ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 14, Nomor 1, Tahun 2025, Halaman 179 - 187
Online di: https://ejournal3.undip.ac.id/index.php/gaussian/

PENGGUNAAN *INDEX CALINSKI-HARABASZ* PADA *CLUSTERING K–MEDOIDS ALGORITHM* UNTUK PENGGOLONGAN KABUPATEN/ KOTA DI PROVINSI JAWA TENGAH BERDASARKAN KARAKTERISTIK PENDUDUK MISKIN

Leviana Agita Sari[1*], Arief Rachman Hakim[2], Agus Rusgiyono[3]
[1,2,3] Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro
***e-mail** : lagitas89@gmail.com

DOI: 10.14710/j.gauss.14.1.179-187

Article Info:
Received: 2024-08-12
Accepted: 2025-07-26
Available Online: 2025-07-26

Keywords:
Poverty ,Clustering, K-Medoids Clustering, Index Calinski-Harabasz

Abstract: Poverty is a problem that occurs almost every year. The government is trying hard to reduce poverty because the poverty rate is a measure of the success of a region. Clustering analysis can assist the government in providing targeted assistance. The k-medoids method is a non-hierarchical clustering method for classifying n objects into k clusters based on similar characteristics. This clustering algorithm uses the medoid as its cluster center. The k-medoids method used to overcome the problem of outliers and determine the optimal number of clusters using cluster validation. This research used Index Calinksi-Harabasz. Based on the result of the clustering, the optimal cluster was obtained k=2 using k-medoids method and cluster validation Index Calinksi-Harabasz of 31,53654**.** Cluster 1 consists of 22 districts or cities and in cluster 2 consists of 13 districts or cities.

1. PENDAHULUAN

Kemiskinan merupakan permasalahan yang tidak bisa dianggap sepele, karena masalah ini masih menghantui hampir semua negara setiap tahunnya, terutama negara berkembang seperti Indonesia. Menurut Badan Pusat Statistik (BPS), kemiskinan adalah kondisi seseorang yang tidak mampu untuk memenuhi kebutuhan dasar makanan dan bukan makanan dalam segi ekonomi. Keadaan kemiskinan di Indonesia saat ini mengalami lonjakan dimana proporsi penduduk miskin menjadi 9,57%. Salah satu provinsi yang juga mengalami kenaikan yaitu Provinsi Jawa Tengah. Muh.Saichudin, kepala tim statistik BPS Jawa Tengah mengungkapkan bahwa pada September 2022, jumlah penduduk miskin meningkat 26,79 ribu dan proporsi penduduk miskin meningkat 0,05% dibandingkan Maret 2022. Angka kemiskinan dijadikan sebagai tolak ukur dalam menentukan berhasil tidaknya pemerintah dalam memimpin daerah yang dibawahinya, oleh karena itu pemerintah berlomba-lomba dalam menekan laju kenaikan angka kemiskinan. Berikut adalah beberapa program pemerintah yang diluncurkan dalam menanggulangi masalah kemiskinan yaitu Jaminan Kesehatan Masyarakat (Jamkesmas), bantuan program pendidikan, Bantuan Langsung Tunai (BLT), Program Keluarga Harapan (PHK) dan lain sebagainya. Dalam menunjang program pemerintah agar bantuan yang diberikan tepat sasaran, maka dibutuhkan suatu penggolongan daerah yang menjadi prioritas pemerintah dalam menanggulangi masalah kemiskinan.

Clustering adalah salah satu metode yang digunakan dalam menggolongkan suatu wilayah berdasarkan pada kemiripan karakteristiknya (Primartha,2021). Salah satu metode *clustering* yaitu metode *K-Medoids.* Metode ini bertujuan menggolongkan suatu *n* objek ke dalam *k cluster* berdasarkan pada *medoid* (Vercelis,2009). Penggolongan dengan menggunakan metode ini hampir mirip dengan *K-Means* tetapi pada *K-Medoids* menggunakan objek yang dijadikan pusat *cluster (medoid)* yang didasarkan pada konsep median, sehingga metode ini *robust* terhadap *outlier* (Kaufman dan Rousseuw,1990)

namun disisi lain juga memiliki kelemahan dimana belum mampu dalam menentukan jumlah *cluster* yang optimal secara langsung. Oleh karena itu, dibutuhkan suatu uji validasi *cluster* guna mengetahui jumlah *cluster* yang optimal. Dalam penelitian ini menggunakan metode validasi *Index Calinski-Harabasz* dimana memiliki keunggulan *robust* terhadap data *outlier* (Guerra *et al*., 2012).

2. TINJAUAN PUSTAKA

Menurut Badan Pusat Statistik (2016) mendefinisikan kemiskinan sebagai kondisi seseorang yang tidak mampu dalam memenuhi kebutuhan makanan dan bukan makanan secara finansial berdasarkan pengeluarannya. Kemiskinan terbagi menjadi dua jenis yaitu kemiskinan sementara dan kemiskinan kronis. Kemiskinan sementara adalah kemiskinan yang dialami seseorang dalam waktu jangka pendek, sebaliknya kemiskinan kronis adalah keadaan seseorang yang tergolong miskin dalam waktu jangka panjang tanpa adanya perubahan taraf hidup yang lebih baik (Sony Harry,2007). Suatu permasalahan muncul pasti terdapat sebab dan akibatnya, begitu juga dengan masalah kemiskinan. Menurut Suwandi (2013) mengatakan bahwa kemiskinan disebabkan oleh dua hal, yaitu sifat alami dan kebijakan pemerintah. Sifat alami yang dimaksud adalah sifat yang lahir dari perilaku manusianya itu sendiri, sedangkan untuk kebijakan pemerintah dibagi menjadi 3 aspek yaitu pembangunan pemerintah yang belum merata, dimana masih ada keterbelakangan pembangunan di daerah-daerah terpencil, ketimpangan pembangunan di desa dan di kota dan tidak memberikan perhatian lebih kepada usaha kecil masyarakat miskin. Banyak faktor yang melatarbelakangi terjadinya kemiskinan. Penyebab terjadinya kemiskinan masih terus diperbincangkan baik pada kalangan akademisi maupun ditingkat kebijakan pembangunan (Suryawati,2004). Salah satu hal yang menjadi perbincangan dan perdebatan adalah menggolongan orang yang masuk dalam kriteria penduduk miskin. Dalam penelitian ini menggunakan beberapa karakteristik penduduk miskin yang merupakan bagian dari publikasi Badan Pusat Statistik pada data dan informasi kemiskinan Provinsi Jawa Tengah pada tahun 2022.

Clustering adalah metode penggolongan suatu objek kedalam beberapa *cluster* berdasarkan pada kemiripan karakteristiknya (Primartha,2021). Pada analisis *cluster* terdapat dua asumsi yang harus terpenuhi yaitu:

1. Asumsi *representativeness of the sample*

Uji asumsi ini digunakan untuk mengetahui apakah sampel yang digunakan dalam penelitian dapat mewakili populasi data. Uji asumsi *representativeness of the sample* dapat diidentifikasi dengan menggunakan *Kaiser Mayer Olkin* (KMO). Sampel dapat mewakili populasi jika nilai KMO berkisar antara 0,50-1,00. Berikut merupakan rumus dari Kaiser Mayer Olkin (KMO) menurut Widarjono (2010):

$$KMO = \frac{\sum_{j=1}^{p}\sum_{h=1,h\neq j}^{p} r^2{}_{x_j x_h}}{\sum_{j=1}^{p}\sum_{h=1,h\neq j}^{p} r^2{}_{x_j x_h} + \sum_{j=1}^{p}\sum_{h=1,h\neq j}^{p} \rho^2{}_{x_j x_h, x_m}} \qquad (1)$$

dengan, p adalah banyaknya variabel, $r_{x_j x_h}$ adalah korelasi antara variabel x_j dan x_h, $\rho_{x_j x_h, x_m}$ adalah korelasi parsial antara $x_j$dan x_h dengan menjaga agar x_m konstan.

2. Asumsi non-multikolinieritas

Uji asumsi ini digunakan untuk mengetahui ada tidaknya hubungan linier antara variabel bebas. Dalam analisis *cluster* lebih baik tidak terjadi multikolinieritas antar variabel karena dapat mempengaruhi hasil *cluster* yang terbentuk (Hair *et al.*,2010). Pengujian asumsi ini dilakukan dengan menghitung nilai *Variance Inflation Factor* (VIF). Berikut adalah rumus VIF menurut Gujarati (2009):

$$VIF_j = \frac{1}{(1 - R^2{}_j)} \quad (2)$$

dengan, $R^2{}_j$ adalah koefisien determinasi variabel x_j.
Dikatakan terjadi hubungan yang linier antar variabel bebas jika nilai VIF ≥ 10.

Dalam analisis *cluster* untuk mengukur kemiripan antar objek dibutuhkan ukuran perbedaan antar objek. Salah satu perhitungan ukuran perbedan antar objek yang dapat digunakan yaitu dengan jarak *euclidean.* Pengukuran ini sederhana dan sering dipakai untuk mengukur ukuran perbedaan pada data numerik (Hair *et al*, 2010). Prinsip utama dari jarak *euclidean* adalah semakin kecil ukuran perbedaan antar objek, maka akan semakin mirip suatu objek. Rumus dari jarak *euclidean* adalah sebagai berikut:

$$d_{is} = \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{sj})^2} \quad (3)$$

dengan, p adalah banyaknya variabel, x_{ij} adalah objek data ke-*i* pada variabel ke-*j,* x_{sj} adalah objek data ke-*s* pada variabel ke-*j.*

K-Medoids clustering atau disebut dengan *Partitioning Around Medoid* (PAM) adalah metode penggolongan objek yang hadir sebagai upaya dalam mengatasi kekurangan dari metode *K-Means clustering* yang *sensitive* pada data *outlier*. Dalam penelitian ini untuk mendeteksi ada tidaknya *outlier* pada data yang tergolong *multivariate* atau lebih dari dua variabel dapat menggunakan kuadrat jarak mahalanobis (Folzmiser,2005). Keberadaan pencilan dapat menyebabkan hasil *cluster* yang tidak *representatif* (Barnet dan Lewis, 1994). Rumus dari kuadrat jarak mahalanobis adalah sebagai berikut:

$$d^2{}_{MD(i)} = (\boldsymbol{x_i} - \overline{\boldsymbol{x}})^T \sum^{-1} (\boldsymbol{x_i} - \overline{\boldsymbol{x}}) \quad (4)$$

dengan, $d^2{}_{MD(i)}$ adalah kuadrat jarak mahalanobis pada objek ke-*i*, $\boldsymbol{x_i}$ adalah vektor data ke-*i* yang berukuran $p \; x \; 1$, $\sum$ adalah matriks kovarian berukuran $p \; x \; p$, $\overline{\boldsymbol{x}}$ adalah vektor rata-rata setiap variabel yang berukuran $p \; x \; 1$ dan p adalah banyaknya variabel.
Menurut Johson dan Wichem (2007) menyatakan bahwa data dikatakan *outlier* jika $d^2{}_{MD(i)} >$ *chi-square* $\chi^2{}_{p(1-\alpha)}$.
Prosedur dalam analisis *K-Medoids clustering* adalah sebagai berikut (Suyanto,2017) :

a. Menentukan jumlah *k cluster* terlebih dahulu.
b. Menentukan objek secara random yang akan dijadikan sebagai *medoid* awal.
c. Menghitung ukuran perbedaan objek *non-medoid* dengan objek *medoid* pada iterasi ke-*t*
d. Menggolongkan setiap objek ke dalam golongan *cluster* berdasarkan pada ukuran perbedaan terkecil.
e. Menghitung total ukuran perbedaan $E_{(t)}$ dengan persamaan sebagai berikut:

$$E_{(t)} = \sum_{i=1}^{n} D_i \quad (5)$$

Dengan, D_i adalah ukuran perbedaan terkecil antara objek ke-*i* dan *medoid* pada masing-masing *cluster,* n adalah banyaknya objek.

f. Menentukan kembali secara random objek *non-medoid* pada iterasi ke-*t* sebagai kandidat *medoid* baru dengan syarat tidak menggunakan kembali objek yang sudah pernah menjadi *medoid.*
g. Menghitung ukuran perbedaan antara objek *non-medoid* dengan *medoid* baru.

h. Menggolongkan setiap objek kedalam golongan *cluster* berdasarkan pada ukuran perbedaan terkecil, kemudian hitunglah total ukuran perbedaannya dengan persamaan (5).
i. Apabila $E_{(t)} < E_{(t-1)}$ maka ulangi langkah (f)-(h). Iterasi akan berhenti jika $E_{(t)} \geq E_{(t-1)}$.

Hal penting lainnya yang harus diperhatikan dalam *clustering* adalah menentukan jumlah *cluster* yang optimal. Dalam menentukan jumlah *cluster* yang optimal diperlukan uji validasi *cluster* dan pada penelitian ini menggunakan uji validasi *Index Calinski-Harabasz* karena memiliki keunggulan dimana *robust* terhadap data *outlier* (Guerra *et al*.,2012). Semakin besar nilai *Index CH* menunjukkan jumlah *cluster* yang semakin baik (Khairati A.F, *et al*., 2019). Berikut adalah rumus dari *Index Calinski-Harabasz:*

$$Index\ Calinski-Harabasz\ (CH) = \frac{SSB}{SSW} \times \frac{(n-k)}{(k-1)} \qquad (6)$$

$$SSW = \sum_{l=1}^{k} \sum_{X_{i \in C_l}} \left(X_{ij} - \bar{X}_{jl}\right)^2$$

$$SSB = \sum_{l=1}^{k} n_{jl} \left(\bar{X}_{jl} - \bar{X}_j\right)^2$$

$$\bar{X}_{jl} = \frac{1}{n_{jl}} \sum_{X_{ij \in C_l}} X_{ij}$$

$$\bar{x}_j = \frac{1}{n_j} \sum_{i}^{n_j} X_{ij}$$

CH : nilai *Index Calinski-Harabasz*
SSB : *sum of square between cluster* (tingkat perbedaan objek pada *cluster* yang berbeda)
SSW : *sum of square within cluster* (tingkat perbedaan objek dalam satu *cluster)*
n_j : banyaknya objek pada variabel ke-*j*
n_{lj} : banyaknya objek pada variabel ke-*j cluster* ke-*l*
k : jumlah *cluster*
C_l : *cluster* ke-*l*
X_{ij} : objek ke-*i* variabel ke-*j* pada *cluster* ke-*l*
$\bar{X}_{jl}$: nilai rata-rata variabel ke-*j* pada *cluster* ke-*l*

Setelah jumlah *cluster* yang optimal terbentuk, maka langkah terakhir dalam *clustering* adalah interprestasi dan *profiling* hasil *cluster*. Interprestasi dan *profiling* dilakukan guna mendefinisikan karakteristik dari tiap golongan untuk mengetahui profil dari setiap golongan (Simamora,2005).

3. METODE PENELITIAN

Data yang digunakan dalam penelitian ini bersumber dari website resmi Badan Pusat Statistik Provinsi (BPS) Jawa Tengah https://jateng.bps.go.id/publication.html. Data yang dipilih adalah data yang diterbitkan tahun 2022 tentang kemiskinan yang merupakan bagian dari publikasi data dan informasi kemiskinan Kabupaten/ Kota di Provinsi Jawa Tengah. Sedangkan unit observasinya terdapat 35 Kabupaten/ Kota di seluruh Provinsi Jawa Tengah. Variabel yang digunakan dalam penelitian ini adalah persentase penduduk

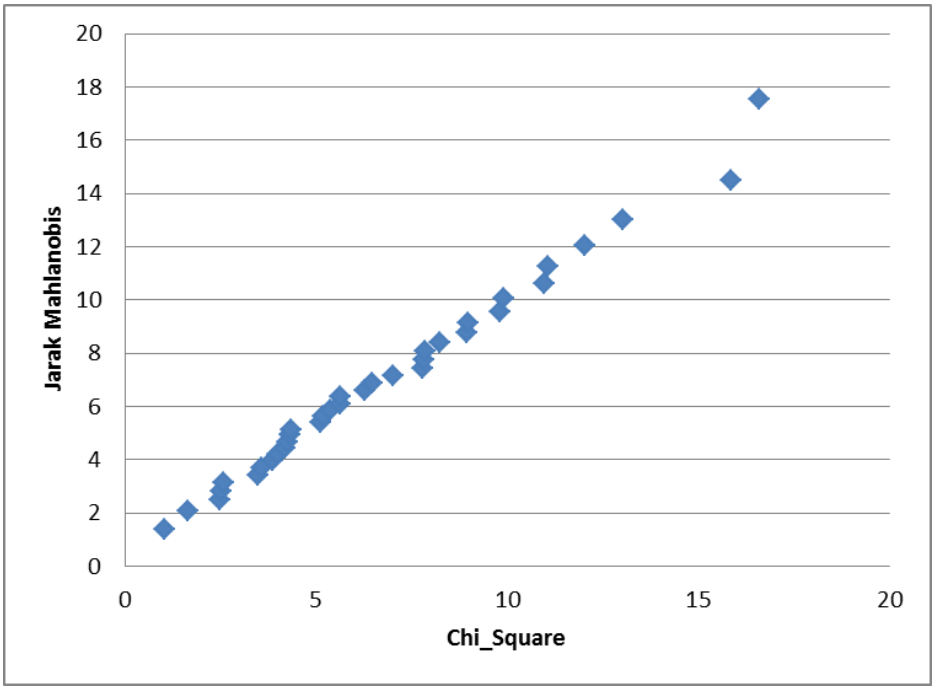
miskin, Angka Partisipasi Sekolah (APS) usia 13-15 tahun, Angka Melek Huruf (AMH) usia 15-55 tahun, persentase penduduk miskin usia 15 tahun keatas yang tidak bekerja (termasuk pengangguran dan bukan angkatan kerja), persentase penduduk miskin usia 15 tahun keatas yang bekerja di sektor pertanian, persentase penduduk miskin usia 15 tahun keatas yang bekerja disektor informal dan persentase pengeluaran perkapita perbulan rumah tangga status miskin untuk komoditi makanan.

Prosedur analisis dalam penelitian ini adalah sebagai berikut:

1. Input data dan analisis statistika deskriptif.
2. Melakukan pendeteksian *outlier* dengan kuadrat jarak mahalanobis.
3. Melakukan uji asumsi *representativeness of the sample.*
4. Melakukan uji asumsi non-multikolinieritas.
5. Menetapkan jumlah *cluster* yang ingin dibentuk yaitu k = 2,3,4 dan 5.
6. Melakukan analisis *K-Medoids clustering.*
7. Menentukan jumlah *cluster* yang optimal dengan uji validasi *Index Calinski-Harabasz.*
8. Interprestasi dan *profiling* hasil *cluster.*

4. HASIL DAN PEMBAHASAN

Dalam menemukan pencilan pada data karakteristik penduduk miskin di Provinsi Jawa Tengah dapat menggunakan kuadrat jarak mahalanobis karena termasuk kasus multivariat. Data tergolong *outlier* apabila $d^2_{MD(i)}$ > *chi-kuadrat* $\chi^2_{p(1-\alpha)}$. Pada penelitian ini menggunakan 7 variabel karakteristik penduduk miskin dengan nilai α = 5%, sehingga diperoleh *chi-kuadrat* $\chi^2_{(1-0.05),7}$ = 14,06714045. Berdasarkan pada hasil perhitungan kuadrat jarak mahalanobis ditemukan dua kabupaten/ kota di Provinsi Jawa Tengah yang merupakan pencilan. Dua kabupaten/ kota tersebut adalah Kota Semarang dan Kota Pekalongan. Berikut adalah gambar *chi-square* plot untuk melihat pencilan secara visual.



Gambar 1. *Chi-square* Plot

Menghapus data pencilan dalam penelitian bukanlah solusi yang bijaksana, karena terkadang data pencilan memberikan informasi yang tidak diberikan oleh data lainnya. Oleh karena itu peneliti melakukan penggolongan objek kedalam beberapa *cluster* dengan

menggunakan metode *K-Medoids clustering*, karena memiliki keunggulan *robust* terhadap data *outlier*.

Terdapat dua asumsi yang harus dipenuhi dalam analisis *clustering* yaitu asumsi *representativenes of the sample* dan asumsi non-multikolinieritas. Uji asumsi *representativenes of the sample* dapat menggunakan nilai KMO dan berdasarkan pada *output* didapatkan nilai KMO sebesar 0,5838689 ≥ 0,50 maka sampel sudah mampu mewakili populasi. Selanjutnya asumsi yang kedua adalah uji non-multikolinieritas dengan menggunakan nilai VIF. Berikut hasil perhitungan nilai VIF yang dapat dilihat pada Tabel A.

Tabel 1. Nilai VIF Masing-masing Variabel

Variabel	Koefisien Determinasi	VIF
X_1	0,3904	1,64042
X_2	0,1819	1,222344
X_3	0,3586	1,559089
X_4	0,429	1,751313
X_5	0,8576	7,022472
X_6	0,8446	6,435006
X_7	0,4113	1,698658

Berdasarkan pada Tabel 1 dapat disimpulkan bahwa asumsi non-multikolinieirtas terpenuhi karena pada semua variabel memiliki nilai VIF < 10.

Berikut hasil penggolongan metode *K-Medoids clustering* dengan ukuran perbedaan menggunakan jarak *euclidean* dan diuji validasi *cluster* dengan *Index Calinski Harabasz* pada k = 2,3,4 dan 5.

Tabel 2. Hasil *K-Medoids Clustering*

Jumlah *Cluster*	*Cluster* ke-	Jumlah Anggota	*Medoid*
2	1	22	Pengamatan ke-12
	2	13	Pengamatan ke-11
3	1	16	Pengamatan ke-12
	2	10	Pengamatan ke-21
	3	9	Pengamatan ke-11
4	1	11	Pengamatan ke -12
	2	10	Pengamatan ke -21
	3	5	Pengamatan ke-8
	4	9	Pengamatan ke -11
5	1	11	Pengamatan ke -12
	2	7	Pengamatan ke -21
	3	5	Pengamatan ke-8
	4	8	Pengamatan ke-11
	5	4	Pengamatan ke -26

K-Medoids clustering memiliki kelemahan dimana belum mampu untuk menentukan jumlah *cluster* optimal secara langsung, oleh karena itu diperlukan uji validasi *cluster* dalam menentukan jumlah k optimal. Dalam kasus ini menggunakan uji validasi *Index Calinski-Harabasz*. *Index Calinski-Harabasz* mempunyai kelebihan tidak *sensitive* atau *robust* terhadap data *outlier* (Guerra *et al.,*2012). Semakin tinggi nilai *index CH* maka akan semakin baik jumlah *cluster* yang terbentuk. Besaran nilai *index CH* pada k = 2,3,4 dan 5 dapat dilihat pada Tabel 3.

Tabel 3. Nilai *Index Calinski-Harabasz*

Jumlah *k* Klaster	*Index CH*
2	**31,53654**
3	27,48157
4	24,90851
5	26,3003

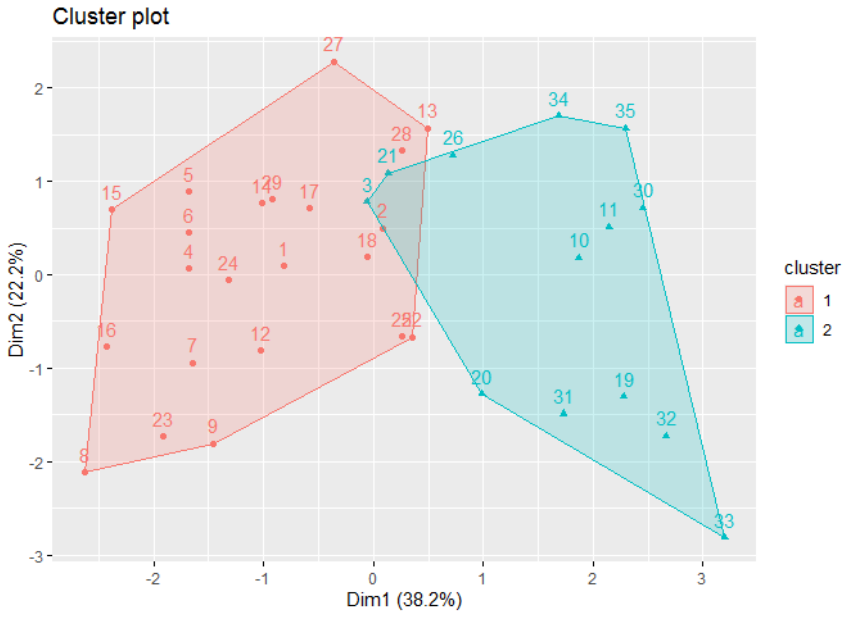
Keterangan: **Bold** mengartikan *Index CH* tertinggi

Berdasarkan pada Tabel 3 diperoleh nilai *Index Calinski-Harabasz* tertinggi ada pada $k = 2$ yaitu sebesar 31,53654 sehingga disimpulkan bahwa jumlah *cluster* yang optimal ada pada $k = 2$. *Profiling* hasil *cluster* terbaik dapat dilihat pada Tabel 4.

Tabel 4. *Profiling* Hasil *Cluster* $k = 2$

Cluster ke-	*Medoid*	Jumlah Anggota *Cluster*	Kabupaten/ Kota
1	Pengamatan ke-12 (Kabupaten Wonogiri)	22	Cilacap, Banyumas, Banjarnegara, Kebumen, Purworejo, Wonosobo, Magelang, Boyolali, Wonogiri, Karanganyar, Sragen, Grobogan, Blora, Rembang, Pati,Semarang, Temanggung, Kendal, Batang, Pemalang, Tegal, Brebes.
2	Pengamatan ke 11 (Kabupaten Sukoharjo)	13	Purbalingga, Klaten, Sukoharjo, Kudus, Jepara, Demak, Pekalongan, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang, Kota Pekalongan, Kota Tegal.

Berikut penggolongan pada $k = 2$ yang ditampilkan pada plot Gambar 2.



Gambar 2. Plot penggolongan $k = 2$

Langkah terakhir dalam proses penggolongan menggunakan *K-Medoids clustering* yaitu interprestasi hasil *cluster*. Pada proses ini dapat direpresentasikan dengan menggunakan *centroid* pada masing-masing variabel. *Centroid* merupakan rata-rata nilai objek yang terdapat dalam setiap *cluster* pada masing-masing variabel. Berikut *output* hasil pengolahan data yang dapat dilihat pada Tabel 5.

Tabel 5. *Centroid* Masing-masing Variabel

Variabel	Keterangan	*Cluster* 1	*Cluster* 2
X_1	Persentase penduduk miskin	11,783636	8,547692
X_2	Angka Partisipasi Sekolah (APS) usia 13-15 tahun	91,98818	94,07846
X_3	Angka Melek Huruf (AMH) usia 15-55 tahun	97,32273	97,86846
X_4	Persentase penduduk miskin usia 15 tahun ke atas yang tidak bekerja (termasuk pengangguran dan bukan angkatan kerja)	39,95273	43,50769
X_5	Persentase penduduk miskin usia 15 tahun ke atas yang bekerja disektor pertanian	30,857273	7,399231
Variabel	**Keterangan**	***Cluster* 1**	***Cluster* 2**
X_6	Persentase penduduk miskin usia 15 tahun ke atas yang bekerja disektor informal	42,46273	24,66692
X_7	Persentase pengeluaran perkapita perbulan rumah tangga status miskin untuk komoditi makanan	61.87545	60.59000

Berdasarkan pada Gambar 2 dan Tabel E dapat diinterprestasikan sebagai berikut:

Cluster 1 adalah golongan wilayah dengan tingkat kemiskinan yang lebih tinggi. *Cluster* 1 memiliki persentase penduduk miskin lebih tinggi jika dibandingkan dengan *cluster* 2, dengan karakteristik penduduk miskin yang paling dominan yaitu pada karakteristik ketenagakerjaan dengan penduduk miskin usia 15 tahun keatas yang bekerja disektor pertanian dan penduduk miskin usia 15 tahun ketas yang bekerja disektor informal memiliki *centroid* yang terpaut jauh lebih tinggi jika dibandingkan dengan *cluster* 2 yaitu sebesar 30,8572735% dan 42,46273%. Pada *cluster* ini, memiliki *centroid* yang sedikit lebih tinggi pada pengeluaran perkapita penduduk miskin pada komoditi makanan yaitu sebesar 61,87545%.

Pada *cluster* 1 (Wilayah berwarna merah) terdiri dari 22 kabupaten/kota yaitu Kabupaten Cilacap, Banyumas, Banjarnegara, Kebumen, Purworejo, Wonosobo, Magelang, Boyolali, Wonogiri, Karanganyar, Grobogan, Sragen, Blora, Rembang, Pati, Semarang, Kendal, Temanggung, Batang, Pemalang, Tegal dan Kabupaten Brebes.

Cluster 2 adalah golongan wilayah dengan tingkat kemiskinan rendah. *Cluster* 2 memiliki persentase penduduk miskin yang lebih rendah jika dibandingkan dengan *cluster* 1, dengan karakteristik penduduk miskin yang terpaut jauh lebih rendah pada karakteristik ketenagakerjaan. Pada penduduk usia 15 tahun keatas yang bekerja disektor pertanian 23,458042% lebih rendah jika dibandingkan dengan *cluster* 1 dan 17,79581% lebih rendah pada penduduk miskin usia 15 tahun keatas yang bekerja disektor informal. Selain itu memiliki *centroid* yang lebih rendah pada pengeluaran perkapita penduduk miskin pada komoditi makanan yaitu sebesar 60.59%.

Pada *cluster* 2 (Wilayah berwarna hijau) terdiri dari 13 kabupaten/ kota yaitu Kabupaten Purbalingga, Klaten, Sukoharjo, Kudus, Jepara, Demak, Pekalongan, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang, Kota Pekalongan dan Kota Tegal.

5. KESIMPULAN

Penggolongan kabupaten/ kota di Provinsi Jawa Tengah dengan menerapkan metode *K-Medoids clustering* dengan uji validasi *cluster Index Calinski-Harabasz* yang berdasarkan karakteristik penduduk miskin diperoleh jumlah *cluster* optimal adalah k =2 dengan *Index Calinski-Harabasz* tertinggi sebesar 31,53654. Pada golongan *cluster* 1 terdiri dari 22 kabupaten/ kota. Golongan *cluster* ini merupakan wilayah yang lebih menjadi prioritas pemerintah dalam masalah kemiskinan dengan tingkat kemiskinan yang

lebih tinggi jika dibandingkan dengan *cluster* 2, pemerintah diharapkan dapat memprioritaskan wilayah *cluster* ini pada karakteristik ketenagakerjaan dengan memberikan program bantuan usaha serta menunjang usaha mikro masyarakat miskin untuk menurunkan persentase penduduk miskin usia 15 tahun keatas yang bekerja disektor informal. Selain itu meluncurkan program peningkatan kesejahteraan tani seperti pelatihan petani dan pelaku agribisnis guna menurunkan persentase penduduk miskin usia 15 tahun ketas yang bekerja disektor pertanian.

DAFTAR PUSTAKA

Afandi, Ta'fif Lukman; Warsito, Budi; Santoso, Rukun.2022. *Implementasi K-Medoids* dan Model *Weighted Length Recency Frequency (W-LRFM) Untuk Segmentasi Pelanggan Dilengkapi Gui R*. Jurnal Gaussian, Vol. 11, No. 3.

Azizah, Anestasya Nur; Widiharih, Tatik; Hakim, Arief Rahman.2022.*Kernel K-Means Clustering Untuk Pengelompokan Sungai di Kota Semarang Berdasarkan Faktor Pencemaran Air.* Jurnal Gaussian, Vol. 11, No. 2.

Arora, P., Deepali, dan Varshney, S. 2015. *Analysis of K-Means and K-Medoids Algorithm for Big Data*. *Procedia Computer Science* Vol. 78.

[BPS]. Badan Pusat Statistik.2022. *Data dan Informasi Kemiskinan Kabupaten/ Kota 2022*. Jawa Tengah.

Barnett, V., dan Lewis, T. 1994. *Outliers in Statistical Data, Wiley Series in Probability and Mathematical Statistics*, New York.

Folzmiser, P. 2005. *Identification of Multivariate Outliers. A Performance Study*. *Australian Journal of Statistics* Vol 34. No. 2.

Gujarati, Damodar. 2009. *Dasar-Dasar Ekonometrika Edisi 5.* Jakarta. Salemba empat

Guerra, L., Robles, V., Bielza, C., & Larranaga, P. 2012. *A Comparion of Clustering Quality Indices Using Outliers and Noise. Intelligent Data Anlysis.* 16.

Harmadi, Sonny Harry B. 2007. *Pengangguran, Kemiskinan dan Pertumbuhan Ekonomi Indonesia*. Dalam Warta Demografi tahun ke 37 No 3. Jakarta. Lembaga Demografi Universitas Indonesia.

Hair, J. F., Black, W. C., Babin, B. J., dan Anderson, R. E. 2010. *Multivariate Data Analysis, Seventh edition*. New Jersey: Prentice Hall International, Inc.

Johnson, R.A., & Wichen, D.W. 2007. *Applied Multivariate Statistical Analysis Sixth Edition. New Jersey: Prentice Hall*

Khairati, A. F., Adlina, A. A., Hertono, G. F., Handari, B. D. 2019. *Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA, PRISMA, Prosiding Seminar Nasional Matematika 2*. ISSN 2613-9189.

Kaufman, L. dan Rousseeuw,P.J.1990. *Finding Groups in Data. An Introduction to Cluster Analysis.* John Wiley, New York.

Primartha, R. 2021. *Algoritma Machine Learning*. Bandung. Informatika.

Suryawati. 2004. *Teori Ekonomi Mikro*. UPP. AMP YKPN. Yogyakarta. Jarnasy.

Suwadi, Widodo. 2013. *Ekonomi Sumber Daya Dan Lingkungan*.

Suyanto.2017. *Data Mining untuk Klasifikasi dan Klasterisasi Data*.Bandung.Informatika.

Simamora, B. 2005. *Analisis Multivariat Pemasaran*. Jakarta. PT Gramedia Pustaka Utama

Vercellis, Carlo. 2009. *Business intelligence: Datamining and optimization for decision making.* Chichester: John Wiley& Sons.

Widarjono, A. 2010. *Analisis Statistika Multivariat Terapan*. Yogyakarta: UPP STIM YKPN.