66,54%

Hasil penggerombolan yang diperoleh dari kedua metode menunjukkan mayoritas capaian kabupaten/kota berdasarkan indikator IPM sesuai dengan pengelompokan IPM yang dilakukan oleh BPS. Hal itu dikarenakan perhitungan IPM oleh BPS menggunakan formula rata-rata geometrik sehingga dapat menyeimbangkan perhitungan nilai IPM terhadap capaian dari setiap indikator. Nilai dari satu indikator tidak dapat menutupi kekurangan dari nilai indikator lainnya dalam menghitung IPM (BPS 2015). Oleh sebab itu, penggerombolan yang bekerja menggunakan konsep jarak memiliki kemungkinan menghasilkan gerombol yang secara mayoritas mirip dengan pengelompokan yang dilakukan oleh BPS berdasarkan selang nilai IPM. Namun demikian, terdapat beberapa pergeseran pada banyak kabupaten/kota yang sebelumnya dikategorikan sebagai IPM yang sedang kemudian digerombolkan sebagai indikator IPM yang tinggi begitupun untuk beberapa kategori lainnya. Kabupaten/kota yang mengalami pergeseran dari kategori sedang ke tinggi tersebut merupakan kabupaten/kota yang memiliki capaian angka harapan hidup dan capaian pada dimensi pengetahuan yang lebih tinggi dibanding kabupaten/kota lainnya pada kategori IPM sedang, hal itu terlihat berdasarkan hasil penggerombolan yang diperoleh dari penggerombolan ukuran gerombol enam.

3.4 Diagram Pencar Komponen Utama Hasil Penggerombolan

Diagram pencar komponen utama pada Gambar 6 memperlihatkan penggerombol dengan ukuran gerombol empat memberikan hasil penggerombolan yang hampir sama untuk *k-means* dan *fuzzy c-means*. Kabupaten/kota dengan capaian indikator IPM yang rendah berada pada sisi kiri bawah diagram dengan wilayah merupakan kabupaten/kota yang berasal dari Provinsi Papua. Wilayah pada sisi tengah atas dan bawah merupakan kabupaten/kota dengan indikator IPM sedang dan tinggi sedangkan

pada sisi paling kanan merupakan kabupaten/kota dengan indikator IPM sangat tinggi. Hasil *fuzzy c-means* dengan ukuran gerombol enam mampu mengurai karakteristik wilayah pada kategori sedang dan tinggi, sementara pada kategori rendah dan sangat tinggi menghasilkan gerombol dengan wilayah yang hampir sama.



Gambar 6 Diagram pencar komponen utama (a) Metode *k-means* ($c = 4$), (b) Metode *fuzzy c-means* ($c = 4$), (c) Metode *fuzzy c-means* ($c = 6$).

3.5 Evaluasi Hasil Penggerombolan

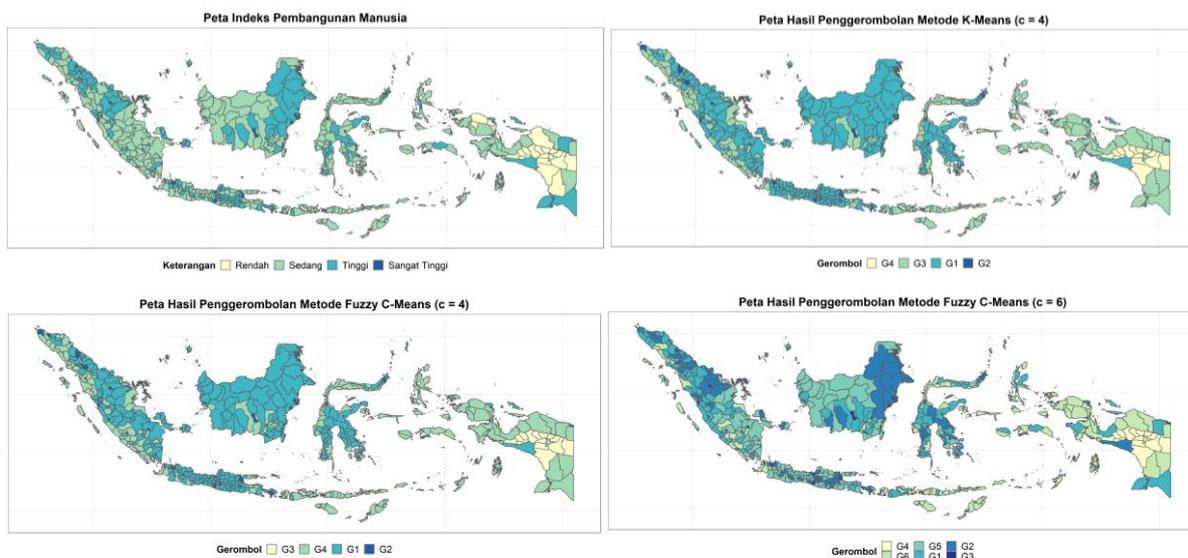
Tabel 10 menunjukkan penggerombolan yang dihasilkan menggunakan dua komponen utama membentuk gerombol yang cukup optimal karena memiliki nilai validitas rasio simpangan baku dan indeks Davies-Bouldin yang lebih rendah serta nilai koefisien siluet yang lebih tinggi jika dibandingkan dengan penggerombolan yang dilakukan untuk inisiasi yang sama menggunakan peubah asal, tiga maupun empat komponen utama. Selain itu, penggerombolan *k-means* juga menghasilkan gerombol yang lebih optimal dibanding penggerombolan *fuzzy c-means* karena memiliki nilai rasio simpangan baku dan indeks Davies-Bouldin yang lebih rendah serta nilai koefisien siluet yang lebih tinggi.

Tabel 10 Perbandingan validitas hasil penggerombolan.

Metode	c	Peubah	RSS	Koefisien siluet	IDB	Jumlah iterasi	Kesesuaian gerombol
<i>K-means</i>	4	Peubah asal	0,4059	0,2885	1,1316	9	337/514
	4	2 KU	0,4047	0,2898	1,1268	12	342/514
	4	3 KU	0,4065	0,2887	1,1313	8	335/514
	4	4 KU	0,4059	0,2885	1,1316	9	337/514
<i>Fuzzy c-means</i>	6	Peubah asal	0,4458	0,1884	1,4192	31	427/514
	6	2 KU	0,4313	0,2107	1,3580	51	421/514
	6	3 KU	0,4409	0,1974	1,3906	38	425/514
	6	4 KU	0,4458	0,1884	1,4192	31	427/514

3.6 Pemetaan Hasil Penggerombolan

Gambar 7 menampilkan peta persebaran kabupaten/kota di Indonesia berdasarkan pengelompokan selang capaian IPM dan penggerombolan berdasarkan indikator IPM. Berdasarkan capaian IPM mayoritas kabupaten/kota di Pulau Sumatera, Kalimantan, Sulawesi, Nusa Tenggara dan Maluku mayoritas memiliki capaian IPM pada kategori sedang sedangkan di Pulau Jawa mayoritas dengan capaian pada kategori tinggi. Penggerombolan terbaik berdasarkan evaluasi yang dilakukan sebelumnya adalah penggerombolan metode *k-means* dengan ukuran gerombol empat. Hasil yang diperoleh menunjukkan mayoritas kabupaten/kota di Pulau Sumatera, Jawa, Kalimantan, dan Sulawesi memiliki capaian indikator IPM yang tinggi sedangkan di Kepulauan Nusa Tenggara (kecuali kabupaten/kota di Provinsi Bali), Maluku dan Papua mayoritas dengan capaian indikator IPM sedang. Kedua kategori ini lebih dicirikan oleh perbedaan capaian angka harapan hidup dan pengeluaran perkapita sehingga gerombol yang dihasilkan menunjukkan sebagian besar kabupaten/kota di Pulau Sumatera, Jawa, Kalimantan dan Sulawesi memiliki capaian angka harapan hidup dan pengeluaran perkapita yang lebih tinggi dibanding banyak kabupaten/kota di Kepulauan Nusa Tenggara (kecuali di Bali), Maluku dan Papua. Selain itu, hasil yang diperoleh dari setiap penggerombolan menunjukkan kondisi kurang baik terjadi di Provinsi Papua karena sebanyak 14 dari 29 kabupaten/kota digerombolkan dengan indikator IPM yang rendah sedangkan untuk kondisi yang sangat baik pada setiap indikator IPM didominasi oleh kabupaten/kota yang menjadi lokasi keberadaan ibu kota dari setiap provinsi di Indonesia.



Gambar 7 Peta persebaran wilayah berdasarkan IPM dan hasil gerombol.

4. Simpulan

Penggerombolan kabupaten/kota di Indonesia berdasarkan indikator IPM menunjukkan ukuran jarak terbaik adalah jarak Euclidean. Ukuran gerombol optimal yang dihasilkan adalah sebesar empat untuk metode *k-means* dan sebesar enam untuk *fuzzy c-means*. Gerombol terbaik adalah hasil yang diperoleh pada metode *k-means* dengan mengelompokkan kabupaten/kota ke dalam empat kategori yaitu dengan capaian indikator IPM rendah, sedang, tinggi dan sangat tinggi. Penggerombolan yang dilakukan menggunakan konsep jarak tersebut menghasilkan

pengelompokan yang secara mayoritas sesuai dengan pengelompokan IPM yang dihasilkan oleh BPS. Hasil penggerombolan yang diperoleh menunjukkan sebagian besar kabupaten/kota di Pulau Jawa, Sumatera, Kalimantan dan Sulawesi memiliki capaian angka harapan hidup dan pengeluaran perkapita yang lebih tinggi dibandingkan beberapa kabupaten/kota di Kepulauan Nusa Tenggara (kecuali kabupaten/kota di Provinsi Bali), Maluku, dan Papua. Capaian yang sangat tinggi pada indikator IPM didominasi oleh kabupaten/kota yang menjadi lokasi keberadaan ibu kota dari setiap provinsi di Indonesia. Lebih lanjut, hasil penggerombolan juga menunjukkan terdapat banyak kabupaten/kota di Provinsi Papua yang masih belum memenuhi kondisi yang diharapkan pada setiap dimensi IPM.

Daftar Pustaka

- [BPS] Badan Pusat Statistik. (2015). *Indeks Pembangunan Manusia 2014 Metode Baru*, Jakarta : Badan Pusat Statistik.
- Han, J., Kamber, M., dan Pei, J. 2012. *Data Mining Concepts and Techniques. Third Edition*, Waltham (MA): Morgan Kaufmann.
- Jolliffe, I. T. (2002). *Principal Component Analysis*. Second Edition, New York (US): Springer-Verlag.
- Khairunnisa. (2017). Kajian Perbandingan Metode K-Means dan Fuzzy C-Means untuk Menggerombolan Kecamatan di Provinsi Jawa Barat Berdasarkan Data Kemiskinan [tesis], Bogor: Institut Pertanian Bogor.
- Lathifaturrahman. (2014). Perbandingan Hasil Penggerombolan K-Means, Fuzzy C-Means, dan Two Step Clustering, *JPM IAIN Antasari* 2(1): 39-62.
- Mattjik, A. A. dan Sumertajaya I. M. 2011. *Sidik Peubah Ganda dengan Menggunakan SAS*, Bogor: IPB Press.
- Muhammad, A. F. (2015). Klasterisasi proses seleksi pemain menggunakan algoritma k-means (study kasus: tim hockey Kabupaten Kendal). *Jurnal Teknik Informatika FIK UDINUS*. 1-5.
- Syafrina. (2015). Penggerombolan Daerah di Indonesia Berdasarkan Peubah IPM dengan Fuzzy k-rataan dan k-medoid [tesis]. Bogor: Institut Pertanian Bogor.
- Ross, T. J. (2010). *Fuzzy Logic with Engineering Applications*. Third Edition, New York (US): John Wiley and Sons.
- Johnson, R. A. and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. Third Edition, New Jersey (US): Pearson Education.