

# Penerapan Algoritma Fuzzy C-Means (FCM) dalam Pengelompokan Provinsi di Indonesia berdasarkan Indikator Penyakit Menular Manusia

FAHRIZA NOVIANTI<sup>1</sup>, YASIRAH R. A. YASMIN<sup>2</sup>, DIAN C. R. NOVITASARI<sup>3</sup>

<sup>1,2,3</sup> Prodi Matematika, UIN Sunan Ampel Surabaya  
Email : diancrini@uinsby.ac.id

## ABSTRAK

*Penyakit menular merupakan penyakit yang dapat tersebar dari manusia yang sakit ke manusia yang sehat diantaranya adalah HIV, AIDS, TBC dan demam berdarah. Tujuan dari penelitian ini ialah mengelompokkan provinsi di Indonesia berdasarkan indikator penyakit menular dan kepadatan penduduk sebagaiantisipasi pemerintah dalam melakukan pemerataan fasilitas kesehatan. Metode clustering yang digunakan pada penelitian ini adalah Fuzzy C-Means (FCM). Hasil pengelompokan provinsi di Indonesia pada tahun 2019 dan 2020 pengelompokan provinsi di Indonesia berdasarkan indikator penyakit menular dan kepadatan penduduk dibagi menjadi 3 Cluster (tinggi, sedang, rendah) dengan nilai Silhouette Coefficient pada tahun 2019 sebesar 0.9151 dan pada tahun 2020 sebesar 0.91899 yang menunjukkan kekuatan cluster mempunyai struktur kuat. Hasil analisis cluster menunjukkan pada tahun 2019 hanya Jakarta yang mempunyai tingkat penyebaran tinggi, meningkat di tahun 2020 di Jakarta yaitu Jawa Timur dan Bali, sedangkan Papua, Jawa Tengah dan Jawa Barat berada pada tingkat penyebaran sedang dan provinsi lainnya dengan tingkat penyebaran rendah.*

**Kata kunci:** Fuzzy C-Means, Indonesia, Pengelompokan, Penyakit Menular Manusia

## ABSTRACT

*Infectious diseases can spread from sick to healthy people, including HIV, AIDS, tuberculosis, and dengue fever. This study aims to group provinces in Indonesia based on indicators of infectious disease and population density as the government's anticipation of distributing health facilities. The clustering method used in this research is Fuzzy C-Means (FCM). The results of the grouping of provinces in Indonesia in 2019 and 2020, the grouping of provinces in Indonesia based on indicators of infectious diseases and population density is divided into 3 Clusters (high, medium, low) with a Silhouette Coefficient value in 2019 of 0.9151 and in 2020 of 0.91899 which shows strength cluster has a strong structure. The cluster analysis results show that in 2019 only Jakarta has a high distribution rate. In 2020, the high distribution rate increased in East Java and Bali, while Papua, Central Java, and West Java are at a moderate distribution level and other provinces are still safe with a low level of distribution.*

**Keywords:** Fuzzy C-Means, Indonesia, Clustering, Human Infectious Disease

## 1. PENDAHULUAN

Penyakit menular merupakan penyakit yang dapat ditularkan (berpindah dari orang yang satu ke orang yang lain, baik secara langsung maupun melalui perantara). Penyakit ini disebabkan oleh perpindahan *infectious agent*/produk toksinnya dari seseorang/*reservoir* ke orang lain/*susceptable host* (Irwan, 2017). Jenis penyakit menular diantaranya yaitu HIV/AIDS, TBC dan demam berdarah. Penyakit menular menjadi salah satu penyebab utama kematian di Dunia. Indonesia sebagai negara kepulauan yang memiliki banyak penduduk, menyebabkan Indonesia menanggung beban berlebih dalam penanggulangan penyakit (Andika et al., 2020).

Penyebaran penyakit disebabkan oleh beberapa faktor diantaranya yaitu faktor daya tahan tubuh, faktor lingkungan, letak geografis, kondisi demografi, dan jenis penyakit. Berdasarkan data berupa jenis penyakit serta kepadatan penduduk maka dapat dilakukan pengelompokan dengan melakukan analisis klaster. Dalam analisis klaster, pengelompokan dilakukan berdasarkan kesaamaan/kemiripan antar objek (Dwitiyanti, Selvia & Andrari, 2019). Algoritma yang sering digunakan dalam *clustering* adalah *K-Means* dan *Fuzzy C-Means* (FCM). *Fuzzy clustering* melakukan pembobotan terhadap tingkat keanggotaan himpunan *fuzzy*-nya. FCM merupakan metode pengelompokan yang memungkinkan satu bagian data memiliki dua atau lebih kelompok. Dengan FCM, objek akan cenderung dibagi menjadi anggota suatu kelompok sesuai derajat keanggotaannya. Sedangkan *K-means* (KM) merupakan algoritma *clustering* yang bekerja dengan cara membagi data ke dalam partisi data (Tamaela et al., 2017).

Pada penelitian sebelumnya, (Firdaus et al., 2021) membandingkan metode *Fuzzy C-Means* dan *K-Means* dalam melakukan pemetaan daerah rawan kriminalitas di Kota Semarang diperoleh hasil bahwa metode *Fuzzy C-Means* lebih unggul karena hasil uji validasi *Partition Coefficient Index* (PCI) dari metode *Fuzzy C-Means* lebih mendekati 1 yaitu sebesar 0,818 sedangkan hasil uji pengolahan *Silhouette Indeks* (SI) dari metode *K-Means* sebesar 0,569. Selanjutnya, (Pamungkas et al., 2021) melakukan penelitian yang membandingkan metode *Fuzzy C-Means* dan *K-Means* untuk mengelompokkan tingkat buta huruf berdasarkan provinsi di Indonesia diperoleh hasil bahwa metode *Fuzzy C-Means* lebih unggul karena hasil uji validasi *Partition Coefficient Index* (PCI) dari metode *Fuzzy C-Means* lebih mendekati 1 yaitu sebesar 0,7684 sedangkan hasil uji pengolahan *Silhouette Indeks* (SI) dari metode *K-Means* sebesar 0,4966. (Bastian, Sujadi & Febrianto, 2018) melakukan analisis cluster penyakit menular di Kabupaten Majalengka menggunakan metode *K-Means*. Penelitian sebelumnya telah dilakukan pengelompokan wilayah dengan metode *K-Means*, dimana pada penentuan objek dalam cluster tertentu didasarkan pada rata-rata terdekat. Sedangkan pada penelitian ini, digunakan metode *Fuzzy C-Means* dalam pengelompokan provinsi di Indonesia dengan menentukan objek *cluster* berdasarkan pada pusat *cluster* dan derajat keanggotaan pada setiap klaster.

Penelitian ini bertujuan untuk melakukan pengelompokan provinsi di Indonesia menurut indikator penyakit menular menggunakan metode *Fuzzy C-Means Clustering*. Diharapkan dengan menggunakan metode ini akan diperoleh nilai keakurasian yang lebih baik

## 2. METODOLOGI

### 2.1. Pengumpulan Data

Data yang digunakan berupa data HIV, AIDS, Tuberculosis, DBD serta kepadatan penduduk. Data penyakit menular HIV, AIDS, Tuberculosis, dan DBD didapatkan dari website resmi

pusat data informasi kementerian kesehatan Indonesia (<https://pusdatin.kemkes.go.id>) Indonesia (<https://www.bps.go.id>). Sampel data tersebut dapat dilihat pada Tabel 1 dan Tabel 2.

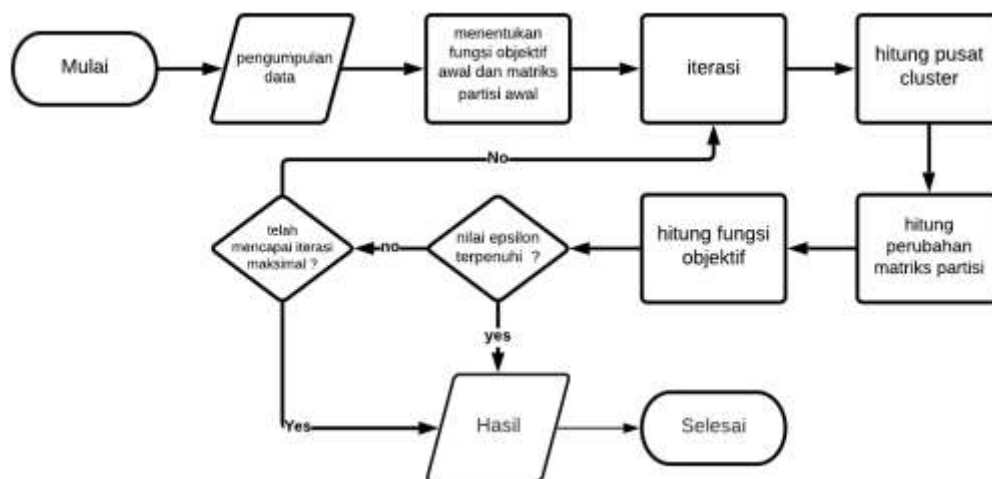
**Tabel 1. Contoh Sampel Jumlah Penderita Penyakit Menular di Indonesia Tahun 2019**

Provinsi	HIV	AIDS	Tuberculosis	DBD	Kepadatan Penduduk
ACEH	174	49	123	45	91,7299
BANGKA BELITUNG	234	44	107	68	88,3521
SUMATERA SELATAN	601	207	116	34	92,7719
JAWA TIMUR	8935	958	110	47	831,4855
KALIMANTAN TENGAH	222	53	87	60	17,25529

**Tabel 2. Contoh Sampel Jumlah Penderita Penyakit Menular di Indonesia Tahun 2020**

Provinsi	HIV	AIDS	Tuberculosis	DBD	Kepadatan Penduduk
ACEH	831	584	123	527	92,96880392
BANGKA BELITUNG	1659	563	107	695	89,49066187
SUMATERA SELATAN	4233	1410	116	1749	93,90295683
JAWA TIMUR	60417	21016	110	5104	835,9018614
KALIMANTAN TENGAH	1129	348	87	603	17,49297526

Data pada Tabel 1 dan Tabel 2 akan diolah sesuai dengan metode Fuzzy C-Means. Proses singkat metode tersebut dipresentasikan dalam diagram alir pada Gambar 1.



**Gambar 1. Diagram Alir Proses *Clustering* Penyakit Menular menggunakan Fuzzy C-Means**

## 2.2. Fuzzy C-Means

Fuzzy C-Means (FCM) pertama kali di perkenalkan oleh Jim Bezdek pada tahun 1981. FCM merupakan suatu teknik pengelompokan data yang keberadaan setiap titik data dalam suatu *cluster* di tentukan oleh derajat keanggotaan (Mulyaningsih, 2019). Output dari FCM berupa deretan pusat clusster dan beberapa derajat keanggotaan untuk setiap titik data. Informasi ini dapat digunakan untuk membangun suatu Fuzzy inference system (Puspitasari, Widians & Pohny, 2018).

Pada kondisi awal, pusat *cluster* masih belum akurat sehingga dibutuhkan perbaikan pusat *cluster* secara berulang hingga berada pada titik yang tepat sehingga setiap data akan memiliki derajat keanggotaan untuk setiap *cluster* (Afifah, Rini & Lubab, 2016).

Kelebihan metode Fuzzy C-Means adalah penempatan pusat *cluster* yang tepat dibandingkan dengan metode *cluster* yang lain (Ais et al., 2022). Algoritma dari Fuzzy C-Means yaitu,

1. Input data yang akan di *cluster*  $X$ , dalam bentuk matriks berukuran  $n \times m$  ( $n$  = jumlah sampel data,  $m$  = atribut setiap data).  $X_{ij}$  = data sampel ke- $i$  ( $i = 1, 2, 3, \dots, n$ ), atribut ke- $j$  ( $j = 1, 2, \dots, m$ ).
2. Parameter yang dibutuhkan, yaitu:
  - a. Jumlah klaster (c)
  - b. Pangkat/bobot (w)
  - c. Maksimum iterasi (MaxIter)
  - d. Error terkecil ( $\varepsilon$ )
  - e. Fungsi objektif ( $P_o = 0$ )
  - f. Iterasi awal ( $t = 1$ )
3. Bentuk bilangan random  $\mu_{ik}$  ( $i = 1, 2, \dots, n$  dan  $k = 1, 2, \dots, c$ ) yang merupakan elemen-elemen matriks partisi awal U. kemudian, hitung jumlah setiap kolom dengan persamaan (1).

$$Q_i = \sum_{k=1}^c \mu_{ik} \quad (1)$$

Dimana :

$\mu_{ik}$  = derajat eanggotaan

$Q_i$  = Jumlah nilai derajat keanggotaan perkolom = 1 dengan  $i=1, 2, \dots, n$ ;

Tentukan nilai matriks partisi awalmenggunakan persamaan (2).

$$\mu_{ik} = \frac{\mu_{ik}}{Q_i} \quad (2)$$

4. Hitung pusat *cluster* ke- $k$  :  $V_{kj}$ , dimana  $k = 1, 2, \dots, c$  dan  $j = 1, 2, \dots, n$  menggunakan persamaan (3).

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \quad (3)$$

Dimana :

$V$  = pusat klaster

$X_i$  = parameter ke- $i$

5. Hitung fungsi objektif pada iterasi ke- $t$  ( $P_t$ ) dengan persamaan (4).

$$P_t = \sum_{i=1}^n \sum_{k=1}^c ([\sum_{j=1}^m (X_{ij} - V_{kj})^2] (\mu_{ik})^w) \quad (4)$$

Dimana  $P_t$  adalah nilai fungsi objektif iterasi ke- $t$  .

6. Hitung perubahan matriks partisi menggunakan persamaan (5).

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{\frac{-1}{w-1}}}{\sum_{k=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{\frac{-1}{w-1}}} \quad (5)$$

Dimana  $i = 1, 2, \dots, n$  dan  $k = 1, 2, \dots, c$ .

7. Cek kondisi berhenti, dengan ketentuan yaitu :

- $(P_t - (P_t - 1)) < \varepsilon$  atau  $(t > MaxIter)$  maka berhenti
- jika tidak :  $t = t + 1$  maka ulangi langkah ke-4.

### 2.3. Uji Kualitas menggunakan *Sillhoutte Coefficient*

Untuk mengetahui kualitas suatu *cluster* perlu dilakukan suatu pengujian. Salah satu caranya yaitu dengan pengujian *Sillhouette Coefficient* (Swindiarto, 2018). Pengujian ini dilakukan dengan membandingkan jarak data pada *cluster* yang sama dengan jarak data di *cluster* lainnya (Kaufman & Rousseeuw, 2005). Teknik ini memberikan representasi singkat tentang seberapa baik objek pada klasternya. *Silhouette Coefficient* mengevaluasi objek pada setiap *cluster* dengan memperhatikan nilai *silhouettenya*. Nilai *silhouette Coefficient* tiap objek dapat bervariasi antara -1 sampai dengan 1 (Ulinuha, 2020). Berdasarkan nilai yang didapatkan dari *Silhouette Coefficient*, hasilnya bernilai positif saat jarak data ke *cluster* yang lain lebih besar dibandingkan jarak data ke *cluster* data itu sendiri, dan sebaliknya akan bernilai negatif saat jarak data ke *cluster* yang lain lebih dekat dibandingkan jarak data ke *cluster* data itu sendiri (Fasich, 2017). Perhitungan *Silhouette Coefficient* ditunjukkan pada persamaan (6).

$$s_i = \frac{(b_i - a_i)}{\max(b_i, a_i)} \quad (6)$$

$s_i$  = Silhouette Coefficient

$a_i$  = rata-rata jarak dari objek  $i$  dengan seluruh objek yang berada pada *cluster* yang sama

$b_i$  = nilai terkecil dari rata-rata jarak objek  $i$  dengan objek lain pada *cluster* yang berbeda

**Tabel 3. Penjelasan Subjektif *Silhouette Coefficient***

Nilai <i>Silhouette Coefficient</i>	Definisi
0.71 – 1.00	Struktur yang terbentuk sudah kuat
0.51 – 0.70	Struktur yang terbentuk masuk akal (struktur masih bisa berubah)
0.26 – 0.50	Struktur yang terbentuk lemah. Coba metode tambahan untuk data yang digunakan
$\leq 0.25$	Struktur data tidak bisa ditemukan

Sumber : *Finding Groups in Data: An Introduction to Cluster Analysis*

Tabel 3 menunjukkan ketika nilai rata-rata *Silhouette Coefficient* mendekati 1 maka struktur dari *cluster* tersebut sudah kuat sehingga hasil hasil klastering sudah tepat, sedangkan ketika nilai mendekati 0 maka kemungkinan besar terjadi misclustering atau data yang tidak dapat dikelompokkan.

## 3. HASIL PENELITIAN DAN ANALISIS

Langkah pertama dalam pengelompokan data menggunakan FCM ialah menentukan komponen-komponen yang diperlukan. Komponen tersebut dijabarkan pada poin 3.1.

### 3.1. Komponen dalam FCM

- Jumlah klaster : 3
- Pangkat pembobot : 2
- Maksimum iterasi : 100
- Error terkecil yang diharapkan :  $1 \times 10^{-9}$
- fungsi objektif awal : 0
- Iterasi awal : 1

Langkah selanjutnya adalah melakukan pengelompokan wilayah pada tiap tahunnya, yaitu tahun 2019 dan 2020. Pengelompokan pada tahun 2019 dijabarkan pada point 4.2. dan pengelompokan pada tahun 2020 dijabarkan pada poin 4.3.

### 3.2. Penyebaran Penyakit Menular Tahun 2019

Dalam menentukan keoptimalan suatu klaster, perlu dilakukan suatu validasi. Pada penelitian ini, metode yang digunakan dalam melakukan validasi ialah *Silhouette Coefficient*. Jika nilai yang didapatkan mendekati 1 maka hasil pengelompokan akan semakin bagus.

**Tabel 4** *Validity Index Cluster*

Jumlah <i>Cluster</i>	<i>Fuzzy Silhouette Index</i>
2	0.8799
3	0.9151
4	0.8305

Berdasarkan Tabel 4. pada kasus penyebaran penyakit menular tahun 2019 nilai paling optimum berada pada jumlah *cluster* sebanyak 3, dengan nilai Silhouette Coefficient sebesar 0,9151 dan iterasi berhenti pada iterasi ke-34 dengan pusat titik *cluster* ditampilkan pada Tabel 5. Tabel berwarna merah menunjukkan titik pusat *cluster* dengan penyebaran tinggi, warna kuning menunjukkan titik pusat *cluster* dengan penyebaran sedang, dan warna hijau menunjukkan titik pusat *cluster* dengan rendah.

**Tabel 5. Titik Pusat Cluster Penyebaran Penyakit Tahun 2019**

	HIV	AIDS	DBD	Tuberculosis	Kepadatan Penduduk
<i>cluster 1</i>	616.5731	91.52544	71.91753	121.8357	191.7841
<i>cluster 2</i>	6700.0546	585.25226	82.95046 2	35.8729	15798.6077
<i>cluster 3</i>	6332.1600	948.46389	38.67668	143.9407	976.2155

Berdasarkan Tabel 5. Dari pusat *cluster* masing-masing variabel dapat diketahui bahwa *cluster 1* dikategorikan sebagai *cluster* yang memiliki tingkat penyebaran penyakit rendah, *cluster 2* merupakan tingkat penyebaran penyakit tinggi, sedangkan *cluster 3* merupakan tingkat penyebaran penyakit sedang.

Setelah menentukan pusat cluster, langkah selanjutnya adalah menentukan derajat keanggotaan pada tiap provinsi. Hasil dari derajat keanggotaan tersebut dapat dilihat pada Tabel 6.

**Tabel 6. Derajat Keanggotaan tiap Provinsi Tahun 2019**

Provinsi	<i>cluster 1</i>	<i>cluster 2</i>	<i>cluster 3</i>	<i>Tingkat Penyebaran</i>
ACEH	0.994042	0.000715	0.005243	Rendah
SUMATERA UTARA	0.819269	0.010706	0.170025	Rendah
SUMATERA BARAT	0.998326	0.000186	0.001488	Rendah
RIAU	0.999138	0.000095	0.000767	Rendah
KEPULAUAN RIAU	0.993988	0.000605	0.005407	Rendah
JAMBI	0.992731	0.000882	0.006387	Rendah
SUMATERA SELATAN	0.999186	0.000087	0.000727	Rendah

Provinsi	<i>cluster 1</i>	<i>cluster 2</i>	<i>cluster 3</i>	<i>Tingkat Penyebaran</i>
BANGKA BELITUNG	0.995357	0.000550	0.004094	Rendah
BENGKULU	0.994148	0.000702	0.005150	Rendah
LAMPUNG	0.999739	0.000029	0.000232	Rendah
BANTEN	0.899137	0.008877	0.091986	Rendah
DKI JAKARTA	0.000002	0.999997	0.000002	Tinggi
JAWA BARAT	0.020144	0.003016	0.976841	Sedang
JAWA TENGAH	0.033765	0.004292	0.961943	Sedang
DI YOGYAKARTA	0.963044	0.004266	0.032690	Rendah
JAWA TIMUR	0.085777	0.026335	0.887888	Sedang
KALIMANTAN BARAT	0.998931	0.000112	0.000957	Rendah
KALIMANTAN TENGAH	0.994543	0.000645	0.004812	Rendah
KALIMANTAN SELATAN	0.998744	0.000141	0.001115	Rendah
KALIMANTAN TIMUR	0.979176	0.001830	0.018994	Rendah
KALIMANTAN UTARA	0.993235	0.000804	0.005961	Rendah
SULAWESI UTARA	0.999654	0.000037	0.000310	Rendah
GORONTALO	0.990527	0.001174	0.008299	Rendah
SULAWESI TENGAH	0.995574	0.000520	0.003906	Rendah
SULAWESI SELATAN	0.963270	0.003050	0.033680	Rendah
SULAWESI BARAT	0.991134	0.001092	0.007774	Rendah
SULAWESI TENGGARA	0.994063	0.000711	0.005226	Rendah
BALI	0.835622	0.010602	0.153776	Rendah
NUSA TENGGARA BARAT	0.995870	0.000492	0.003638	Rendah
NUSA TENGGARA TIMUR	0.998162	0.000188	0.001650	Rendah
MALUKU	0.996982	0.000346	0.002671	Rendah
MALUKU UTARA	0.989041	0.001365	0.009594	Rendah
PAPUA	0.405654	0.017015	0.577331	Sedang
PAPUA BARAT	0.998315	0.000177	0.001508	Rendah

Dari hasil Tabel 6. diperoleh bahwa dalam *cluster 1* terdiri dari 29 provinsi, *cluster 2* hanya terdiri dari 1 provinsi yakni DKI Jakarta, dan *cluster 3* terdiri dari 4 Provinsi yaitu Jawa Barat, Jawa tengah, Jawa Timur, dan Papua.

### 3.3. Penyebaran Penyakit Menular Tahun 2020

Agar penyebaran penyakit menular pada tahun 2020 dapat dibandingkan dengan penyebaran penyakit menular pada tahun 2019, maka jumlah cluster yang digunakan harus sama yakni berjumlah 3 *cluster*. Pada Tabel 7 ditunjukkan nilai *Silhouette Coefficient* dengan jumlah *cluster 3* sebesar 0.91899. Dimana nilai tersebut masuk dalam hasil dengan struktur yang kuat seperti yang tertera pada Tabel 3.

**Tabel 7. *Validity Index Cluster***

Jumlah <i>Cluster</i>	<i>Fuzzy Silhouette Index</i>
2	0.93497
3	0.91899
4	0.70404

Pada kasus ini terdapat 3 *cluster* dengan perhitungan berhenti di iterasi ke-54. Titik pusat cluster dapat dilihat pada Tabel 8. Tabel berwarna merah menunjukkan titik pusat cluster dengan penyebaran tinggi, warna kuning menunjukkan titik pusat cluster dengan penyebaran sedang, dan warna hijau menunjukkan titik pusat cluster dengan rendah.

**Tabel 8. Titik Pusat Cluster Penyebaran Penyakit Tahun 2020**

	HIV	AIDS	DBD	Tuberculosis	Kepadatan Penduduk
<i>cluster 1</i>	4199.767	1405.509	1362.158	121.7987	187.5745
<i>cluster 2</i>	63517.631	15376.973	4392.922	177.2904	8569.0908
<i>cluster 3</i>	34806.456	12781.590	4609.676	149.7042	811.2939

Berdasarkan Tabel 8. Dari pusat cluster masing masing variabel dapat diketahui bahwa *cluster 1* dikategorikan sebagai *cluster* yang memiliki tingkat penyebaran penyakit rendah, *cluster 2* merupakan tingkat penyebaran penyakit tinggi, sedangkan *cluster 3* merupakan tingkat penyebaran penyakit sedang.

Setelah menentukan pusat cluster, langkah selanjutnya adalah menentukan derajat keanggotaan pada tiap provinsi. Hasil dari derajat keanggotaan tersebut dapat dilihat pada Tabel 9.

**Tabel 9. Derajat Keanggotaan tiap Provinsi Tahun 2020**

Provinsi	<i>cluster 1</i>	<i>cluster 2</i>	<i>cluster 3</i>	<i>Tingkat Penyebaran</i>
ACEH	0.987511	0.002968	0.009521	Rendah
SUMATERA UTARA	0.481094	0.063986	0.454920	Rendah
SUMATERA BARAT	0.998553	0.000320	0.001127	Rendah
RIAU	0.993448	0.001375	0.005177	Rendah
KEPULAUAN RIAU	0.939825	0.011511	0.048663	Rendah
JAMBI	0.994188	0.001348	0.004464	Rendah
SUMATERA SELATAN	0.999809	0.000042	0.000149	Rendah
BANGKA BELITUNG	0.992188	0.001830	0.005982	Rendah
BENGKULU	0.988088	0.002827	0.009085	Rendah
LAMPUNG	0.984589	0.003426	0.011985	Rendah
BANTEN	0.924593	0.014052	0.061355	Rendah
DKI JAKARTA	0.020320	0.912790	0.066889	Tinggi
JAWA BARAT	0.060935	0.178311	0.760754	Sedang
JAWA TENGAH	0.004569	0.006521	0.988910	Sedang
DI YOGYAKARTA	0.988194	0.002476	0.009330	Rendah
JAWA TIMUR	0.024446	0.855364	0.120190	Tinggi

Penerapan Algoritma Fuzzy C-Means (FCM) dalam Pengelompokan Provinsi di Indonesia berdasarkan Indikator Penyakit Menular Manusia

Provinsi	<i>cluster 1</i>	<i>cluster 2</i>	<i>cluster 3</i>	<i>Tingkat Penyebaran</i>
KALIMANTAN BARAT	0.972506	0.005482	0.022012	Rendah
KALIMANTAN TENGAH	0.988927	0.002623	0.008450	Rendah
KALIMANTAN SELATAN	0.997037	0.000679	0.002284	Rendah
KALIMANTAN TIMUR	0.977039	0.004627	0.018334	Rendah
KALIMANTAN UTARA	0.986907	0.003113	0.009980	Rendah
SULAWESI UTARA	0.997499	0.000539	0.001962	Rendah
GORONTALO	0.983718	0.003914	0.012368	Rendah
SULAWESI TENGAH	0.992773	0.001686	0.005541	Rendah
SULAWESI SELATAN	0.914161	0.015586	0.070253	Rendah
SULAWESI BARAT	0.982800	0.004147	0.013053	Rendah
SULAWESI TENGGARA	0.989581	0.002461	0.007958	Rendah
BALI	0.280519	0.068077	0.651404	Tinggi
NUSA TENGGARA BARAT	0.988405	0.002675	0.008920	Rendah
NUSA TENGGARA TIMUR	0.976393	0.005017	0.018590	Rendah
MALUKU	0.997301	0.000601	0.002098	Rendah
MALUKU UTARA	0.989129	0.002567	0.008303	Rendah
PAPUA	0.072011	0.140288	0.787701	Sedang
PAPUA BARAT	0.993938	0.001293	0.004769	Rendah

Dari hasil Tabel 9. diperoleh bahwa dalam *cluster 1* terdiri dari 28 provinsi, *cluster 2* terdiri dari 3 *cluster* yaitu DKI Jakarta, Jawa Timur dan Bali. *Cluster 3* terdiri dari 3 provinsi yaitu Jawa Barat, Jawa Tengah, dan Papua.

Hasil *clustering* pada tahun 2019 dan 2020, divisualisasikan pada peta penyebaran penyakit menular tahun 2019 yang dapat dilihat pada Gambar 2 dan peta penyebaran penyakit menular tahun 2020 dapat dilihat pada Gambar 3.



**Gambar 2. Peta penyebaran Penyakit Menular tahun 2019**



**Gambar 3. Peta penyebaran Penyakit Menular tahun 2020**

Dari Gambar 2 dan Gambar 3 dapat dijelaskan bahwa hasil clustering menunjukkan pada provinsi Jawa Timur yang pada tahun 2019 berada pada *cluster* dengan tingkat penyebaran penyakit menular sedang dan Provinsi Bali yang pada tahun 2019 berada pada *cluster* dengan tingkat penyebaran penyakit menular rendah keduanya mengalami perubahan pada tahun 2020, yaitu berada pada *cluster* dengan penyebaran penyakit menular tingkat tinggi bersama dengan DKI Jakarta. Perubahan posisi *cluster* tersebut dapat disebabkan oleh perubahan jumlah penduduk dan perubahan jumlah kasus penyakit menular pada provinsi tersebut.

#### 4. KESIMPULAN

Metode Fuzzy C-Means (FCM) dapat diterapkan untuk pengelompokan wilayah Indonesia berdasarkan indeks penyakit menular pada manusia. Cluster yang dihasilkan untuk data pada kedua tahun yakni 2019 dan 2020 ialah sebanyak 3 *cluster*. Penentuan *cluster* tersebut dilakukan berdasarkan pengujian validasi dengan menggunakan *Shillouette Coefficient*. Pada data tahun 2019 nilai validasi sebesar 0.9151 dan pada tahun 2020 didapatkan nilai validasi sebesar 0.91899.

Pada penelitian ini 3 cluster terdiri dari cluster tingkat rendah, sedang dan tinggi. Pada tahun 2019, DKI Jakarta merupakan satu-satunya provinsi yang memiliki tingkat penyebaran penyakit menular yang tinggi, selain itu provinsi Jawa Timur, Jawa Barat, Jawa Tengah dan Papua menduduki provinsi dengan penyebaran penyakit menular tingkat sedang. Pada tahun 2020, terjadi perubahan di Jawa Timur dan Bali menduduki penyebaran penyakit menular tingkat tinggi bersamaan dengan DKI Jakarta yang semenjak tahun 2019 berada dalam kategori tinggi. Untuk penelitian selanjutnya diharapkan dapat mengimplementasikan metode-metode *clustering* yang memiliki kinerja lebih baik, seperti K-Medoids (**Azhari et al., 2021**).

#### REFERENSI

Afifah, N., Rini, D. C., & Lubab, A. (2016). Pengklasteran Lahan Sawah Di Indonesia Sebagai

- Evaluasi Ketersediaan Produksi Pangan Menggunakan Fuzzy C-Means. *Jurnal Matematika "MANTIK,"* 2(1), 40. <https://doi.org/10.15642/mantik.2016.2.1.40-45>
- Ais, C., Hamid, A., & Novitasari, D. C. R. (2022). Analysis of Livestock Meat Production in Indonesia Using Fuzzy C-Means Clustering. *Jurnal Ilmu Komputer Dan Informasi*, 15(1), 1–8.
- Andika, F., Safira, A., Mustina, N., & Marniati. (2020). Edukasi Tentang Pemberantasan Penyakit Menular Pada Siswa di SMA Negeri 5 Kota Banda Aceh. *Jurnal Pengabdian Masyarakat*, 2(1), 29–33.
- Azhari, S. M., Pudjiantoro, T. H., & Santikarama, I. (2021). Klasterisasi Outlet Berdasarkan Data Penjualan Dengan Menggunakan Algoritma K-Medoids. *JUMANJI (Jurnal Masyarakat Informatika Unjani)*, 5(2), 69. <https://doi.org/10.26874/jumanji.v5i2.93>
- Bastian, A., Sujadi, H., & Febrianto, G. (2018). Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka). 1, 26–32.
- Dwitiyanti, N., Selvia, N., & Andrari, F. R. (2019). Penerapan Fuzzy C-Means Cluster dalam Pengelompokan Provinsi Indonesia Menurut Indikator Kesejahteraan Rakyat. *Faktor Exacta*, 12(3), 201–209.
- Fasich, D. A. (2017). *Klastering Emosi Berdasarkan Gelombang Otak Sinyal EEG Menggunakan Fuzzy C-Means Clustering*. Institut Teknologi Sepuluh November.
- Firdaus, H. S., Nugraha, A. L., Sasmito, B., Awaluddin, M., & Nanda, C. A. (2021). Perbandingan Metode Fuzzy C-Means dan K-Means untuk Pemetaan Daerah Rawan Kriminalitas di Kota Semarang. 04(01), 58–64.
- Irwan. (2017). Epidemiologi Penyakit Menular. In *CV. Absolute Media* (Vol. 109, Issue 1).
- Kaufman, L., & Rousseeuw, P. J. (2005). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons.
- Mulyaningsih, W. S. (2019). *Implementasi Fuzzy C-Means Dan Fuzzy Possibilistic C-Means Untuk Pengelompokan Kabupaten/Kota Di Provinsi Banten*.
- Pamungkas, M. A., Oktavianto, H., & Umilasari, R. (2021). Perbandingan Fuzzy C-Means Dan K-Means Untuk Mengelompokkan Tingkat Buta Huruf Berdasarkan Provinsi Di Indonesia.
- Puspitasari, N., Widians, J. A., & pohny. (2018). a Clustering of Generative and Infectious Diseases Using Fuzzy C-Means. *ITSMART: Jurnal Ilmiah Teknologi Dan Informasi*, 7(1), 22–28.
- Swindiarto, V. T. P. (2018). Integration of Fuzzy C-Means Clustering and TOPSIS ( FCM-TOPSIS ) with Silhouette Analysis for Multi Criteria Parameter Data. *2018 International Seminar on Application for Technology of Information and Communication*, 463–468.
- Tamaela, J., Sedyono, E., & Setiawan, A. (2017). Cluster Analysis Menggunakan Algoritma Fuzzy C-means dan K-means Untuk Klasterisasi dan Pemetaan Lahan Pertanian di Minahasa Tenggara. *Jurnal Buana Informatika*, 8(3), 151–160. <https://doi.org/10.24002/jbi.v8i3.1317>
- Ulinuha, N. (2020). Provincial Clustering in Indonesia Based on Plantation Production Using Fuzzy C-Means. *Jurnal Ilmiah Teknologi Dan Informasi*, 9(1), 8–12.