

Penggunaan Metode *Fuzzy C-Means* untuk Pengelompokan Kabupaten/Kota Berdasarkan Jumlah Tenaga Kesehatan

Surya Cahyadi Syam^{1*}, Georgina Maria Tinungki², Anisa³

¹²³Departemen Statistika, Fakultas MIPA, Universitas Hasanuddin, Makassar, 90245, Indonesia

* Corresponding author, email: suryacahyadi.syam@gmail.com

Abstract

Clustering is the process of partitioning data into a number of clusters, which in the cluster have similar characteristics to one another and have differences with other clusters. In this study, clustering is used to group district/cities based on the number of health workers in South Sulawesi using fuzzy c-means. To get the results of grouping, there are 11 health worker variables used. In determining the optimum number of clusters, the Pseudo F-Statistic is used. In the first test, the optimum cluster was obtained in 5 groups and the second test was in 6 groups without makassar and bone. From the clustering analysis in the first test, it was found that groups 2 and 3 only have one member. In the second test, group 1 had 4 members, groups 2 and 6 had 3 members, group 3 had 2 members, and groups 4 and 5 had 5 members.

Keywords: *Clustering, Fuzzy C-Means, Group, Health Workers, Pseudo F-Statistic.*

Abstrak

Clustering adalah proses partisi data ke dalam sejumlah klaster, yang mana di dalam klaster memiliki kemiripan karakteristik antar satu sama lainnya dan memiliki perbedaan dengan klaster lainnya. Dalam penelitian ini clustering digunakan untuk pengelompokan kabupaten/kota berdasarkan jumlah tenaga kesehatan di Sulawesi Selatan dengan menggunakan fuzzy c-means. Untuk mendapatkan hasil pengelompokan, terdapat 11 variabel tenaga kesehatan yang digunakan. Dalam menentukan jumlah klaster optimum digunakan Pseudo F-Statistic. Dalam pengujian pertama didapatkan klaster optimum pada 5 kelompok dan pengujian kedua pada 6 kelompok tanpa makassar dan bone. Dari hasil analisis pengelompokan pada pengujian pertama didapatkan bahwa kelompok 2 dan 3 hanya memiliki satu anggota. Pada pengujian kedua, kelompok 1 memiliki 4 anggota, kelompok 2 dan 6 memiliki 3 anggota, kelompok 3 memiliki 2 anggota, dan kelompok 4 dan 5 memiliki 5 anggota.

Kata Kunci: *Clustering, Fuzzy C-Means, Kelompok, Pseudo F-Statistic, Tenaga Kesehatan.*

1. Pendahuluan

Pengelompokan adalah proses pemisahan, pemecahan, atau segmentasi data ke dalam sejumlah klaster agar objek-objek yang tergabung dalam sebuah kelompok merupakan objek-objek yang mirip satu sama lain dan berbeda dengan objek dalam kelompok lainnya [1,7]. Metode pengelompokan terdiri atas metode hirarki dan metode non-hirarki. Adapun metode hirarki adalah metode pengelompokan yang terstruktur dan

bertahap di mana jumlah kelompok belum diketahui sejak awal. Sedangkan, metode non-hirarki dimulai dengan menentukan terlebih dahulu jumlah kelompok yang diinginkan [8]. Salah satu metode pengelompokan non-hirarki yang populer adalah *Fuzzy C-Means* (FCM).

Fuzzy C-Means (FCM) atau yang juga dikenal sebagai *Fuzzy ISODATA*, merupakan salah satu metode pengelompokan yang merupakan perluasan dari metode *Hard K-Means* [15]. FCM menggunakan model pengelompokan *fuzzy* sehingga suatu data dapat menjadi anggota dari semua kluster yang terbentuk dengan derajat atau tingkat keanggotaan yang berbeda-beda pada rentang 0 hingga 1. Tingkat keberadaan data dalam suatu kelas atau kluster ditentukan oleh derajat keanggotaannya. Teknik ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981 [2,6].

Dalam menentukan jumlah kluster (c) yang akan digunakan pada metode pengelompokan non-hirarki, terdapat berbagai cara objektif, salah satunya dengan menggunakan *Calinski-Harabasz Pseudo F-Statistic* [14]. *Calinski-Harabasz Pseudo F-Statistic* dapat digunakan untuk menentukan jumlah kluster terbaik, di mana nilai *Calinski-Harabasz* yang lebih tinggi berasosiasi dengan model kluster yang lebih jelas dan terpisah dengan baik. *Calinski-Harabasz* didefinisikan sebagai rasio rata-rata sebaran antar kluster (varian eksternal) dan sebaran dalam kluster (varian internal) [3].

Penelitian ini bertujuan untuk menentukan jumlah kluster optimum menggunakan *Calinski-Harabasz Pseudo F-Statistic* dan melakukan pengelompokan kabupaten/kota di Provinsi Sulawesi Selatan berdasarkan jumlah tenaga kesehatan menggunakan *Fuzzy C-Means*. Analisis dan pemetaan tenaga medis ini dinilai sangat penting untuk mendukung pemerataan sumber daya kesehatan berbasis data di tingkat regional [12], [13].

2. Material dan Metode

2.1. Statistika Deskriptif

Statistika deskriptif (*descriptive statistics*) berkaitan dengan penerapan metode statistik untuk mengumpulkan, mengolah, menyajikan, dan menganalisis data kuantitatif secara deskriptif. Penelitian kami digunakan pemusatan data berupa rata-rata untuk mengetahui karakteristik dari setiap jumlah tenaga kesehatan atau setiap variabel. Berikut persamaan (1) yang merupakan rumus rata-rata dari tiap variabel.

$$\bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n} \quad (1)$$

dengan:

- \bar{x}_j : rata-rata tiap variable
- x_{ij} : kabupaten/kota ke- i dan variabel ke- j
- n : jumlah kabupaten/kota

Kemudian untuk ukuran keseragaman data tiap variabel digunakan standar deviasi. Dengan keterangan yang sama seperti rata-rata, berikut rumus dari standar deviasi tiap variabel pada persamaan (2).

$$S_j = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n - 1}} \quad (2)$$

Bentuk sekelompok data, ukuran pemusatan data, dan penyebaran data yang diamati dapat digambarkan dalam bentuk *box plot*. Box plot atau boxplot (juga dikenal sebagai diagram *box-and-whisker*) merupakan suatu *box* (kotak berbentuk bujur sangkar). Dalam *boxplot* juga ditunjukkan, jika ada, nilai *outlier* dari observasi. *Boxplot* dapat digunakan untuk menunjukkan perbedaan antara populasi tanpa menggunakan asumsi distribusi statistik yang mendasarinya. Karenanya, *boxplot* tergolong dalam statistik non-parametrik. Jarak antara bagian-bagian dari *box* menunjukkan derajat dispersi (penyebaran) dan *skewness* (kecondongan) dalam data. Dalam penggambarannya, boxplot dapat digambarkan secara horizontal maupun vertikal. Nilai yang berada di atas atau dibawah *whisker* pada *boxplot* dinamakan nilai *outlier* atau ekstrim

Outlier atau pencilan adalah titik sampel yang memiliki karakteristik unik dan dapat diidentifikasi secara jelas karena berbeda dengan mayoritas titik sampel lainnya. *Outlier* dapat dideteksi secara multivariate (berdasarkan *Mahalanobis Distences*) dengan hipotesis berikut ini:

H_0 : tidak terdapat *outlier* pada objek ke- i ($i = 1, 2, 3, \dots, n$)

H_1 : terdapat *outlier* pada objek ke- i

Statistik Uji:

$$d_i^2 = (\mathbf{X}_i - \bar{\mathbf{X}})' \mathbf{S}^{-1} (\mathbf{X}_i - \bar{\mathbf{X}}) \quad (3)$$

dengan:

$$\mathbf{S} = \frac{1}{n - 1} \sum_{i=1}^n \left(\begin{bmatrix} (x_{i1} - \bar{x}_1) \\ (x_{i2} - \bar{x}_2) \\ \vdots \\ (x_{im} - \bar{x}_m) \end{bmatrix} [x_{i1} - \bar{x}_1 \quad x_{i2} - \bar{x}_2 \quad \dots \quad x_{im} - \bar{x}_m]' \right)$$

$$\mathbf{S} = \begin{bmatrix} S_{1^2} & S_{12} & \dots & S_{1m} \\ S_{12} & S_{2^2} & \dots & S_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ S_{1m} & S_{2m} & \dots & S_{m^2} \end{bmatrix} \quad (4)$$

dengan:

- n : banyaknya objek
- p : banyaknya variabel
- d_i^2 : jarak mahalanobis
- S : matriks varians-kovarians

Untuk dapat menolak H_0 apabila $d_i^2 > \chi^2(0, 05;p)$.

2.2. Calinski-Harabasz Pseudo F-Statistic

Pseudo F-Statistic digunakan untuk mengetahui jumlah kluster optimum yang akan digunakan dalam metode *clustering*. Ada berbagai macam metode yang dapat digunakan dalam menentukan banyaknya kelompok optimum, Calinski-Harabasz *Pseudo F-Statistic* adalah salah satunya.

Berikut ini adalah persamaan yang digunakan untuk menghitung nilai *Pseudo F-Statistic*:

$$C - H \text{ Pseudo } F - \text{Statistic} = \frac{\left(\frac{R^2}{k-1}\right)}{\left(\frac{1-R^2}{n-k}\right)} \quad (5)$$

dengan:

$$R^2 = \left(\frac{SST - SSW}{SST}\right)$$

$$SST = \sum_{i=1}^n \sum_{j=1}^p \sum_{k=1}^c (x_{ijk} - \bar{x}_j)^2$$

$$SSW = \sum_{i=1}^n \sum_{j=1}^p \sum_{k=1}^c (x_{ijk} - \bar{x}_{jk})^2$$

Keterangan:

- SST : Total jumlah dari kuadrat jarak sampel terhadap rata-rata keseluruhan
- SSW : Total jumlah dari kuadrat jarak sampel terhadap rata-rata kelompoknya
- n : banyaknya sampel
- p : banyaknya variabel
- c : banyaknya kelompok
- x_{ijk} : sampel ke- i pada variabel ke- j kelompok ke- k
- \bar{x}_j : rata-rata seluruh sampel pada variabel ke- j
- \bar{x}_{jk} : rata-rata seluruh sampel pada variabel ke- j dan kelompok ke- k

Nilai *Pseudo F-Statistic* tertinggi menunjukkan bahwa jumlah kelompok yang digunakan untuk partisi data telah optimal [4].

2.3. Fuzzy C-Means

Fuzzy C-Means Clustering merupakan suatu teknik pengklasteran data yang mana keberadaan tiap-tiap data dalam suatu klaster ditentukan oleh nilai keanggotaan. Konsep dasar *Fuzzy C-Means Clustering*, pertama kali adalah menentukan pusat klaster yang akan menandai lokasi rata-rata untuk tiap klaster. Pada kondisi awal, pusat klaster masih belum akurat. Tiap-tiap data yang memiliki derajat keanggotaan untuk tiap-tiap klaster. Dengan cara memperbaiki pusat klaster dan nilai keanggotaan tiap-tiap data secara berulang, maka akan dapat dilihat bahwa pusat klaster akan bergerak menuju lokasi yang tepat [5].

Langkah-langkah algoritma FCM [6]:

1. Input data yang akan dikelompokkan (X), berupa matriks berukuran $n \times p$ (n = jumlah sampel data, p = variabel setiap data).
2. Menentukan jumlah *cluster* (c), pangkat pembobot ($m > 1$), maksimum iterasi (*MaxIter*), *error* terkecil yang diharapkan (ϵ), Fungsi objektif awal ($P_0 = 0$).
3. Bangkitkan bilangan random U_{ik} , dimana $i = 1, 2, 3, \dots, n; k = 1, 2, \dots, c$;
4. Menghitung pusat *cluster* ke- k pada variabel ke- j ; v_{kj} , dengan $k = 1, 2, \dots, c; j = 1, 2, \dots, p$;

$$V = [v_{kj}] = \frac{\sum_{i=1}^n (u_{ik})^m x_{ij}}{\sum_{i=1}^n (u_{ik})^m}$$

5. Menghitung nilai fungsi objektif pada iterasi ke- t (P_t) dengan formula:

$$P_t = \sum_{i=1}^n \sum_{k=1}^c (u_{ik})^m d_{ik}(x_i, v_{kj})$$

6. Menghitung perubahan matriks keanggotaan u_{ik} dengan formula:

$$u_{ik} = \left[\frac{[\sum_{j=1}^p d_{jk}]^{\frac{1}{m-1}}}{\sum_{k=1}^c [\sum_{j=1}^p d_{jk}]^{\frac{1}{m-1}}} \right]^{-1}$$

Memeriksa kondisi berhenti: Jika ($|P_t - P_{t-1}| < \epsilon$ atau ($t > \text{MaxIter}$) maka iterasi berhenti; Jika tidak, maka $t = t + 1$, kembali ke langkah 4.

3. Hasil dan Diskusi

3.1 Statistika Deskriptif

Statistika deskriptif digunakan untuk mengetahui gambaran secara umum karakteristik masing-masing variabel yang digunakan mengenai jumlah tenaga

kesehatan di setiap Kabupaten/Kota di Provinsi Sulawesi Selatan. Variabel dari penelitian ini adalah peubah prediktor (X) yang diantara lain X_1 = tenaga medis; X_2 = perawat; X_3 = bidan; X_4 = tenaga farmasi; X_5 = tenaga kesehatan masyarakat; X_6 = tenaga kesehatan lingkungan; X_7 = ahli gizi; X_8 = tenaga keterampilan fisik; X_9 = tenaga keteknisan medis; X_{10} = tenaga teknik biomedika; X_{11} = tenaga penunjang kesehatan. Deskripsi kabupaten/kota di Sulawesi Selatan dapat diketahui berdasarkan ukuran pemusatan dan ukuran penyebaran. Ukuran pemusatan yang digunakan adalah rata-rata, minimum dan maksimum. Sedangkan untuk penyebaran yang digunakan adalah standar deviasi.

3.1.1 Deskripsi Variabel

Tabel 1. Deskripsi Variabel

Variabel	Minimum	Q ₁	Median	Rata-Rata	Q ₃	Maksimum	SD
X_1	47	80,75	96,5	142,46	108,25	1133	213,831
X_2	231	367,8	410,5	582,8	550,2	3120	573,1946
X_3	183	241	290	384	73	1472	288,3535
X_4	33	57,5	71,5	97,5	96,75	558	102,3965
X_5	19	54,75	79	85,55	104	266	51,79837
X_6	19	27,75	32	39,8	46,75	131	23,50701
X_7	24	32	39	49,5	45,5	222	40,82305
X_8	2	5,75	7,5	11,46	10,25	90	17,17425
X_9	12	23,75	34	46,96	57,25	226	43,48761
X_{10}	25	40	54,5	66,46	67	402	73,47048
X_{11}	70	148,8	182,5	289,8	309,8	1847	351,8829

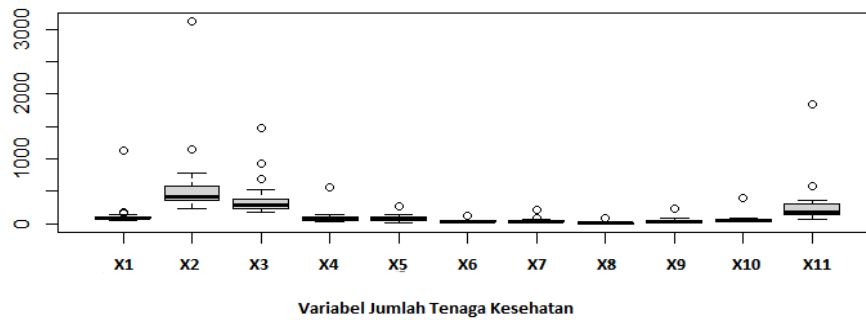
Tabel 1 menunjukkan bahwa tenaga perawat secara rata-rata merupakan tenaga kesehatan tertinggi di Sulawesi Selatan, diikuti dengan tenaga kebidanan dengan rata-rata tertinggi kedua dan tenaga penunjang kesehatan rata-rata tertinggi ketiga di Sulawesi Selatan pada tahun 2019. Untuk tenaga kesehatan terendah secara rata-rata yaitu tenaga kesehatan lingkungan.

Jumlah tenaga kesehatan tertinggi yaitu tenaga perawat, jumlah terendah terdapat di Kabupaten Enrekang sebanyak 231 orang dan jumlah tertinggi terdapat di Kota Makassar sebanyak 3.120 orang. Sementara untuk jumlah tenaga kesehatan terendah yaitu tenaga teknik keterampilan fisik, jumlah terendah terdapat di Kabupaten Jeneponto sebanyak 2 orang dan jumlah tertinggi terdapat di Kota Makassar sebanyak 90 orang.

3.1.2 Uji Outlier secara Multivariat dengan Metode Mahalanobis Distance

Dari hasil statistik deskripsi, kita dapat mengetahui ada ketimpangan atau ketidakmerataan jumlah tenaga kesehatan di beberapa variabel, diantaranya tenaga medis, keterampilan fisik, teknik biomedika dan tenaga penunjang kesehatan. Kita juga dapat melihat bahwa dari 11 variabel, Kota Makassar memiliki nilai maksimum yang paling mendominasi yakni pada 10 variabel.

Selanjutnya akan dilakukan pengujian apakah terdapat Kabupaten/Kota yang menjadi *outlier* pada data jumlah tenaga kesehatan yang diteliti. Berikut adalah *Boxplot* dari tiap variabel untuk melihat *outlier* dari tiap variabel.



Gambar 1 *Boxplot* Jumlah Tenaga Kesehatan

Boxplot pada Gambar 1 menunjukkan keragaman dari masing-masing variabel jumlah tenaga kesehatan. Dapat dilihat pula bahwa semua variabel jumlah tenaga kesehatan memiliki *outlier*. Apabila dilihat dengan pengujian *outlier* secara multivariat atau keseluruhan variabel dengan hipotesis

H_0 : tidak terdapat *outlier* pada obyek ke- i ($i = 1, 2, \dots, 24$)

H_1 : terdapat *outlier* pada obyek ke- i

H_0 ditolak apabila $d_i^2 > X^2_{(0,05;11)}$. Berdasarkan nilai dari *Mahalanobis Distance* pada tabel 4.2 berikut ini didapatkan hasil bahwa dari 24 kabupaten/kota di Provinsi Sulawesi Selatan terdapat 2 kabupaten/kota yang *outlier* yakni Kabupaten Bone dan Kota Makassar.

Tabel 2. Hasil Perhitungang *Mahalanobis Distance*

Kabupaten/Kota	d_i^2
KEPULAUAN SELAYAR	5,382683

BULUKUMBA	13,18307
BANTAENG	15,036353
JENEPONTO	10,220797
TAKALAR	12,44394
GOWA	8,560775
SINJAI	8,042424
MAROS	15,007304
PANGKAJENE DAN KEPULAUAN	14,883567
BARRU	6,662536
BONE	19,760025
SOPPENG	8,299709
WAJO	1,842154
SIDENRENG RAPPANG	14,997314
PINRANG	14,844393
ENREKANG	7,55456
LUWU	5,290664
TANA TORAJA	8,564079
LUWU UTARA	6,077621
LUWU TIMUR	6,698035
TORAJA UTARA	6,928405
KOTA MAKASSAR	21,79637
KOTA PAREPARE	11,170328
KOTA PALOPO	9,752895

***bold** : Data *Outlier*

3.2 Pengelompokan Kabupaten/Kota di Sulawesi Selatan menggunakan dengan Metode Fuzzy C-Means

Setelah mendeskripsikan karakteristik pada kabupaten/kota di Sulawesi Selatan berdasarkan jumlah tenaga kesehatan, langkah selanjutnya adalah melakukan pengelompokan kabupaten/kota di Sulawesi Selatan berdasarkan jumlah tenaga kesehatan menggunakan metode *Fuzzy C-Means*. Batasan Jumlah kelompok yang digunakan pada penelitian ini yakni 2 kelompok hingga 6 kelompok. Dalam menentukan jumlah kelompok optimum digunakan metode Pseudo F. Nilai *pseudo F* yang terbesar menunjukkan bahwa kelompok tersebut mampu memberikan hasil pengelompokan yang optimal, dimana keragaman dalam kelompok sangat homogen sebaliknya antar kelompok sangat heterogen. Untuk mencari nilai C-H *Pseudo F-Statistics* menggunakan persamaan (5) sebagai berikut:

$$C - H \text{ Pseudo } F - \text{Statistic} = \frac{\left(\frac{R^2}{k-1}\right)}{\left(\frac{1-R^2}{n-k}\right)}$$

$$R^2 = \frac{SST - SSW}{SST}$$

$$SST = \sum_{i=1}^{24} \sum_{j=1}^{11} \sum_{k=1}^2 (x_{ijk} - \bar{x}_j)^2 = 13903017,628$$

$$SSW = \sum_{i=1}^{24} \sum_{j=1}^{11} \sum_{k=1}^2 (x_{ijk} - \bar{x}_{jk})^2 = 2858842,957$$

$$R^2 = \frac{SST - SSW}{SST} = \frac{13903017,628 - 2858842,957}{13903017,628} = 0,794282247$$

$$C - H \text{ Pseudo } F - \text{Statistic} = \frac{\left(\frac{R^2}{k-1}\right)}{\left(\frac{1-R^2}{n-k}\right)} = \frac{\left(\frac{0,794282247}{2-1}\right)}{\left(\frac{1-0,794282247}{24-2}\right)} = 84,94301$$

Untuk hasil perhitungan *pseudo F* selanjutnya dapat dilihat pada tabel berikut.

Tabel 3. *Pseudo F Statistics*

Kelompok	Nilai Pseudo F
2	84,94301
3	126,8443
4	150,9231
5	161,8689
6	152,3278

***bold** : nilai pseudo F tertinggi

Tabel 3 menunjukkan bahwa nilai pseudo F tertinggi antara 2 kelompok hingga 6 kelompok terdapat pada 5 kelompok yakni sebesar 161,8689, sehingga hasil pengelompokan kabupaten/kota di Sulawesi Selatan optimum dengan menggunakan 5 kelompok/*cluster* (c). Selanjutnya dilakukan pengelompokan dengan metode *Fuzzy C-Means*.

Data jumlah tenaga kesehatan Provinsi Sulawesi Selatan dilakukan pengelompokan atau *clustering* dengan menggunakan *Fuzzy C-Means* sebagai berikut:

1. Input data yang akan dikelompokkan X , berupa matriks berukuran $n \times m$ (dimana n adalah jumlah sampel data atau jumlah Kabupaten/Kota dan p adalah jumlah atribut/variabel, dengan $n = 24$ dan $p = 11$).
2. Menentukan nilai parameter awal
 - a. Jumlah *cluster* (c) : 5
 - b. Pangkat (w) : 2
 - c. Maksimum Iterasi (MaxIter) : 100
 - d. Error terkecil yang diharapkan (ϵ) : 10^{-9}
 - e. Fungsi obyektif awal (P_0) : 0
 - f. Iterasi awal (t) : 1
3. Membangkitkan bilangan random U_{ik} , dimana $i = 1,2,3,\dots,n; k = 1,2,\dots,c$;
Berikut adalah matriks partisi awal U yang terbentuk (secara random) :

$$U_{0(24 \times 5)} = \begin{pmatrix} 0,10040 & 0,25720 & 0,32303 & 0,29741 & 0,02193 \\ 0,18833 & 0,26387 & 0,12797 & 0,19119 & 0,22862 \\ 0,26205 & 0,11124 & 0,33945 & 0,22344 & 0,06380 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0,05089 & 0,05895 & 0,21303 & 0,45854 & 0,21857 \\ 0,19319 & 0,35857 & 0,03706 & 0,40354 & 0,00762 \\ 0,228431 & 0,09708 & 0,23286 & 0,11543 & 0,32617 \end{pmatrix}$$

4. Menghitung pusat *cluster* ke- k pada variabel ke- j ; v_{kj} , dengan $k = 1,2,\dots,c; j = 1,2,\dots,p$;

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

$$V_{5 \times 11} = \begin{pmatrix} 125,437 & 538,791 & 379,698 & \dots & 44,146 & 61,692 & 253,193 \\ 123,855 & 533,425 & 384,354 & \dots & 44,269 & 60,857 & 278,648 \\ 145,267 & 596,994 & 401,367 & \dots & 46,488 & 68,420 & 297,255 \\ 163,122 & 625,311 & 378,071 & \dots & 50,516 & 73,763 & 315,011 \\ 147,997 & 590,325 & 389,360 & \dots & 48,670 & 65,865 & 294,215 \end{pmatrix}$$

5. Menghitung Fungsi objektif (P_t)
Fungsi objektif pada iterasi pertama P_1 dihitung dengan menggunakan persamaan:

$$P_1 = \sum_{i=1}^{24} \sum_{k=1}^5 \left(\left[\sum_{j=1}^{11} (x_{ij} - v_{kj})^2 \right] (\mu_{ik})^2 \right) = 29000735,75$$

6. Menghitung Perubahan Matriks Partisi (U)
Perubahan matriks partisi dihitung menggunakan persamaan berikut ini:

$$\mu_{ik} = \frac{\left[\sum_{j=1}^p (x_{ij} - v_{kj})^2 \right]^{\frac{-1}{m-1}}}{\sum_{k=1}^c \left[\sum_{j=1}^p (x_{ij} - v_{kj})^2 \right]^{\frac{-1}{m-1}}}$$

Berikut adalah hasil perhitungan perubahan matriks partisi baru (U_1) untuk iterasi pertama:

$$U_{1_{24 \times 5}} = \begin{pmatrix} 0,097 & 0,183 & 0,300 & 0,346 & 0,074 \\ 0,308 & 0,265 & 0,128 & 0,123 & 0,177 \\ 0,148 & 0,121 & 0,295 & 0,312 & 0,125 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0,114 & 0,121 & 0,222 & 0,318 & 0,226 \\ 0,253 & 0,315 & 0,105 & 0,282 & 0,044 \\ 0,271 & 0,160 & 0,223 & 0,109 & 0,237 \end{pmatrix}$$

7. Mengecek kondisi berhenti :

Karena $|P_1 - P_0| = |29000735,75 - 0| = 29000735,75 \gg \varepsilon(10^{-9})$, dan Iterasi = 1 < $MaxIter(= 100)$, maka proses dilanjutkan ke iterasi ke-2.

Iterasi dilanjutkan hingga diperoleh ($|P_t - P_{t-1}| < \varepsilon$) atau ($t > MaxIter$). Pada tahap selanjutnya digunakan *syntax* Algoritma *Fuzzy C-Means* pada aplikasi *RStudio* dalam menentukan kelompok tiap obyek.

Output dari *syntax* yang digunakan memberikan informasi diantaranya:

1. Hasil pertama yaitu hasil perhitungan nilai fungsional dan diperoleh bahwa diperlukan sebanyak 100 kali iterasi sebelum memperoleh solusi optimum bagi nilai fungsional $J_m(U, V)$ sebesar 277225.
2. Hasil yang kedua yaitu hasil perhitungan nilai-nilai V_{kj} sebagai berikut :

$$V_{5 \times 11} = \begin{pmatrix} 92,061 & 358,301 & 260,561 & \dots & 40,416 & 45,817 & 157,282 \\ 159,726 & 113,877 & 1456,697 & \dots & 80,177 & 75,924 & 368,209 \\ 1132,834 & 3119,604 & 932,955 & \dots & 225,977 & 401,945 & 1846,785 \\ 125,080 & 625,480 & 315,023 & \dots & 32,628 & 63,308 & 334,542 \\ 85,412 & 435,088 & 346,685 & \dots & 32,632 & 50,750 & 197,343 \end{pmatrix}$$

Nilai-nilai tersebut merupakan nilai dari “koordinat” titik pusat dari ke 5 *cluster*.

3. Hasil yang ketiga merupakan perhitungan dari nilai-nilai U_{ik} . Berdasarkan dari nilai-nilai U_{ik} tersebut dapat diperoleh hasil pengelompokan pada Tabel berikut.

Tabel 4. Hasil Pengelompokan seluruh Kabupaten/Kota

Kabupaten/Kota	Kelompok	Jumlah
Kep. Selayar, Jeneponto, Takalar, Gowa, Sinjai, Pangkep, Soppeng, Wajo, Enrekang, Luwu, Luwu Timur	1	11
Bone	2	1
Kota Makassar	3	1
Bulukumba, Maros, Kota Pare-Pare, Kota Palopo	4	4
Bantaeng, Barru, Sidrap, Pinrang, Tana Toraja, Luwu Utara, Toraja Utara	5	7

Jumlah anggota kelompok berdasarkan Tabel 4 yakni kelompok 1 terdiri dari 6 kabupaten/kota, kelompok 4 terdiri dari 15 kabupaten kota, masing-masing kelompok 2 dan 3 terdiri dari 1 kabupaten/kota. Karena terdapat kabupaten/kota yang membentuk kelompok sendiri, sesuai dengan langkah analisis data yang dituliskan maka data *outlier* dikeluarkan dan dilakukan pengelompokan ulang.

3.3 Pengelompokan Kabupaten/Kota di Sulawesi Selatan menggunakan Fuzzy C-Means Tanpa Makassar dan Bone

Jumlah kabupaten/kota yang dikelompokkan tanpa memasukkan Makassar dan Bone yakni sebanyak 22 kabupaten/kota. Dalam menentukan jumlah kelompok optimum digunakan metode Pseudo F.

Tabel 5. *Pseudo F Statistics* tanpa Makassar dan Bone

Kelompok	Nilai Pseudo F
2	16,14331
3	15,9761
4	13,86785
5	13,58486
6	16,18428

***bold** : nilai pseudo F tertinggi

Tabel 5 menunjukkan bahwa nilai pseudo F tertinggi antara 2 kelompok hingga 6 kelompok terdapat pada 6 kelompok yakni sebesar 11.07455, sehingga hasil pengelompokan kabupaten/kota di Sulawesi Selatan optimum dengan menggunakan 6 kelompok/*cluster* (c). Selanjutnya dilakukan pengelompokkan dengan metode *Fuzzy C-Means*.

1. Selanjutnya dilakukan pengelompokan dengan menggunakan metode *Fuzzy C-Means* tanpa Makassar dan Bone. Sehingga jumlah kabupaten/kota dan jumlah variabel yang digunakan masing-masing adalah $n = 22$ dan $m = 11$. Kemudian menentukan nilai parameter awal sebagai berikut :
 - a. Jumlah *cluster* (c) : 6
 - b. Pangkat (w) : 2
 - c. Maksimum Iterasi (MaxIter) : 100
 - d. Error terkecil yang diharapkan (ϵ) : 10^{-9}
 - e. Fungsi obyektif awal (P_0) : 0
 - f. Iterasi awal (t) : 1
2. Kemudian membangkitkan bilangan random U_{ik} , dimana $i = 1,2,3,\dots,n$; $k=1,2,\dots,c$;

Berikut adalah matriks partisi awal U yang terbentuk (secara random) :

$$U_{0(22 \times 6)} = \begin{pmatrix} 0,330990252 & 0,20741530 & \dots & 0,087587039 & 0,22004443 \\ 0,001842248 & 0,37917143 & \dots & 0,047356130 & 0,01261982 \\ 0,222681459 & 0,07643545 & \dots & 0,016620881 & 0,26023731 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0,216966095 & 0,18852581 & \dots & 0,195978184 & 0,07113391 \\ 0,052898707 & 0,15511309 & \dots & 0,119477444 & 0,20612260 \\ 0,100688956 & 0,04639261 & \dots & 0,233189318 & 0,28678654 \end{pmatrix}$$

3. Pada tahap selanjutnya akan digunakan Algoritma *Fuzzy C-Means* pada aplikasi RStudio dalam menentukan kelompok tiap obyek.

Output dari *syntax* yang digunakan memberikan informasi diantaranya:

- a. Hasil perhitungan nilai fungsional dan diperoleh bahwa diperlukan sebanyak 98 kali iterasi sebelum memperoleh solusi optimum bagi nilai fungsional $J_m(U, V)$ sebesar 137874,1.
- b. Hasil perhitungan nilai-nilai V_{kj} .

$$V_{(6 \times 11)} = \begin{pmatrix} 95,820 & 333,820 & 237,961 & \dots & 47,137 & 47,918 & 148,569 \\ 92,464 & 528,212 & 387,439 & \dots & 44,775 & 55,074 & 299,075 \\ 58,988 & 418,270 & 505,230 & \dots & 16,550 & 36,875 & 169,626 \\ 92,663 & 411,850 & 258,198 & \dots & 31,326 & 48,720 & 203,954 \\ 82,884 & 380,170 & 300,800 & \dots & 32,174 & 39,513 & 138,886 \\ 139,788 & 636,530 & 257,441 & \dots & 29,748 & 66,692 & 324,706 \end{pmatrix}$$

- c. Hasil perhitungan dari nilai-nilai U_{ik} . Berdasarkan dari nilai-nilai U_{ik} tersebut dapat diperoleh hasil pengelompokan pada Tabel berikut.

Tabel 6 Hasil Pengelompokan Tanpa Makassar dan Bone

Kabupaten/Kota	Kelompok	Jumlah
Takalar, Sinjai, Pangkep, Enrekang	1	4
Maros, Pinrang, Luwu Utara	2	3

Bantaeng, Toraja Utara	3	2
Gowa, Barru, Soppeng, Wajo, Sidrap	4	5
Kep. Selayar, Jeneponto, Luwu, Tana Toraja, Luwu Timur	5	5
Bulukumba, Kota Pare- Pare, Kota Palopo	6	3

4 Kesimpulan

Setelah melakukan analisis dan pembahasan mengenai pengelompokan kabupaten/kota di Sulawesi Selatan berdasarkan jumlah tenaga kesehatan, maka diperoleh kesimpulan sebagai berikut.

1. Penentuan jumlah kluster optimum secara keseluruhan kabupaten/kota didapatkan jumlah kluster terbaik adalah 5 dengan nilai Pseudo F-Statistic sebesar 161,8689. Sementara itu, tanpa Kota Makassar dan Kabupaten Bone didapatkan jumlah kluster terbaik adalah 6 dengan nilai Pseudo F-Statistic sebesar 16,18428.
2. Pada pengelompokan kabupaten/kota di Sulawesi Selatan berdasarkan jumlah tenaga kesehatan dengan menggunakan metode Fuzzy C-Means yang membentuk 5 kelompok sebagai berikut : Kelompok 1 terdiri dari 11 kabupaten/kota meliputi, Kep. Selayar, Jeneponto, Takalar, Gowa, Sinjai, Pangkep, Soppeng, Wajo, Enrekang, Luwu, dan Luwu Timur. Pada Kelompok 2 dan 3 terdiri masing-masing 1 kabupaten/kota. Kelompok 4 dan 5 terdiri dari masing-masing 4 dan 7 kabupaten/kota.
3. Pada pengelompokan kabupaten/kota di Sulawesi Selatan tanpa Kota Makassar dan Kabupaten Bone berdasarkan jumlah tenaga kesehatan dengan menggunakan metode Fuzzy C-Means yang membentuk 6 kelompok, anggota dan karakteristik masing-masing kelompok adalah sebagai berikut: Kelompok 1 terdiri dari 4 kabupaten/kota meliputi, Takalar, Sinjai, Pangkep, dan Enrekang. Kelompok 2 dan 6 terdiri dari masing-masing 3 kabupaten/kota. Kelompok 3 terdapat 2 kabupaten/kota yaitu Bantaeng dan Toraja. Kelompok 4 dan 5 terdiri dari masing-masing 5 kabupaen/kota.

Daftar Pustaka

- [1] Prasetyo, E. *Data Mining: Konsep dan Aplikasi menggunakan MATLAB*. Andi Offset, 2012.
- [2] Kusumadewi, S., & Purnomo, H. *Aplikasi Logika Fuzzy untuk Pendukung Keputusan*. Graha Ilmu, 2010.
- [3] Calinski, T., & Harabasz, J. A Dendrite Method for Cluster Analysis. *Communications in Statistics - Theory and Methods*, 3(1), 1974.

- [4] Orpin, A. R., & Kostylev, V. E. *Toward a Statistically Valid Method of Textural Sea Floor Characterization of Benthic Habitats*. *Marine Geology*, 225(1-4), 2006.
- [5] Kusumadewi, S., Hartati, S., Harjoko, S., & Wardoyo, R. *Fuzzy Multi-attribute Decision Making*. Graha Ilmu, 2006.
- [6] Bezdek, J. C. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, 1981.
- [7] Jain, A. K., Murty, M. N., & Flynn, P. J. Data Clustering: A Review. *ACM Computing Surveys*, 31(3), 1999
- [8] MacQueen, J. Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, 1(14), 1967.
- [9] Johnson, R. A., & Wichern, D. W. *Applied Multivariate Statistical Analysis*. Pearson Prentice Hall, 2007.
- [10] Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. *Probability & Statistics for Engineers & Scientists*. Pearson Education, 2012.
- [11] Rousseeuw, P. J., & Hubert, M. Robust statistics for outlier detection. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1), 2011.
- [12] Arisanti, R., Wahyuni, S., & Hakim, A. Penerapan Metode Fuzzy C-Means untuk Pengelompokan Kabupaten/Kota di Indonesia Berdasarkan Indikator Kesehatan. *Jurnal Biometrika dan Kependudukan*, 8(2), 2019.
- [13] Kementerian Kesehatan Republik Indonesia. Profil Kesehatan Indonesia Tahun 2021. Kemenkes RI, 2022.
- [14] Maulidah, N., & Susanti, Y. Penentuan Jumlah Kluster Optimum Pada Fuzzy C-Means Menggunakan Calinski-Harabasz Index. *Jurnal Statistika*, 12(2), 2020.
- [15] Duda, R. O., Hart, P. E., & Stork, D. G. *Pattern Classification*. John Wiley & Sons, 2001.