

PERBANDINGAN METODE FUZZY C-MEANS DAN GUSTAFSON-KESSEL DALAM PENENTUAN CLUSTER TINGKAT RISIKO PENULARAN TUBERCULOSIS TERHADAP PENYAKIT DI JAWA TIMUR

Atika Nurzida^{1*}, Iut Tri Utami², Masithoh Yessi Rochayani³

^{1,2,3}Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

*e-mail: atika.nurzida@alumni.undip.ac.id

DOI: 10.14710/j.gauss.13.2.373-382

Article Info:

Received: 2023-06-12

Accepted: 2024-11-29

Available Online: 2024-12-2

Keywords:

Clustering, Fuzzy C-Means,
Gustafson-Kessel, Dunn Index,
Tuberculosis

Abstract: Tuberculosis (TB) is a type of disease that is easily transmitted to people with weak immune systems. The disease factors that are most susceptible to TB transmission in East Java Province are Diabetes Mellitus (DM) sufferers, malnutrition toddlers, and HIV sufferers. The aim of this study was to form clusters in districts/cities with DM sufferers, malnutrition toddlers, and HIV sufferers to minimize the risk of TB transmission. Fuzzy clustering analyzes that are often used are Fuzzy C-Means and Gustafson-Kessel. Fuzzy C-Means uses the Euclidean distance squared function to form spherical clusters, while Gustafson-Kessel uses the Mahalanobis distance squared function to form ellipsoid clusters. The optimal number of clusters is determined by the Dunn Index (DI) value. The results showed that the formation of regional clusters was optimal in the Gustafson-Kessel method with 2 clusters as indicated by a DI value of 0.974603. Cluster 1 consists of 4 regencies/cities with an average of DM sufferers, malnutrition toddlers, and HIV sufferers namely 86.448 people, 2.841 people, and 87 people. Cluster 2 consists of 34 districts/cities with an average of DM sufferers, malnutrition toddlers, and HIV sufferers namely 10.449 people, 3.364 people, and 72 people.

1. PENDAHULUAN

Penyakit Tuberculosis (TB) sering disebut sebagai salah satu penyakit akibat bakteri *Mycobacterium tuberculosis*. Bakteri *M. tuberculosis* dapat menyerang paru-paru dan organ tubuh lainnya. Penyakit TB disebut sebagai penyebab kematian utama di dunia. Menurut WHO (2023), TB mudah menular pada orang dengan sistem imunitas tubuh yang lemah seperti penderita Diabetes Melitus, penderita HIV, orang yang kekurangan gizi, dan perokok. Menurut publikasi data Dinkes Provinsi Jawa Timur, penyakit dengan angka paling tinggi tertular TB adalah Diabetes Melitus dan HIV, sedangkan angka balita dengan gizi kurang juga yang paling tinggi dari kelompok anak yang rentan terhadap TB. UNICEF (2018) mengatakan bahwa balita dengan sistem kekebalan tubuh yang lemah atau kekurangan gizi sangat rentan terhadap penyakit Tuberculosis yang ditunjukkan oleh prevalensi kejadian TB sebesar 20 persen. Penderita Diabetes Melitus berisiko 8,9 kali tertular TB (Mihardja *et al.*, 2015), penderita HIV berisiko 11,7 kali tertular TB (Tulu *et al.*, 2014), dan balita dengan gizi kurang 1,78 kali tertular TB (Widyastuti *et al.*, 2021). Provinsi Jawa Timur tahun 2021 berada di urutan ke-3 daerah dengan kasus TB terbanyak di Indonesia, tetapi sebanyak 53 persen kabupaten/kota di Provinsi Jawa Timur belum mampu mencapai 90 persen keberhasilan pengobatan TB (Kominfo, 2022).

Tingginya kasus TB dapat berpotensi menyebabkan endemi TB. Pengelompokan wilayah berdasarkan faktor penyakit yang rentan terhadap penularan TB dapat membantu meminimalisir risiko penularan TB serta mencegah endemi TB. *Clustering* dalam penelitian ini digunakan untuk memetakan titik data penderita Diabetes Melitus, balita dengan gizi kurang, dan penderita HIV agar dapat diketahui prioritas wilayah kabupaten/kota yang

segera ditangani. Metode *clustering* yang digunakan adalah metode Fuzzy C-Means dan Gustafson-Kessel.

Metode yang sering digunakan untuk *clustering* adalah metode Fuzzy C-Means dan Gustafson-Kessel. Metode Fuzzy C-Means adalah metode *clustering* yang memungkinkan suatu bagian dari data memiliki dua atau lebih *cluster* di mana tiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaan. Nilai derajat keanggotaan yang tinggi mengindikasikan bahwa semakin dekat jarak titik data terhadap pusat *cluster*. Metode Gustafson-Kessel adalah salah satu metode *fuzzy clustering* yang mengadaptasi algoritma standar pada Fuzzy C-Means dengan menggunakan fungsi perhitungan jarak yang berbeda. Fuzzy C-Means menggunakan fungsi perhitungan kuadrat jarak Euclidean yang membentuk *cluster spherical*, sedangkan Gustafson-Kessel menggunakan fungsi perhitungan kuadrat jarak Mahalanobis yang membentuk *cluster ellipsoid* (Lesot dan Kruse, 2008).

Inisialisasi parameter c (banyak *cluster*) dan m (bobot fuzzy) sangat penting sebagai Langkah awal proses *clustering*. Perhitungan menggunakan parameter c dapat dilakukan secara berulang kali untuk memperoleh *cluster* yang optimal. Penentuan *cluster* yang optimal dengan menggunakan validasi, salah satunya validasi *clustering* Dunn Index. Validasi Dunn Index berdasarkan kriteria internal yang menghitung nilai minimum dari perbandingan jarak antara dua *cluster* sebagai pemisah (*separation*) dengan nilai maksimum dari diameter *cluster* sebagai ukuran kedekatan (*compactness*) (Khairati *et al.*, 2019).

Penelitian sebelumnya tentang metode Fuzzy C-Means dan Gustafson-Kessel yaitu mengelompokkan saham menggunakan indeks LQ45 yang dilakukan oleh Rahmatika *et al.* (2015) diperoleh hasil bahwa metode terbaik adalah metode Gustafson-Kessel. Amalia *et al.* (2016) juga melakukan penelitian tentang perbandingan metode Fuzzy C-Means dan Gustafson-Kessel untuk mengelompokkan perusahaan sub sektor perbankan berdasarkan rasio keuangan CAMELS dan diperoleh hasil metode Gustafson-Kessel yang terbaik dengan jumlah *cluster* sebanyak 2 *cluster*.

Tujuan penelitian ini adalah untuk mengelompokkan wilayah Kabupaten/Kota di Provinsi Jawa Timur yang rentan tertular penyakit TB. Perbandingan metode Fuzzy C-Means dan Gustafson-Kessel dalam penelitian ini diharapkan dapat menunjukkan pembentukan *cluster* yang optimal. Hasil penelitian ini juga diharapkan dapat menjadi referensi bagi pemerintah Provinsi Jawa Timur untuk membuat strategi baru terkait pencegahan dan penanggulangan penyakit Tuberculosis. Data yang digunakan adalah data publikasi Dinas Kesehatan Provinsi Jawa Timur tahun 2021. Variabel yang digunakan diantaranya: jumlah penderita Diabetes Melitus, balita dengan gizi kurang, dan penderita HIV. Pengolahan data menggunakan *software* GUI R.

2. TINJAUAN PUSTAKA

Clustering adalah metode yang dikategorikan dalam *unsupervised learning* yang mana dalam prosesnya mengelompokkan objek data ke dalam *cluster* tertentu. Tiap data pengamatan dalam suatu *cluster* mempunyai kesamaan (similaritas) yang tinggi satu sama lainnya, tetapi mempunyai kesamaan (similaritas) yang rendah pada objek antar-*cluster* (Tan *et al.*, 2006). Kriteria kemiripan tersebut diukur dengan jarak sehingga proses ini disebut juga sebagai *distance-based clustering*.

Fuzzy clustering dikategorikan sebagai *soft clustering* dimana tiap titik data dapat menjadi dua atau lebih *cluster* (Putri *et al.*, 2019). *Fuzzy clustering* menggunakan konsep berpikir dengan *logika fuzzy* di mana nilai dapat mengandung nilai intermediet yaitu sebagian benar atau sebagian salah. *Fuzzy clustering* tidak memaksakan titik data untuk masuk ke dalam suatu *cluster*, melainkan diberikan nilai derajat keanggotaan. Nilai derajat keanggotaan berada pada rentang 0 hingga 1.

Ada asumsi yang harus dipenuhi dalam *clustering* yaitu sampel yang mewakili (*representative sample*) dan asumsi non-multikolinearitas. Uji *representative sample* menunjukkan apakah sampel dapat mewakili populasi atau tidak. Asumsi *representative sample* dapat dipenuhi apabila nilai Kaiser-Meyer Olkin (KMO) berada di antara 0,5 sampai 1 (Widarjono, 2010). Rumus KMO ditunjukkan pada Persamaan (1).

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2} \quad (1)$$

Uji non-multikolinearitas dilakukan untuk mengetahui ada atau tidaknya korelasi antarvariabel independen dengan nilai koefisien korelasi. Nilai koefisien korelasi harus di bawah 0,8 agar tidak terjadi multikolinearitas (Santoso, 2012). Rumus koefisien korelasi ditunjukkan pada Persamaan (2).

$$r_{ab} = \frac{n \sum_{i=1}^n ab - (\sum_{i=1}^n a)(\sum_{i=1}^n b)}{\sqrt{\{n \sum_{i=1}^n a^2 - (\sum_{i=1}^n a^2)\}\{n \sum_{i=1}^n b^2 - (\sum_{i=1}^n b^2)\}}} \quad (2)$$

Normalisasi data merupakan proses menskalakan nilai data agar berada pada rentang nilai yang diharapkan. Salah satu metode normalisasi data yang sering digunakan adalah *min-max normalization* yang mana cara kerjanya adalah dengan mengubah rentang nilai data untuk memperoleh nilai perbandingan yang seimbang saat sebelum dan sesudah proses normalisasi data. Nilai *min-max normalization* ini berkisar antara 0 hingga 1 (Fenny, 2019). Rumus normalisasi min-max ditunjukkan pada Persamaan (3).

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3)$$

Metode Fuzzy C-Means menggunakan fungsi jarak Euclidean untuk mengukur jarak dari data pengamatan ke pusat *cluster*. Metode ini memungkinkan suatu bagian dari data memiliki dua atau lebih kelompok, dalam hal ini nilai derajat keanggotaan menentukan tiap titik data dalam suatu *cluster*. Nilai derajat keanggotaan setiap data diperbaiki secara berulang hingga diperoleh fungsi objektif yang minimal. Nilai derajat keanggotaan yang tinggi mengindikasikan bahwa semakin dekat jarak titik data terhadap pusat *cluster*.

Berikut ini penerapan algoritma Fuzzy C-Means.

1. Menginisialisasi nilai parameter awal, diantaranya: banyaknya *cluster* yang diperlukan ($c \geq 2$), *fuzzifier* ($m > 1$), iterasi maksimum (t_{max}), iterasi awal ($t = 1$), error terkecil (ϵ), serta fungsi objektif awal *cluster* ke-k sebesar 0
2. Membangkitkan bilangan random $\mu_{ik}, i = 1, 2, \dots, n; k = 1, 2, \dots, c$ yang disimbolkan dengan matriks U_0 .

$$U_0 = \begin{bmatrix} \mu_{11} & \dots & \mu_{1c} \\ \vdots & \ddots & \vdots \\ \mu_{1n} & \dots & \mu_{nc} \end{bmatrix} \quad (4)$$

Matriks U ini sebagai fungsi keanggotaan yang mana dipilih secara acak. Nilai ini berada pada kisaran 0 hingga 1. Perhatikan rumus $\sum \mu_{ik}^m = 1$

3. Menghitung nilai pusat *cluster* ke-k (v_k) pada Persamaan (5)

$$v_{kj} = \frac{\sum_{i=1}^n (\mu_{ik})^m x_{ij}}{\sum_{i=1}^n (\mu_{ik})^m} \quad (5)$$

4. Menghitung nilai kuadrat jarak Euclidean pada Persamaan (6)

$$D_{ik}^2 = \|x_{ij} - v_{kj}\|^2 = \left(\sqrt{\sum_{j=1}^p (x_{ij} - v_{kj})^2} \right)^2 \quad (6)$$

5. Menghitung nilai fungsi objektif pada iterasi ke-t pada Persamaan (7)

$$J(t) = \sum_{k=1}^c \sum_{j=1}^p (\mu_{ik})^m D_{ik}^2 \quad (7)$$

6. Menghitung nilai dari fungsi keanggotaan yang baru pada Persamaan (8)

$$U_{(t+1)} = \frac{[\sum_{j=1}^p D_{ik}^2]^{-\frac{1}{m-1}}}{\sum_{k=1}^c [\sum_{j=1}^p D_{ik}^2]^{-\frac{1}{m-1}}} \quad (8)$$

7. Mengecek kondisi berhenti yaitu:

- Apabila $|J(t) - J(t-1)| < \varepsilon$ atau bisa dengan $(t > t_{max})$, maka dapat dikatakan sudah konvergen.
- Apabila $|J(t) - J(t-1)| \geq \varepsilon$, maka kembali ke langkah 3 yaitu mencari nilai pusat *cluster*.

Gustafson-Kessel merupakan bentuk adaptasi dari algoritma standar dari Fuzzy C-Means dengan menggunakan fungsi perhitungan jarak yang berbeda. Fuzzy C-Means menggunakan fungsi perhitungan kuadrat jarak Euclidean sehingga bentuk *cluster* adalah bulat sempurna (*spherical*), sedangkan Gustafson-Kessel menggunakan fungsi perhitungan kuadrat jarak Mahalanobis sehingga bentuk *cluster* adalah *ellipsoid* (Lesot dan Kruse, 2008). Fungsi perhitungan kuadrat jarak mahalanobis pada Gustafson-Kessel selalu diperbaharui pada setiap iterasi dengan menggunakan nilai determinan matriks *fuzzy covariance* (Gustafson dan Kessel, 1979).

Berikut ini penerapan algoritma metode Gustafson-Kessel.

- Menginisialisasi nilai parameter awal, diantaranya: banyaknya *cluster* yang diperlukan ($c \geq 2$), *fuzzifier* ($m > 1$), iterasi maksimum (t_{max}), error terkecil (ε), serta iterasi awal ($t = 1$)
- Membangkitkan bilangan random μ_{ik} , $i = 1, 2, \dots, n$; $k = 1, 2, \dots, c$ yang disimbolkan dengan matriks U_0 .

$$U_0 = \begin{bmatrix} \mu_{11} & \cdots & \mu_{1c} \\ \vdots & \ddots & \vdots \\ \mu_{1n} & \cdots & \mu_{nc} \end{bmatrix} \quad (9)$$

Matriks U ini sebagai fungsi keanggotaan yang mana dipilih secara acak. Nilai ini berada pada kisaran 0 hingga 1. Perhatikan rumus $\sum \mu_{ik}^m = 1$.

3. Menghitung nilai pusat *cluster* ke-k (v_k) pada Persamaan (10)

$$v_{kj} = \frac{\sum_{i=1}^n (\mu_{ik})^m x_{ij}}{\sum_{i=1}^n (\mu_{ik})^m} \quad (10)$$

4. Menghitung nilai matriks kovarian *cluster* pada Persamaan (11)

$$F_k = \frac{\sum_{i=1}^n (\mu_{ik})^m (x_{ij} - v_{kj})(x_{ij} - v_{kj})^T}{\sum_{i=1}^n (\mu_{ik})^m}, k = 1, 2, \dots, c \quad (11)$$

5. Mencari nilai kuadrat jarak Mahalanobis pada Persamaan (12)

$$D_{ikAk}^2 = (x_{ij} - v_{kj})^T \left[\frac{\rho_k \mathbf{F}_k^{-1}}{\det(\mathbf{F}_k)^p} \right] (x_{ij} - v_{kj}) \quad (12)$$

6. Menghitung nilai fungsi objektif pada iterasi ke-t pada Persamaan (13)

$$J_{(t)} = \sum_{k=1}^c \sum_{j=1}^p (\mu_{ik})^m D_{ikAk}^2 \quad (13)$$

7. Menghitung nilai dari fungsi keanggotaan yang baru \mathbf{U}_{t+1}

$$\mathbf{U}_{(t+1)} = \frac{[\sum_{j=1}^p D_{ik}^2]^{-\frac{1}{m-1}}}{\sum_{k=1}^c [\sum_{j=1}^p D_{ik}^2]^{-\frac{1}{m-1}}} \quad (14)$$

8. Mengecek kondisi berhenti yaitu dengan membandingkan nilai keanggotaan dalam matriks \mathbf{U} , kondisi konvergen apabila nilai $\|\mathbf{U}_{(t+1)} - \mathbf{U}_{(t)}\| < \varepsilon$ atau ($t > t_{max}$), sedangkan kondisi tidak konvergen apabila nilai $\|\mathbf{U}_{(t+1)} - \mathbf{U}_{(t)}\| \geq \varepsilon$ sehingga kembali ke langkah 3.

Penggunaan jarak merupakan langkah yang sangat penting dalam proses *clustering* karena digunakan untuk mengukur kesamaan (similaritas) objek data. Jarak Euclidean adalah jarak geometris antardua objek data yang ditunjukkan apabila semakin dekat jarak Euclidean, maka semakin mirip suatu objek data. Jarak Mahalanobis digunakan dalam *clustering* untuk mengatasi kelemahan jarak Euclidean. Jarak Euclidean sensitif terhadap besarnya sampel dan sebaran varian, sedangkan jarak Mahalanobis lebih fleksibel atau lebih dapat menyesuaikan pola data karena perhitungannya didasarkan pada korelasi antarvariabel.

Metode validasi *Dunn Index* merupakan salah satu ukuran validasi berdasarkan kriteria internal. Metode validasi *Dunn Index* menghitung nilai minimum dari perbandingan nilai jarak antara dua *cluster* sebagai *separation* dengan *compactness* yaitu nilai maksimum dari diameter *cluster* (Khairati *et al.*, 2019). Menurut Kassambara (2017), nilai *Dunn Index* dapat dihitung sebagai berikut: (1) untuk setiap *cluster*, hitung jarak antara setiap objek di dalam *cluster* dan objek-objek di *cluster* lainnya, (2) gunakan minimum jarak berpasangan sebagai *inter-cluster separation* (min.separation), (3) hitung jarak antarobjek dalam *cluster* yang sama untuk setiap *cluster* yang sama, (4) gunakan jarak maksimal dalam *cluster* (max.diameter) sebagai jarak *intra-cluster compactness*, (5) hitung nilai *Dunn Index* dengan persamaan berikut.

$$DI = \frac{\text{min. separation}}{\text{max. diameter}} \quad (15)$$

Kualitas *cluster* yang baik ditunjukkan dengan nilai *Dunn Index* (DI) yang semakin besar.

3. METODE PENELITIAN

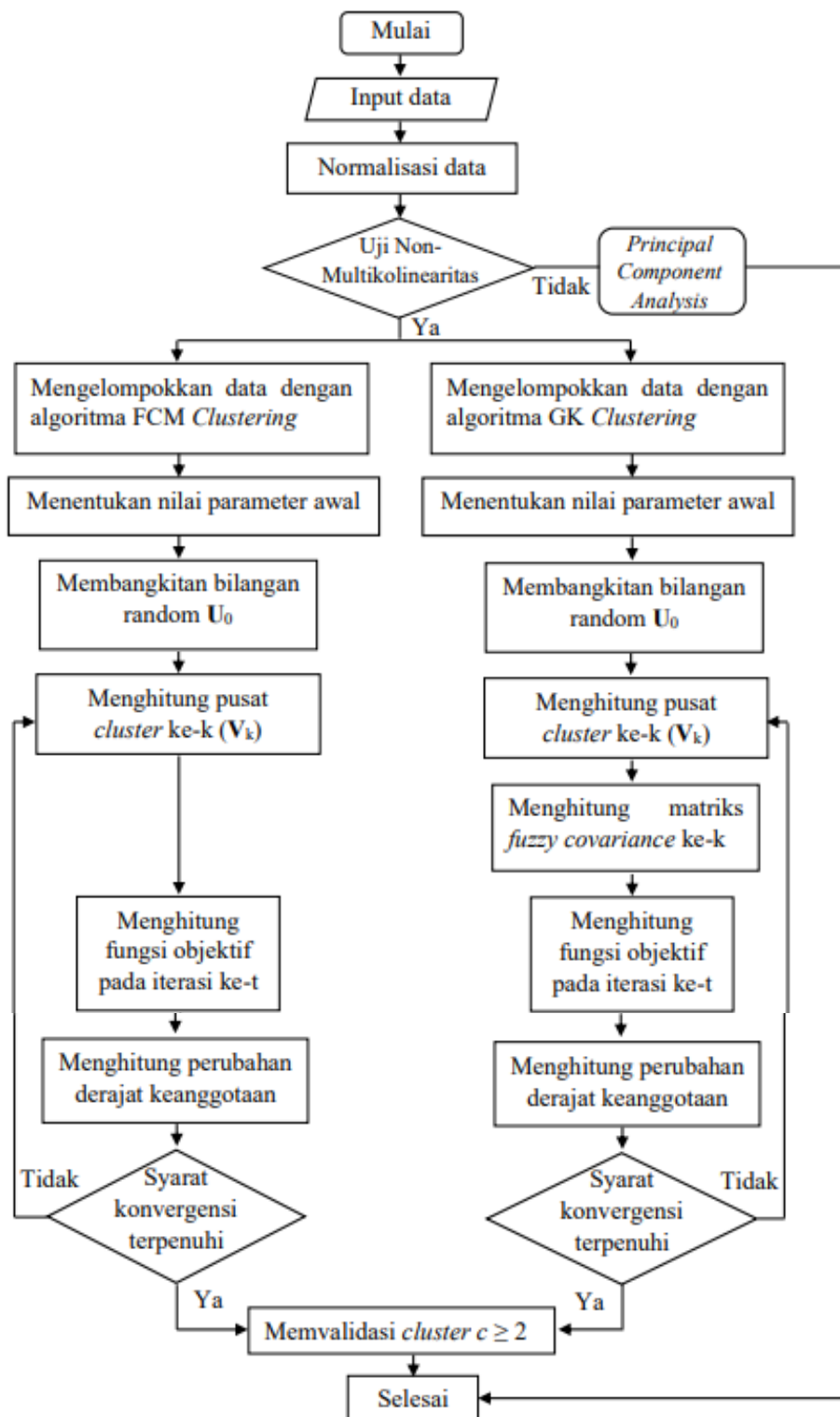
Data penelitian ini diperoleh dari publikasi tahunan “Profil Kesehatan” oleh Dinas Kesehatan Provinsi Jawa Timur tahun 2021 yaitu merupakan faktor risiko penularan TB terhadap penyakit dengan variabelnya:

x_1 : Jumlah penderita Diabetes Melitus

x_2 : Jumlah balita dengan gizi kurang

x_3 : Jumlah penderita HIV

Berikut ini tahapan analisis data yang diringkas dalam diagram alir.



Gambar 1. Flowchart Tahapan Analisis Data

4. HASIL DAN PEMBAHASAN

Hasil ringkasan data melalui statistika deskriptif untuk masing-masing variabel penderita Diabetes Melitus, balita dengan gizi kurang, dan penderita HIV dengan banyak data adalah 38 data ditunjukkan oleh **Tabel 1**.

Tabel 1. Statistika Deskriptif

	N	Minimum	Maximum	Mean	Std.Deviation
Diabetes Melitus	38	4.187	96.280	24.468,68	20.290,42
Gizi Kurang	38	275	18.664	3.694,95	3.443,93
HIV	38	10	493	101,74	99,67
Valid N (listwise)	38				

Normalisasi data yang digunakan adalah metode Min-Max dimana metode ini melakukan perubahan rentang nilai dengan menggunakan nilai tertinggi dan nilai terendah data asli. Perhitungan normalisasi data menggunakan Persamaan (3). Hasil data yang dinormalisasi ditunjukkan oleh **Tabel 2**.

Tabel 2. Data Hasil Normalisasi

No.	Diabetes Melitus	Gizi Kurang	HIV
1	0,03402	0,073522	0,006211
2	0,122984	0,162053	0,10766
⋮	⋮	⋮	⋮
38	0,004854	0,012453	0,026915

Pengujian asumsi non-multikolinearitas dilakukan dengan menggunakan nilai koefisien korelasi antarvariabel independen dengan kriteria non-multikolinearitas adalah nilai koefisien korelasi $< 0,8$. Perhitungan nilai koefisien korelasi menggunakan Persamaan (2).

Tabel 3. Nilai Koefisien Korelasi Antarvariabel Independen

	Diabetes Melitus	Gizi Kurang	HIV
Diabetes Melitus	1	0,2444483	0,5806528
Gizi Kurang	0,2444483	1	0,3021816
HIV	0,5806528	0,3021816	1

Tabel 3 memberikan hasil bahwa nilai koefisien korelasi antarvariabel independen di bawah 0,8 ($< 0,8$), maka kesimpulannya adalah tidak terjadi hubungan antarvariabel independen atau asumsi non-multikolinearitas terpenuhi.

Inisialisasi parameter awal untuk proses *clustering* menggunakan algoritma Fuzzy C-Means dan Gustafson-Kessel adalah sama. Proses *clustering* dilakukan sebanyak 3 kali yaitu dari *cluster* 2 sampai 4 dengan parameter $m = 2$, $t_{max} = 1000$, $\varepsilon = 10^{-9}$, nilai fungsi objektif awal *cluster* ke- k sebesar 0. Hasil *clustering* Fuzzy C-Means ditunjukkan oleh **Tabel 4**, sedangkan hasil *clustering* Gustafson-Kessel ditunjukkan oleh **Tabel 5**.

Tabel 4. Output Hasil *Clustering* dengan Fuzzy C-Means (FCM)

Jumlah <i>Cluster</i> (<i>c</i>)	Iterasi	Fungsi Objektif	Dunn Index
2	45	2,03921	0,739133
3	428	1,27818	0,653394
4	595	0,87040	0,582624

Tabel 5. Output Hasil *Clustering* dengan Gustafson-Kessel (GK)

Jumlah <i>Cluster</i> (<i>c</i>)	Iterasi	Fungsi Objektif	Dunn Index
2	38	4,055058 x 10 ¹⁴	0,974603
3	177	2,375333 x 10 ¹⁴	0,911705
4	216	1,514081 x 10 ¹⁴	0,883825

Hasil *clustering* dari kedua metode yaitu Fuzzy C-Means (FCM) dan Gustafson-Kessel (GK) kemudian dilakukan perbandingan dengan nilai dari validasi Dunn Index. Jumlah *cluster* yang digunakan dengan 2 *cluster*, 3 *cluster*, dan 4 *cluster* ditunjukkan oleh **Tabel 6**.

Tabel 6. Perbandingan Metode FCM dan GK

Jumlah <i>Cluster</i> (<i>c</i>)	Fuzzy C-Means (FCM)	Gustafson-Kessel (GK)
2	0,739133	0,974603
3	0,653394	0,911705
4	0,582624	0,883825

Tabel 6 menunjukkan hasil *clustering* terbaik dalam penelitian ini adalah metode Gustafson-Kessel dan jumlah *cluster* paling optimal adalah ketika pembentukan *cluster* sebanyak 2 *cluster*. Hasil pengolahan menggunakan GUI R menunjukkan nilai fungsi objektif sebesar 4,550584 x 10¹⁴ dan banyak iterasi sebanyak 38 iterasi. Metode Gustafson-Kessel memberikan hasil pusat *cluster* akhir sebagai berikut.

$$v_{kj} = \begin{bmatrix} 86.448 & 2.841 & 87 \\ 10.449 & 3.364 & 72 \end{bmatrix}$$

Interpretasi hasil yang diperoleh dari hasil akhir pusat *cluster* adalah:

- Cluster* 1 terdiri dari 4 kabupaten/kota dengan rata-rata penderita Diabetes Melitus, balita dengan gizi kurang, dan penderita HIV masing-masing adalah 86.448 orang, 2.841 orang, dan 87 orang.
- Cluster* 2 terdiri dari 34 kabupaten/kota dengan rata-rata penderita Diabetes Melitus, balita dengan gizi kurang, dan penderita HIV masing-masing adalah 10.449 orang, 3.364 orang, dan 72 orang.

5. KESIMPULAN

Berdasarkan perbandingan metode Fuzzy C-Means (FCM) dan Gustafson-Kessel (GK) *Clustering* dengan menggunakan validasi Dunn Index, diperoleh metode terbaik yaitu metode Gustafson-Kessel. Analisis *clustering* ini menghasilkan *cluster* optimal yaitu 2 *cluster* dengan karakteristik *cluster* 1 terdiri dari 4 kabupaten/kota dengan rata-rata penderita Diabetes Melitus, balita dengan gizi kurang, dan penderita HIV masing-masing adalah

86.448 orang, 2.841 orang, dan 87 orang, sedangkan *cluster* 2 terdiri dari 34 kabupaten/kota dengan rata-rata penderita Diabetes Melitus, balita dengan gizi kurang, dan penderita HIV masing-masing adalah 10.449 orang, 3.364 orang, dan 72 orang.

DAFTAR PUSTAKA

- Amalia, N.A., Widodo, D.A., & Oktaviana, P.P. 2016. *Analisis Clustering Perusahaan Sub Sektor Perbankan Berdasarkan Rasio Keuangan CAMELS Tahun 2014 Menggunakan Metode Fuzzy C-Means dan Fuzzy Gustafson-Kessel*. Jurnal GARUDA Vol. 5, No. 2: Hal. 321 – 326.
- Dinkes Provinsi Jawa Timur. 2021. *Profil Kesehatan 2021*. Dinkes Jatim.
- Fenny, D. 2019. *Analisis Perbandingan Cosine Normalization dan Min-Max Normalization pada Pengelompokan Terjemahan Ayat Al-Qur'an Menggunakan Algoritma K-Means Clustering*. Skripsi. Jakarta: UIN Syarif Hidayatullah.
- Gustafson, D. & Kessel, W. 1979. *Fuzzy Clustering with a Fuzzy Covariance Matrix*, in: Proceedings of the IEEE CDC, San Diego, CA, USA, pages: 761 – 766.
- Kassambara, A. 2017. *Practical Guide to Cluster Analysis in R 1st Edition*. STHDA.com
- Khairati, A.F., Adlina, A., Hertono, G., & Handari, B. 2019. *Kajian Index Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA* Jurnal PRISMA Vol. 2: Hal. 161 – 170.
- Kominfo Provinsi Jawa Timur. 2022. *Peringati Hari TBC Sedunia, Khofifah Bertekad Jatim Capai Eliminasi TBC di Tahun 2030*. URL: <https://kominfo.jatimprov.go.id/berita/peringati-hari-tbc-sedunia-khofifah-bertekad-jatim-capai-eliminasi-tbc-di-tahun-2030>. Diakses tanggal 23 Februari 2023.
- Lesot, M. J. & Kruse, R. 2008. *Gustafson-Kessel-like Clustering Algorithm Based on Typicality Degrees*. Uncertainty & Intelligence Information Systems Journal., pp. 117 – 130.
- Mihardja, L., Lolong, D.B., & Lannywati, G. 2015. *Prevalensi Diabetes Melitus pada Tuberkulosis dan Masalah Terapi*. Jurnal Ekologi Kesehatan Vol. 14, No. 4: Hal. 350 – 358.
- Putri *et al.*, R. K. 2019. *Implementasi Algoritma Modified Gustafson-Kessel untuk Clustering Tweets pada Akun Twitter Lazada Indonesia*. Jurnal Gaussian Vol. 8, No. 3: Hal. 285 - 295.
- Rahmatika, L., Suparti, & Safitri, D. 2015. *Analisis Kelompok dengan Algoritma Fuzzy C-Means dan Gustafson Kessel Clustering pada Indeks LQ45*. Jurnal Gaussian Vol. 4, No. 3: Hal. 543 – 552.
- Santoso, S. 2012. *Panduan Lengkap SPSS Versi 20*. Jakarta: PT Elex Media Komputindo.
- Tan, P.N., Steinbach, M., & Kumar, V. 2006. *Introduction to Data Mining*. Boston: Pearson Education.
- Tulu, B., Dida, N., Kassa, Y., & Taye, B. 2014. *Smear Positive Pulmonary Tuberculosis and Its Risk Factors Among Tuberculosis Suspect in South East Ethiopia; a Hospital Based Cross-sectional Study*. BMC Res Notes, Vol. 7: Pages. 285.
- UNICEF. 2018. *An Agenda for Action on Childhood Tuberculosis*. URL: https://data.unicef.org/wp-content/uploads/2018/03/TB-Advocacy-Brochure-Final-3_21-high-Res.pdf. Diakses tanggal 24 Juni 2023.
- WHO. 2023. *Tuberculosis*. URL: <https://www.who.int/news-room/fact-sheets/detail/tuberculosis>. Diakses tanggal 2 Maret 2023

- Widarjono, A. 2010. *Analisis Statistika Multivariat Terapan Edisi pertama*. Yogyakarta: UPP STIM YKPN.
- Widyastuti, N., Nugraheni, W., Wahyono, T., & Yovsyah, Y. 2021. *Hubungan Status Gizi dan Kejadian Tuberculosis Paru pada Anak Usia 1-5 Tahun di Indonesia (Analisis Data Riskesdas 2018)*. Buletin Penelitian Sistem Kesehatan. Vol. 24, No. 2, Hal: 89 – 96.