

## Metode *C-Means* untuk Pengelompokan Kabupaten/Kota Provinsi Jawa Timur berdasarkan Indikator Indeks Pembangunan Manusia (IPM)

Marita Qori'atunnadyah

Program Studi Informatika, Institut Teknologi dan Bisnis Widya Gama Lumajang

Corresponding Author: Marita Qori'atunnadyah (maritaqori@gmail.com)

### ARTICLE INFO

Date of entry:

02 April 2023

Revision Date:

27 April 2023

Date Received:

30 April 2023

### ABSTRAK

Peringkat keempat untuk jumlah penduduk terbesar di dunia adalah Negara Indonesia. Jumlah penduduk yang besar ini dapat dijadikan Negara Indonesia sebagai aset untuk meningkatkan kesejahteraan masyarakatnya. Indeks Pembangunan Manusia (IPM) adalah indeks yang mengukur keberhasilan pembangunan kualitas hidup manusia. Pada tahun 2022, nilai IPM Indonesia sebesar 72,91 dimana nilai ini berada dibawah target pemerintah yang memiliki target sebesar 73,41 hingga 73,46. Begitu juga dengan IPM Provinsi Jawa Timur yang nilainya masih berada dibawah nilai IPM Indonesia dan belum mencapai target. Penelitian ini difokuskan pada pengelompokan wilayah Provinsi Jawa Timur dengan metode *c-means* yang didasarkan pada indikator Indeks Pembangunan Manusia. Berdasarkan hasil penelitian, dengan metode *c-means* didapatkan bahwa jumlah kelompok optimum yang terbentuk adalah sebanyak 4. Masing-masing memiliki kriteria tertentu yaitu kelompok 1 merupakan kelompok dengan Indeks Pembangunan Manusia rendah dan perlu mendapatkan perhatian khusus dari pemerintah. Kemudian kategori cukup rendah berada di Kelompok 3. Selanjutnya, kelompok 4 merupakan kelompok dengan Indeks Pembangunan Manusia tinggi dan kelompok 2 dengan Indeks Pembangunan Manusia cukup tinggi.

Keywords: *C-Means*, Indikator, IPM, Pengelompokan



Cite this as: Qori'atunnadyah, M. (2023). Metode *C-Means* untuk Pengelompokan Kabupaten/Kota Provinsi Jawa Timur berdasarkan Indikator Indeks Pembangunan Manusia (IPM). *Journal of Informatics Development*, 2(2), 51-58. <https://doi.org/10.30741/jid.v2i2.1013>

### PENDAHULUAN

Negara peringkat keempat dengan jumlah penduduk terbesar di dunia setelah negara China, India dan Amerika Serikat adalah Indonesia. Hasil sensus penduduk 2020 menunjukkan bahwa jumlah penduduk Indonesia mencapai angka 270.203.917 jiwa. Jumlah penduduk tahun 2020 ini meningkat sebesar 32,56 juta jiwa dibandingkan jumlah penduduk tahun 2010 yang mencapai angka 237.641.326 jiwa. Kemudian untuk rata-rata laju pertumbuhan penduduk selama kurun waktu tahun 2010 hingga 2020 adalah sebesar 1,25% per tahun (BPS, 2021). Jumlah penduduk Indonesia yang besar dapat dijadikan aset untuk meningkatkan produktivitas masyarakatnya. Selanjutnya, dengan peningkatan produktivitas masyarakat maka kesejahteraan dan pertumbuhan ekonomi juga dapat meningkat jika didukung oleh kualitas Sumber Daya Manusia yang baik.

Namun, jika pertumbuhan penduduk tinggi maka juga dapat menimbulkan masalah apabila SDM yang ada tidak dimanfaatkan dengan baik, seperti tingginya angka pengangguran yang tidak diimbangi dengan tingginya jumlah lapangan pekerjaan yang tersedia. Jadi, pembangunan manusia guna memajukan kualitas Sumber Daya Manusia perlu dilakukan agar pertumbuhan ekonomi dan kesejahteraan masyarakat meningkat.

Indikator yang digunakan sebagai ukuran keberhasilan pembangunan kualitas hidup manusia (masyarakat/penduduk) disebut Indeks Pembangunan Manusia (IPM). Indeks tersebut dibentuk melalui tiga dimensi dasar yaitu standar hidup yang layak, pengetahuan, serta umur panjang dan hidup sehat. Indikator standar hidup yang layak digambarkan oleh pengeluaran perkapita disesuaikan, indikator umur panjang dan hidup sehat digambarkan oleh angka harapan hidup, serta indikator pengetahuan digambarkan oleh harapan lama sekolah dan rata-rata lama sekolah (BPS, 2019). IPM Indonesia pada tahun 2022 sebesar 72,91, dimana nilai tersebut belum mencapai target pemerintah yaitu Indeks Pembangunan Manusia 2022 sebesar 73,41 hingga 73,46 (Kementerian PPN/Bappenas, 2021). Dari 38 provinsi di Indonesia tercatat bahwa IPM Provinsi Jawa Timur masih di bawah IPM Indonesia dan belum memenuhi target IPM 2022 yaitu sebesar 72,75 (BPS, 2023). Sehingga perlu dilakukan pengelompokan wilayah Provinsi Jawa Timur berdasarkan indikator IPM agar Pemerintah dapat merencanakan kebijakan yang tepat untuk meningkatkan IPM Provinsi Jawa Timur.

Pengelompokan objek-objek yang didasarkan pada kemiripan karakteristik yang dimilikinya dapat dilakukan menggunakan analisis *cluster*. Pada pengelompokan ini, tingkat kemiripan objek-objek dalam satu kelompok sangat tinggi sedangkan tingkat kemiripan objek-objek antar kelompok satu dengan lainnya rendah (Hair, Black, Babin, & Anderson, 2010). *C-means* adalah metode pengelompokan yang termasuk dalam *non-hierarki* dimana dilakukan dengan membagi data ke dalam satu atau lebih kelompok. Beberapa penelitian sebelumnya yang menggunakan metode *C-means* yaitu pengelompokan wilayah berdasarkan kondisi infrastruktur jalan (Qori'atunnadyah & Rahmawati, 2022) dan pengelompokan wilayah di Kabupaten Lumajang berdasarkan rasio antara guru dan murid (Qori'atunnadyah, 2022). Pada kedua penelitian tersebut belum dilakukan penentuan *cluster* optimum karena hanya menggunakan satu *cluster* dan juga belum dilakukan pengujian terhadap hasil pengelompokan yang didapatkan. Penelitian ini menggunakan metode *C-means*, dimana pada penelitian ini dilakukan penentuan kelompok optimum dan pengujian terhadap hasil pengelompokan yang didapatkan. Selain itu, penelitian ini berfokus pada pengelompokan wilayah Provinsi Jawa Timur berdasarkan indikator IPM pada tahun 2022.

## METODE

Data penelitian bersumber dari Badan Pusat Statistik Provinsi Jawa Timur (BPS Jawa Timur, 2023). Metode yang digunakan adalah *clustering* yang dapat mengelompokkan objek-objek berdasarkan kemiripannya (Hair, Black, Babin, & Anderson, 2010). *C-means* merupakan pengelompokan *non-hierarki* yang sering digunakan (Johnson & Wicherin, 2007). Berikut ini merupakan algoritma atau langkah dari metode *c-means*.

1. Menentukan jumlah kelompok yang dibentuk (besarnya  $c$ )
2. Membagi data secara acak kedalam  $c$  kelompok
3. Menghitung nilai *centroid* rata-rata dari masing-masing kelompok, sesuai rumus yang diberikan pada persamaan (1).

$$v_{cj} = \frac{\sum_{i=1}^{n_c} x_{ij}}{n_c} \quad (1)$$

Keterangan :

$v_{cj}$  : *centroid* kelompok ke- $c$  untuk variabel ke- $j$

$n_c$  : banyak data kelompok ke- $c$

$x_{ij}$  : data ke- $i$  pada variabel ke- $j$

$i$  : indeks objek

$c$  : indeks kelompok

$j$  : indeks variabel

4. Mengelompokkan data ke *centroid* terdekat (menggunakan jarak *euclidean*) dengan rumus :

$$d_{ci} = \sqrt{\sum_{j=1}^I (x_{ij} - v_{cj})^2} \quad (2)$$

5. Mengulangi langkah 3 hingga tidak terjadi perpindahan data dari satu kelompok ke kelompok yang lain (Agusta, 2007).

Nilai *Pseudo F-statistic* digunakan dalam penentuan banyaknya kelompok optimum yang terbentuk dalam pengelompokkan. Rumus *pseudo F-statistic* adalah sebagai berikut (Orpin & Kostylev, 2006).

$$PseudoF - statistic = \frac{\left(\frac{R^2}{c-1}\right)}{\left(\frac{1-R^2}{n-c}\right)} \quad (3)$$

Pengujian normal multivariat menggunakan uji koefisien korelasi memiliki statistik sebagai berikut.

$$r_Q = \frac{\sum_{j=1}^n (x_{(j)} - \bar{x})(q_{(j)} - \bar{q})}{\sqrt{\sum_{j=1}^n (x_{(j)} - \bar{x})^2} \sqrt{\sum_{j=1}^n (q_{(j)} - \bar{q})^2}} \quad (4)$$

Pengujian homogenitas menggunakan Uji Box's M dilakukan dengan tujuan menguji apakah matriks varians kovarians homogen atau tidak. Statistik yang digunakan ditunjukkan pada persamaan (5).

$$\chi_{hitung}^2 = (1-u) \left\{ \left[ \sum_{c=1}^g (n_c - 1) \right] \ln |S_{pooled}| - \sum_{c=1}^g [(n_c - 1) \ln |S_c|] \right\} \quad (5)$$

*One-Way* MANOVA digunakan untuk membandingkan rata-rata dua populasi atau lebih dengan variabel dependen lebih dari satu atau untuk mengkaji pengaruh dari suatu perlakuan terhadap respon (Johnson & Wicherin, 2007). Jika dalam uji normal multivariat dan uji homogenitas terpenuhi maka *One-Way* MANOVA yang digunakan adalah *Wilk's Lambda*. Namun, jika uji normal multivariat dan uji homogenitas tidak terpenuhi maka digunakan satatistik uji *Pillai's Trace*. Selanjutnya, *One-way* ANOVA digunakan untuk uji perbedaan kelompok ketika variabel terikat yang digunakan hanya satu atau uji perbedaan pada variabel-variabel antar anggota kelompok (Johnson & Wicherin, 2007).

## HASIL DAN PEMBAHASAN

Indikator IPM Provinsi Jawa Timur pada tahun 2022 digunakan sebagai dasar pengelompokkan wilayah pada penelitian ini dengan metode *c-means*. Jumlah wilayah di Jawa Timur adalah 38

kabupaten/kota yang kemudian akan dikelompokkan ke dalam beberapa kelompok. Jumlah kelompok yang dipilih pada penelitian ini adalah 2 sampai 5 kelompok dengan hasil seperti pada Tabel 1.

**Tabel 1. Pengelompokan dengan Metode C-Means**

Kelompok ke-	Jumlah Kelompok			
	2	3	4	5
1	21	9	9	9
2	17	10	13	12
3		19	9	6
4			7	7
5				4

Sumber: Data diolah (2023)

Tabel 1 memberikan informasi mengenai jumlah anggota tiap kelompok dari hasil pengelompokan. Penentuan jumlah kelompok yang optimum dilakukan dengan melihat nilai *pseudo f-statistic* dimana nilai paling tinggi merupakan nilai yang terpilih. Tabel 2 di bawah ini menunjukkan nilai tersebut.

**Tabel 2. Nilai Pseudo F-Statistic**

Jumlah Kelompok	Pseudo F-Statistic
2	38,90176
3	37,90929
4	39,19128
5	31,26145

Sumber: Data diolah (2023)

Berdasarkan Tabel 2 diatas, jumlah kelompok optimum yang terpilih adalah 4 kelompok. Hal ini sesuai dengan nilai *pseudo f-statistic* pada jumlah kelompok 4 yang merupakan nilai tertinggi yaitu sebesar 39,19128. Hasil pengelompokan menggunakan 4 kelompok disajikan pada Tabel 3.

**Tabel 3. Daftar Kabupaten/Kota pada 4 Kelompok**

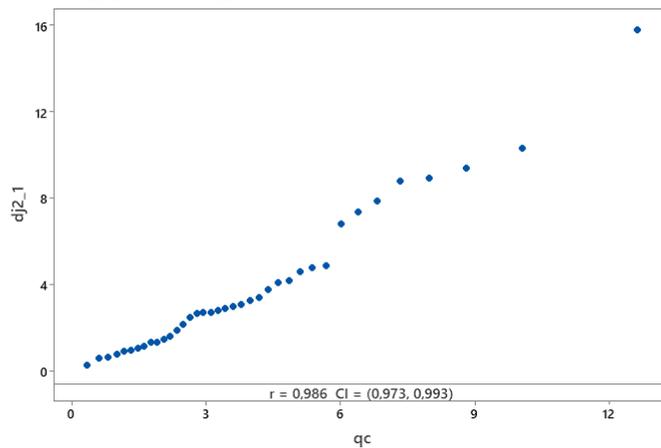
Kelompok	Kabupaten/Kota		
1	Lumajang	Pamekasan	Situbondo
	Probolinggo	Bondowoso	Pasuruan
	Jember	Sampang	Bangkalan
2	Tulungagung	Kediri	Banyuwangi
	Nganjuk	Jombang	Mojokerto
	Gresik	Magetan	Madiun
	Kota Pasuruan Kota Batu	Kota Probolinggo	Lamongan
3	Pacitan	Blitar	Bojonegoro
	Ponorogo	Malang	Tuban
	Trenggalek	Ngawi	Sumenep
4	Kota Malang Sidoarjo	Kota Madiun Kota Kediri	Kota Blitar Kota Surabaya
	Kota Mojokerto		

Sumber: Data diolah (2023)

Berdasarkan Tabel 3, dapat diketahui anggota masing-masing kelompok. Hasil pengelompokan wilayah dengan metode *c-means* diharapkan memiliki perbedaan rata-rata setiap indikator pada setiap kelompok. Selanjutnya, dilakukan pengujian *one-way* MANOVA dan *one-way* ANOVA

untuk menguji apakah ada perbedaan rata-rata pada kelompok yang terbentuk. Sebelumnya dilakukan pengujian normal multivariat dan pengujian homogenitas.

Uji distribusi normal multivariat bertujuan agar peneliti dapat menguji apakah data yang digunakan berdistribusi normal multivariat atau tidak. Hasil uji normal multivariat didapatkan nilai korelasi sebesar 0,986 yang yaitu sebagai berikut.



**Gambar 1. Plot Normal Multivariat**

Sumber: Data diolah (2023)

Nilai korelasi yang didapatkan yaitu sebesar 0,986, dimana nilai ini lebih besar dibandingkan nilai *critical point* dengan alpha 5% yaitu sebesar 0,9700. Sehingga, didapatkan keputusan bahwa gagal tolak  $H_0$  yang berarti data berdistribusi normal multivariat.

Uji homogenitas menggunakan Uji Box's M dengan menggunakan tingkat signifikansi 5%. Kemudian hasil yang didapatkan ditunjukkan pada Tabel di bawah ini.

**Tabel 4. Hasil Pengujian Homogenitas**

	Nilai
Box's M	56,188
F	1,426
df1	30
df2	2198,456
Sig.	0,063

Sumber: Data diolah (2023)

Dengan menggunakan tingkat signifikansi 5% dan derajat bebas sebesar 40, didapatkan nilai  $\chi^2_{\frac{1}{2}(g-1)p(p+1)}$  sebesar 55,758. Nilai Box's M yang diperoleh kurang dari  $\chi^2_{\frac{1}{2}(g-1)p(p+1)}$ . Sehingga

keputusannya adalah gagal tolak  $H_0$  yang memberi kesimpulan bahwa matriks varians kovarians dari data yang digunakan itu adalah homogen.

Berdasarkan pengujian sebelumnya, didapatkan hasil bahwa data berdistribusi normal multivariat dan matriks varians kovarians tidak homogen. Sehingga pengujian dilakukan menggunakan statistik uji *Pillai's Trace*. Tabel 5 berikut ini merupakan hasil pengujian *Pillai's Trace*.

**Tabel 5. Hasil Uji Pillai's Trace**

Nilai <i>Pillai's Trace</i>	F	Derajat bebas	Derajat bebas eror	Sig.	<i>Partial Eta Squared</i>
-----------------------------	---	---------------	--------------------	------	----------------------------

hipotesis					
1,469	7,912	12	99	0,000	0,490

Sumber: Data diolah (2023)

Hasil uji *One-Way* MANOVA ditunjukkan pada Tabel 5, dimana didapatkan nilai  $F$  sebesar 7,912. Jika dibandingkan dengan nilai  $F_{48;396;0,05}$  sebesar 1,390 maka nilai  $F$  lebih besar daripada  $F_{48;396;0,05}$ . Sehingga menghasilkan keputusan tolak  $H_0$  yang berarti terdapat perbedaan pada kelompok yang terbentuk.

Kemudian pengujian perbedaan variabel antar anggota kelompok, dan didapatkan hasil seperti yang ditunjukkan dibawah ini.

**Tabel 6. Hasil Uji *One-Way* ANOVA**

Variabel/Indikator	F	Sig
Angka Harapan Hidup (AHH)	38,529	0,000
Harapan Lama Sekolah (HLS)	23,562	0,000
Rata-rata Lama Sekolah (RLS)	72,012	0,000
Pengeluaran perkapita disesuaikan	39,191	0,000

Sumber: Data diolah (2023)

Berdasarkan Tabel 6, nilai  $F$  dari setiap variabel dibandingkan dengan nilai  $F_{4;33;0,05}$  sebesar 2,659. Ketika dibandingkan dengan  $F_{4;33;0,05}$ , nilai  $F$  keempat variabel lebih besar dari  $F_{4;33;0,05}$  sehingga ditarik keputusan tolak  $H_0$  yang artinya terdapat perbedaan karakteristik keempat variabel terhadap kelompok yang terbentuk sehingga keempat variabel memberikan pengaruh pada terbentuknya kelompok.

Pengelompokan wilayah dengan metode *c-means* pada penelitian ini didapatkan bahwa jumlah kelompok yang terbentuk sebanyak 4 kelompok. Hasil uji *one-way* MANOVA dan *one-way* ANOVA menunjukkan bahwa terdapat perbedaan pada keempat kelompok yang terbentuk dan keempat variabel memberikan pengaruh perbedaan antar kelompok yang terbentuk. Tabel 7 di bawah ini menampilkan rata-rata dari setiap kelompok.

**Tabel 7. Rata-rata Setiap Kelompok**

Kelompok	1	2	3	4
<b>Jumlah Anggota</b>	9	13	9	7
Angka Harapan Hidup	69,1644	72,5108	72,8667	74,0071
Harapan Lama Sekolah	12,8067	13,5823	12,9300	14,8557
Rata-rata Lama Sekolah	6,4089	8,7431	7,4778	10,7914
Pengeluaran perkapita disesuaikan	9920,2222	12338,8462	10303,2222	15346,7143

Sumber: Data diolah (2023)

Berdasarkan Tabel 7 diketahui bahwa kelompok 4 secara rata-rata memiliki nilai tertinggi diantara kelompok lainnya pada masing-masing indikator yang memberikan arti bahwa kelompok 4 memiliki nilai indikator IPM yang paling tinggi yang berarti sudah baik. Selanjutnya, kelompok 2 merupakan kelompok dengan rata-rata tertinggi kedua pada setiap indikator setelah kelompok 4, namun nilai angka harapan hidup kelompok ini sedikit lebih rendah dari kelompok 3 sehingga perlu dilakukan perbaikan pada indikator tersebut. Setelah kelompok 2, ada kelompok 3 yang merupakan kelompok dengan rata-rata terendah pada 3 indikator yaitu rata-rata lama sekolah, pengeluaran perkapita dan harapan lama sekolah sehingga perbaikan perlu difokuskan pada ketiga indikator tersebut. Kemudian, kelompok IPM terendah yaitu kelompok 1 dimana keseluruhan indikator memiliki rata-rata yang paling rendah sehingga perlu dilakukan penanganan atau perbaikan yang terfokus pada keseluruhan indikator tersebut.

## KESIMPULAN

Berdasarkan nilai *pseudo f-statistic*, pengelompokan wilayah pada penelitian ini didapatkan jumlah kelompok optimum adalah sebanyak 4. Pada pengujian *one-way* MANOVA dan *one-way* ANOVA didapatkan hasil bahwa terdapat perbedaan pada kelompok yang terbentuk dan juga keempat variabel memberikan pengaruh terhadap perbedaan rata-rata antar kelompok yang terbentuk. Kelompok 4 memiliki nilai indikator IPM paling tinggi. Kelompok yang menduduki urutan paling tinggi kedua adalah kelompok 2 yang memiliki nilai indikator IPM cukup tinggi namun perlu dilakukan perbaikan pada indikator angka harapan hidup. Kelompok 3 dengan indikator IPM sedang dimana perbaikan perlu difokuskan pada indikator rata-rata lama sekolah, pengeluaran perkapita yang disesuaikan dan harapan lama sekolah. Kelompok 1 adalah kelompok dengan semua indikator IPM paling rendah sehingga perlu adanya perbaikan pada seluruh indikator. Metode pengelompokan wilayah lainnya seperti *fuzzy c-means* yang pernah dilakukan oleh (Harwanti & Rumiati, 2021) juga dapat dilakukan untuk penelitian ini selanjutnya. Selanjutnya, pemilihan metode terbaik juga perlu dilakukan agar hasil pengelompokan dapat digunakan sebagai acuan dalam membuat kebijakan yang tepat seperti yang pernah dilakukan oleh (Kartidan & Irhamah, 2013), (Chusna & Rumiati, 2021) dan (Sikana & Wijayanto, 2021).

## REFERENCES

- Agusta, Y. (2007). K-Means – Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem dan Informatika*, 47-60.
- BPS. (2019). *Statistik Indonesia 2019*. Jakarta: Badan Pusat Statistik.
- BPS. (2021). *Hasil Sensus Penduduk 2020*. Badan Pusat Statistik.
- BPS. (2023). *[Metode Baru] Indeks Pembangunan Manusia menurut Provinsi 2020-2022*. BPS.
- BPS Jawa Timur. (2023). *Angka Harapan Hidup (Tahun), 2020-2022*. Retrieved from <https://jatim.bps.go.id/indicator/26/29/1/angka-harapan-hidup.html>
- Chusna, H. A., & Rumiati, A. T. (2021). Penerapan Metode K-Means dan Fuzzy C-Means untuk Pengelompokan Sekolah Menengah Pertama (SMP) di Indonesia Berdasarkan Standar Nasional Pendidikan (SNP). *Jurnal Sains dan Seni ITS*.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis*. New Jersey: Prentice Hall.
- Harwanti, N. S., & Rumiati, A. T. (2021). Pengelompokan Mutu Sekolah Dasar Di Indonesia Berdasarkan Standar Nasional Pendidikan dengan Metode Fuzzy C-Means. *Jurnal Sains dan Seni ITS*.
- Johnson, R., & Wicherin, D. (2007). *Applied Multivariate Statistical Analysis Sixth Edition*. New Jersey: Prentice Hall.
- Kartidan, H. S., & Irhamah. (2013). Pengelompokan Kabupaten/Kota di Provinsi Jawa Timur Berdasarkan Indikator Pendidikan SMA/SMK/MA dengan Metode C-Means dan Fuzzy C-Means. *Jurnal Sains dan Seni ITS*.
- Kementerian PPN/Bappenas. (2021). *Pemuktahiran Rencana Kerja Pemerintah (RKP) Tahun 2022*.
- Orpin, R. A., & Kostylev, E. V. (2006). Towards a statistically valid method of textural sea floor characterization of benthic habitats. *Marine Geology*.
- Qori'atunnadyah, M. (2022). Pengelompokan Wilayah Berdasarkan Rasio Guru-Murid Pada Jenjang Pendidikan Menggunakan Algoritma K-Means. *Journal of Informatics Development*, 33-38.
- Qori'atunnadyah, M., & Rahmawati, F. D. (2022). Pengelompokan Kabupaten dan Kota Berdasarkan Kondisi Infrastruktur Jalan Menggunakan Hierarchical Clustering. *Journal of Informatics Development*, 1-5.

Sikana, A. M., & Wijayanto, W. A. (2021). Analisis Perbandingan Pengelompokan Indeks Pembangunan Manusia Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarchical Clustering. *Jurnal Ilmu Komputer*.