



**INTEGRASI *CLUSTER VALIDITY INDEX* DAN ALGORITMA *FUZZY C-MEANS* UNTUK PEMETAAN KABUPATEN/KOTA BERDASARKAN FAKTOR STUNTING BALITA DI PROVINSI JAWA TENGAH**

**JURNAL ILMIAH**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Statistika**



**Oleh**

**ATIKAH ALIFAH**

**B2A219013**

**PROGRAM STUDI S1 STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS MUHAMMADIYAH SEMARANG**

**2021**

## PERSETUJUAN PEMBIMBING

Skripsi dengan Judul “*Integrasi Cluster Validity Index dan Algoritma Fuzzy C-Means* untuk pemetaan Kabupaten/Kota berdasarkan faktor Stunting balita di Provinsi Jawa Tengah” yang disusun oleh:

Nama : Atikah Alifah

NIM : B2A219013

Program Studi : S1 Statistika

Telah disetujui oleh Dosen Pembimbing pada tanggal 08 September 2021.



Pembimbing Utama

Pembimbing Pendamping

Dr. Rochdj Wasono, M.Si.  
NIK. 28.6.1026.119

Indah Manfaati Nur, M.Si.  
NIK.28.6.1026.221

## SURAT PERNYATAAN PUBLIKASI KARYA ILMIAH

Nama : Atikah Alifah  
NIM : B2A219013  
Fakultas/Jurusan : S1 Statistika  
Jenis Penelitian : Skripsi  
Judul : Integrasi *Cluster Validity Index* dan Algoritma *Fuzzy C-Means* untuk pemetaan Kabupaten/Kota berdasarkan faktor Stunting balita di Provinsi Jawa Tengah  
Email : atikahfajri@gmail.com

Dengan ini menyatakan bahwa saya menyetujui untuk:

1. Memberikan hak bebas royalti kepada perpustakaan Unimus atas penulisan karya ilmiah saya, demi pengembangan ilmu pengetahuan.
2. Memberikan hak menyimpan, mengalih mediakan atau mengalih formatkan, mengelola dalam bentuk pengakalan data (*database*), mendistribusikannya, serta menampilkannya dalam bentuk *softcopy* untuk kepentingan akademis kepada perpustakaan Unimus, tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis atau pencipta.
3. Bersedia dan menjamin untuk menanggung secara pribadi tanpa melibatkan pihak perpustakaan Unimus, dari semua bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini.

Demikian pernyataan ini saya buat dengan sesungguhnya dan semoga dapat digunakan sebagaimana mestinya.

Semarang, 23 September 2021

Yang Menyatakan,



Atikah Alifah  
NIM. B2A219013

Atikah pernyataan  
keadilan961.jpg  
Type: JPEG Image  
Size: 206 KB  
Dimension: 1654 x

# INTEGRASI *CLUSTER VALIDITY INDEX* DAN ALGORITMA *FUZZY C-MEANS* UNTUK PEMETAAN KABUPATEN/KOTA BERDASARKAN FAKTOR STUNTING BALITA DI PROVINSI JAWA TENGAH

**Atikah Alifah, Rochdi Wasono, Indah Manfaati Nur**

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Muhammadiyah Semarang

e-mail: [atikahfajri18@gmail.com](mailto:atikahfajri18@gmail.com)

## ABSTRAK

*Stunting* masih menjadi tantangan sendiri di Indonesia, termasuk di Jawa Tengah dimana prevalensi *stunting* berada atas rata-rata prevalensi *stunting* nasional. Untuk pengelompokkan faktor *stunting* diusulkan menggunakan Algoritma Fuzzy C-Means. Metode ini memiliki kelebihan dalam penempatan pusat kluster yang lebih tepat dibandingkan dengan metode lain, namun jumlah kluster dalam Algoritma Fuzzy C-Means harus ditentukan dahulu sebelum melakukan penelitian. Untuk menentukan jumlah kluster optimal pada Algoritma Fuzzy C-Means, pada penelitian ini diusulkan metode *Cluster Validity Index* (CVI). CVI adalah sebuah ukuran validitas untuk menemukan jumlah kluster optimal yang sepenuhnya dapat menjelaskan struktur data dengan cara mengukur derajat kekompakan dalam suatu kluster dan separasi anatar kluster. Pada penelitian ini menggunakan *Partition Entropy Index* (PEI), *Partition Coefficient Index* (PCI), *Fukuyama Sugeno Index* (FSI), *Xie Beni Index* (XBI), *Modification Partition Coefficient Index* (MPCI), *Partition Coefficient And Exponential Separation Index* (PCAESI). Hasil penelitian menunjukkan jumlah kluster optimal berdasarkan CVI adalah 2 kluster. Kluster pertama terdiri dari 17 Kabupaten/Kota dan kluster kedua terdiri dari 18 Kabupaten/Kota.

**Kata Kunci :** *Cluster Validity Index, Clustering, Fuzzy C-Means, Stunting.*

## ABSTRACT

*Stunting* is still a challenge in itself in Indonesia, including in Central Java where the *stunting* prevalence is above the national average *stunting* prevalence. For grouping *stunting* factors, it is proposed to use the Fuzzy C-Means Algorithm. This method has the advantage of more precise placement of the cluster center compared to other methods, but the number of clusters in the Fuzzy C-Means Algorithm must be determined before conducting research. To determine the optimal number of clusters in the Fuzzy C-Means Algorithm, this research proposes the *Cluster Validity Index* (CVI) method. CVI is a validity measure to find the optimal number of clusters that can fully explain the data structure by measuring the degree of cohesiveness in a cluster and the separation between clusters. In this study using the *Partition Entropy Index* (PEI), *Partition Coefficient Index* (PCI), *Fukuyama Sugeno Index* (FSI), *Xie Beni Index* (XBI), *Modification Partition Coefficient Index* (MPCI), *Partition Coefficient And Exponential Separation Index* (PCAESI). The results of the research show that the optimal number of clusters based on CVI is 2 clusters. The first cluster consists of 17 regencies/cities and the second cluster consists of 18 regencies/cities.

**Keyword :** *Cluster Validity Index, Clustering, Fuzzy C-Means, Stunting.*

## 1. PENDAHULUAN

Masalah *stunting* sangatlah penting, kebijakan terkait *stunting* pada balita merupakan suatu kebutuhan yang mendesak karena berkaitan dengan masa depan anak-anak dan dapat berpengaruh pada kualitas suatu bangsa. Dalam upaya mengatasi *stunting* balita, pemerintah mencanangkan pengentasan masalah *stunting* pada Sembilan Agenda (Nawacita), selain itu pengentasan masalah *stunting* juga diperlukan untuk merespon agenda global seperti SDGs (*Sustainable Development Goals*). Dari berbagai program pemerintah, stratifikasi provinsi berdasarkan prevalensi *stunting* balita dan tingkat kerawanan pangan telah dilakukan untuk mengurangi jumlah *stunting* pada anak. Kebijakan yang sama dapat diterapkan pada wilayah dengan strata yang sama tanpa melihat karakteristik faktor *stunting* daerah terkait (Istiqomah, 2015).

Pada tahun 2018 *stunting* masih menjadi tantangan sendiri di Indonesia, termasuk di Jawa Tengah dimana prevalensi *stunting* masih berada di angka 34,3% di atas rata-rata prevalensi *stunting* nasional yaitu sebesar 30,8%. Menurut WHO (1995) jika prevalensi *stunting* berada pada kisaran 30-39% maka masalah kesehatan masyarakat dianggap berat dan dianggap serius apabila prevalensi *stunting* lebih besar dari 40%. Provinsi Jawa Tengah terdiri dari 35 kabupaten/kota yang memiliki karakteristik faktor *stunting* balita yang berbeda di setiap wilayah. Berdasarkan hal tersebut diperlukan pengelompokan kabupaten/kota di Jawa Tengah yang didasarkan pada kemiripan karakteristik faktor *stunting* balita sehingga proses monitoring akan dilakukan berdasarkan kelompok-kelompok kabupaten/kota yang terbentuk. Indikator *stunting* balita dapat digunakan untuk mengetahui kondisi yang terjadi di suatu wilayah, maka dari itu pengelompokan kabupaten/kota di Jawa Tengah akan dilakukan berdasarkan pada faktor *stunting* balita di tiap wilayah.

Pemetaan dan pengelompokan daerah berdasarkan karakteristik faktor *stunting* dapat dilakukan menggunakan analisis kluster. Analisis kluster atau klastering merupakan suatu teknik dalam data mining yang bersifat *unsupervised learning*. Klastering bertujuan untuk mengelompokkan data tidak berlabel ke dalam beberapa kelompok berdasarkan karakteristik masing-masing (Mashfuufah & Istiawan, 2018).

## 2. TINJAUAN PUSTAKA

*Stunting* didefinisikan sebagai bentuk kegagalan pertumbuhan yang menyebabkan penundaan baik fisik dan kognitif pada pertumbuhan dan perkembangan anak. *Stunting* juga dikenal sebagai kegagalan pertumbuhan linear, didefinisikan sebagai ketidakmampuan untuk mencapai potensi tinggi badan untuk usia tertentu. Anak-anak yang mengalami *stunting* akan memiliki risiko yang lebih tinggi akibat dari kesehatan yang buruk sepanjang hidupnya (Reinhardt dan Fanzo, 2014).

Pengelompokan faktor *stunting* merupakan salah satu cara untuk mengidentifikasi karakteristik *stunting* di masing-masing Kabupaten/Kota agar dalam mengambil kebijakan dan strategi pembangunan tepat sasaran dan tepat guna. Analisis kluster berguna dalam beberapa eksplorasi pola-analisis, pengelompokan, dan pengambilan keputusan (Hafiludien & Istiawan, 2018).

Algoritma Fuzzy C-Means merupakan salah satu analisis kluster yang paling banyak dikenal (Simhachalam & Ganesan, 2015) dan dipelajari secara luas (Filho, Pimentel, Souza, & Oliveira, 2015). Algoritma Fuzzy C-Means juga diketahui memiliki kemampuan yang baik dalam mendeteksi kluster (Pimentel & de Souza, 2016) dan memiliki ketepatan dalam penempatan pusat kluster (Scitovski, Vidovi, & Bajer, 2016). Namun algoritma Fuzzy C-Means memiliki kelemahan pada penentuan jumlah kluster optimal (Rezaee, 2010) (Yang & Nataliani, 2017) yang harus ditentukan sebelum melakukan penelitian (Wang & Zhang, 2007).

Menurut (Yu, Liu, & Wang, 2013) menentukan jumlah kluster pada analisis kluster adalah hal yang paling menantang dan sulit. Untuk memvalidasi apakah jumlah kluster yang diperoleh dapat menjelaskan keseluruhan data atau tidak dapat menggunakan Cluster Validity Index (CVI) (Wang & Zhang, 2007).

### 3. METODOLOGI PENELITIAN

#### 3.1 Fuzzy C-Means

Algoritma Fuzzy C-Means menghubungkan derajat keanggotaan suatu objek dengan jarak objek tersebut pada pusat kelompoknya. Suatu objek akan cenderung menjadi anggota suatu *cluster* dimana objek itu memiliki derajat keanggotaan tertinggi. Dalam teori fuzzy, keanggotaan sebuah data diberikan dengan suatu nilai derajat keanggotaan yang jangkauan nilainya 0 sampai 1. Semakin tinggi nilai derajat keanggotaannya maka semakin tinggi nilai keanggotaan sebuah data dalam suatu kelompok dan semakin kecil nilai derajat keanggotaannya maka semakin rendah nilai keanggotaan sebuah data dalam suatu kelompok. Tahapan algoritma Fuzzy C-Mean (Zhang & Wang, 2008) sebagai berikut :

1. Menentukan jumlah kluster ( $k \geq 2$ )
2. Menentukan pangkat untuk matriks partisi  $w$  ( $w > 1$ )
3. Menginisialisasi matriks partisi awal  $U_i$
4. Menghitung nilai *centroid*  $c_{ij}$

$$c_{ij} = \frac{\sum_{i=1}^N u_{ij}^w x_{ij}}{\sum_{i=1}^N u_{il}^w}$$

5. Menghitung nilai derajat keanggotaan

$$u_{ij} = \frac{D(x_i, c_j)^{\frac{-2}{w-1}}}{\sum_{l=1}^k D(x_i, c_l)^{\frac{-2}{w-1}}}$$

6. Menghitung nilai fungsi objektif.

$$J_m(\tilde{U}, v) = \sum_{k=1}^n \sum_{i=1}^n (u_{ij})^w D(x_i, c_l)^2$$

Jika perubahan nilai fungsi objektif sudah berada di bawah nilai ambang batas, maka berhenti. Jika masih di atas nilai ambang batas yang ditentukan, ulangi langkah keempat dan kelima.

#### 3.2 Cluster Validity Index (CVI)

*Cluster Validity Index* (CVI) adalah sebuah ukuran validitas untuk menemukan jumlah kluster optimal yang sepenuhnya dapat menjelaskan struktur data (K. L. Wu & Yang, 2005). Secara umum ada dua fungsi CVI. Yang pertama CVI dapat digunakan untuk menentukan jumlah kluster, dan yang kedua adalah CVI dapat menjelaskan partisi terbaik yang dilakukan. Penentuan jumlah kluster pada suatu data merupakan teori penting karena memiliki peran penting untuk mendapatkan hasil dugaan yang baik pada analisis kluster. Jika penentuan jumlah kluster tidak sesuai maka akan berdampak pada hasil partisi yang tidak optimal sehingga dapat mengakibatkan kesalahan ketika pengambilan keputusan (Wang & Zhang, 2007). Penentuan kluster optimal dalam CVI yaitu dengan mengukur derajat *compactness* atau kekompakan dan *separation* atau separasi struktur data pada seluruh kluster. Kluster optimal ditentukan ketika nilai terbesar maupun nilai terkecil dari CVI terbentuk, sesuai kriteria dari masing-masing CVI (Wang & Zhang, 2007). Ada 6 Metode CVI untuk menentukan jumlah kluster optimal diantaranya :

1. *Partition Coefficient Index (PCI)*

Metode ini menentukan nilai yang semakin besar mempunyai arti bahwa kualitas kluster yang didapat semakin baik. Berikut formula untuk menghitung PCI:

$$PCI = \frac{1}{N} \left( \sum_{i=1}^N \sum_{j=1}^K u_{ij}^2 \right)$$

2. *Partition Entropy Index (PEI)*

Metode ini menentukan nilai yang semakin kecil mempunyai arti bahwa kualitas kluster yang didapat semakin baik. Berikut ini formula untuk menghitung PEI:

$$PEI = -\frac{1}{N} \left( \sum_{i=1}^N \sum_{j=1}^K u_{ij} \times \log_2 u_{ij} \right)$$

3. *Modified Partition Coefficient Index (MPCI)*

Metode ini menentukan nilai yang semakin kecil mempunyai arti bahwa kualitas kluster yang didapat semakin baik. Berikut ini formula untuk menghitung MPCI:

$$MPCI = 1 - \frac{K}{K-1} (1 - PCI)$$

4. *Fukuyama Sugeno Index (FSI)*

Metode ini menentukan nilai yang semakin kecil mempunyai arti bahwa kualitas kluster yang didapat semakin baik. Berikut ini formula untuk menghitung FSI:

$$FSI = \sum_{i=1}^N \sum_{j=1}^K u_{ij}^m \times d(x_i, c_j)^2 - \sum_{i=1}^N \sum_{j=1}^K u_{ij}^m \times d(c_j, \bar{x})^2$$

$$FSI = J_m(u, c) - K_m(u, c)$$

5. *Xie Beni Index (XBI)*

Metode ini menentukan nilai yang semakin kecil mempunyai arti bahwa kualitas kluster yang didapat semakin baik. Berikut ini formula untuk menghitung XBI:

$$XBI = \frac{\sum_{j=1}^K \sum_{i=1}^N u_{ij}^m \times d(x_i, c_j)}{N \times \min_{i,j} (d(c_i, c_j)^2)} = \frac{J_m(u, c)/N}{Sep(c)}$$

6. *Partition Coefficient And Exponential Separation Index (PCAESI)*

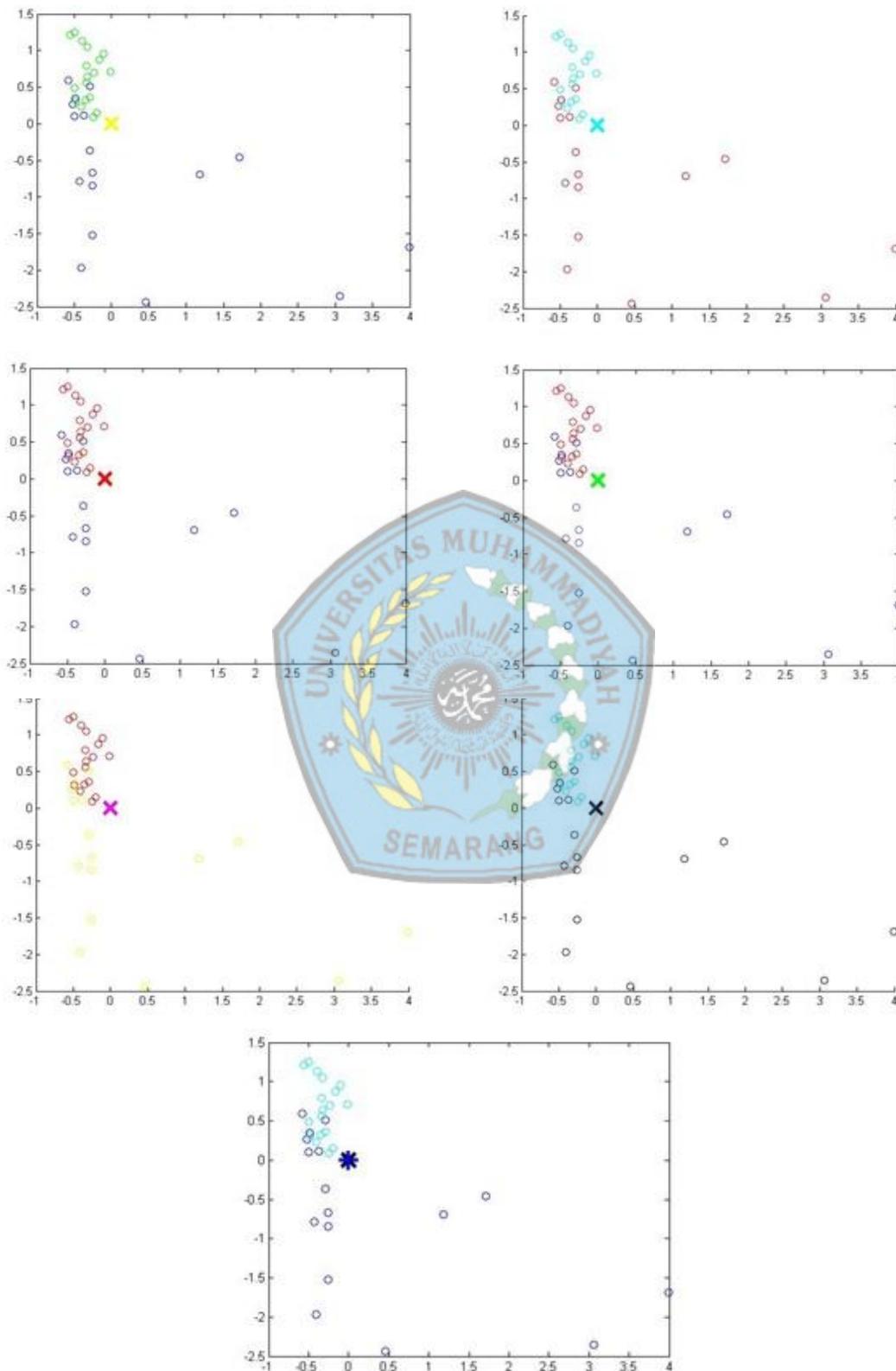
Metode ini menentukan nilai yang semakin besar mempunyai arti bahwa kualitas kluster yang didapat semakin baik. Berikut ini formula untuk menghitung PCAESI:

$$PCAESI = \sum_{j=1}^K PCAESI = Koh - Sep = \sum_{j=1}^K Koh_j - \sum_{j=1}^N Sep_j$$

#### 4. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data sekunder yang diambil dari Publikasi “Indeks Pembangunan Kesehatan Masyarakat 2018”, “Profil Kesehatan Provinsi Jawa Tengah 2018” dari Kementerian Kesehatan dan Survei Sosial Ekonomi Nasional (Susenas) tahun 2018 dari Badan Pusat Statistik (BPS) Provinsi Jawa Tengah. Variabel yang digunakan dalam penelitian ini menggunakan 15 variabel dari faktor *stunting* balita antara lain Pengeluaran rata-rata Ruta Balita (X<sub>1</sub>), Persentase Pendidikan Perempuan < SMA (X<sub>2</sub>), Persentase Balita dengan Imunisasi Lengkap (X<sub>3</sub>), Cakupan Pelayanan Kesehatan Bayi (X<sub>4</sub>), Cakupan Pelayanan Kesehatan Balita (X<sub>5</sub>), Kecukupan Posyandu (X<sub>6</sub>), Kecukupan Puskesmas (X<sub>7</sub>), Kecukupan Dokter (X<sub>8</sub>), Kecukupan Tenaga Ahli gizi (X<sub>9</sub>), Kecukupan Bidan (X<sub>10</sub>),

Persentase BBLR ( $X_{11}$ ), Prevalensi Balita Gizi Buruk dan Kurang ( $X_{12}$ ), Prevalensi Diare Balita ( $X_{13}$ ), Prevalensi ISPA Balita ( $X_{14}$ ), Rata-rata lama pemberian ASI balita ( $X_{15}$ ). Hasil pengelompokan kluster dapat dilihat pada Gambar dan Tabel berikut :



Gambar **Error! No text of specified style in document.**1 Hasil klastering dengan jumlah kluster dua sampai delapan

**Tabel Error! No text of specified style in document..1 Hasil dari Cluster Validity Index (CVI)**

<i>k</i>	PCI	PEI	MPCI	FSI	XBI	PCAESI
2	<b>0.50</b>	<b>1.00</b>	0.0000002620	255.00	<b>495,83</b>	<b>1.96</b>
3	0.33	1.59	0.0000001735	170.00	5,178,700	1.32
4	0.25	2.00	<b>0.0000001376</b>	127.50	6,523,200	1.61
5	0.20	2.32	0.0000001470	102.00	56,105,000	1.69
6	0.17	2.59	0.0000001722	85.00	109,360,000	1.69
7	0.14	2.81	0.0000001462	72.86	55,697,000	0.84
8	0.13	3.00	0.0000001428	<b>63.75</b>	126,920,000	1.10

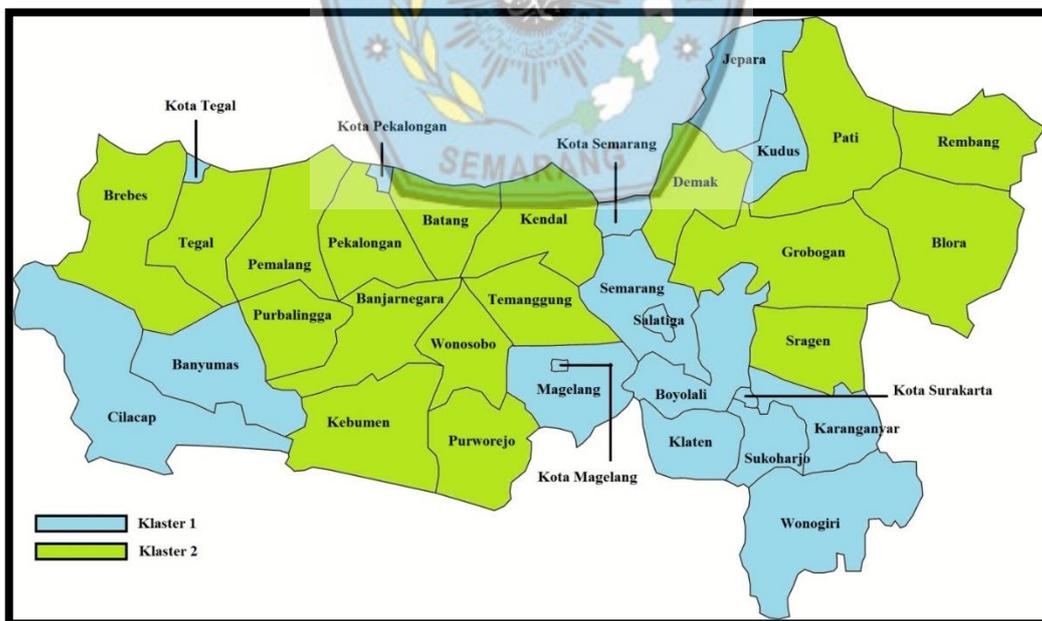
Tabel 4.1 di atas menunjukkan hasil dari masing-masing CVI. PEI, FSI, XBI dan MPCI menunjukkan nilai sekecil mungkin untuk memperoleh kluster optimalnya. Untuk PCI dan PCAESI menunjukkan nilai sebesar mungkin untuk memperoleh kluster optimalnya. Dari keenam CVI tersebut, empat CVI menunjukkan bahwa kluster kedua merupakan kluster optimal. Gambar 4.2 di bawah ini merupakan gambar grafik dari masing-masing CVI.

**Tabel Error! No text of specified style in document..2 Centroid** dari kluster yang terbentuk

<i>Centroid</i>	Kluster 1	Kluster 2
$X_1$	0.2549	-0.2607
$X_2$	-0.8743	0.8627
$X_3$	0.1752	-0.1865
$X_4$	0.3918	-0.3974
$X_5$	0.5034	-0.5092
$X_6$	0.8694	-0.8695
$X_7$	-0.2815	0.2871
$X_8$	0.4928	-0.5042
$X_9$	0.4482	-0.4481
$X_{10}$	-0.6956	0.6955
$X_{11}$	-0.3960	0.3960
$X_{12}$	-0.1714	0.1770
$X_{13}$	-0.5663	0.5720
$X_{14}$	-0.2602	0.2716
$X_{15}$	0.2430	-0.2314

Tabel Error! No text of specified style in document..3 Keanggotaan Klaster

Klaster 1	Klaster 2
Cilacap	Purbalingga
Banyumas	Banjarnegara
Kota Magelang	Kebumen
Boyolali	Purworejo
Klaten	Wonosobo
Sukoharjo	Sragen
Wonogiri	Grobogan
Karanganyar	Blora
Kudus	Rembang
Jejara	Pati
Kota Semarang	Demak
Magelang	Temanggung
Surakarta	Kendal
Salatiga	Batang
Semarang	Pekalongan
Kota Pekalongan	Pemalang
Kota Tegal	Tegal
	Brebes



Gambar Error! No text of specified style in document..2 Peta Tematik Pengelompokan Kabupaten/Kota di Jawa Tengah

### 5. KESIMPULAN

Penerapan *Cluster Validity Index* (CVI) pada Algoritma Fuzzy C-Means pada pengelompokan Kabupaten/Kota berdasarkan faktor *stunting* balita menghasilkan dua

klaster optimal. Klaster pertama terdiri dari 17 Kabupaten/Kota dengan kategori baik dan Klaster kedua terdiri dari 18 Kabupaten/Kota yang dikategorikan kurang baik.

## 6. DAFTAR PUSTAKA

- Babuska, R. (2001), "Fuzzy Clustering", dalam Fuzzy And Neural Control, Babuska, R., Netherland, pp. 51-72.
- Hidayat, R., Wasono, R., & Darsyah, M. Y. (2017). Pengelompokan Kabupaten/Kota Di Jawa Tengah Menggunakan Metode K-Means Dan Fuzzy C-Means. *Seminar Nasional Pendidikan, Sains Dan Teknologi*, 240–250.
- Istiqomah, N. (2015), "Penggerombolan Kabupaten/Kota Berdasarkan Faktor Stunting Menggunakan Metode Penggerombolan Dua Langkah Untuk Data Campuran".
- Mashfuufah, S., & Istiawan, D. (2018). Penerapan Partition Entropy Index, Partition Coefficient Index dan Xie Beni Index untuk Penentuan Jumlah Klaster Optimal pada Algoritma Fuzzy C-Means dalam Pemetaan Tingkat Kesejahteraan Penduduk Jawa Tengah. *The 7th University Research Colloquium*, 51–60.
- Wang, W., & Zhang, Y. (2007). On Fuzzy Cluster Validity Indices. *Fuzzy Sets and Systems*, 158, 2095–2117. <https://doi.org/10.1016/j.fss.2007.03.004>
- Wu, C, Ouyang, C, Chen, L, & Lu, L. (2014). A New Fuzzy Clustering Validity Index with a Median Factor for Centroid-based Clustering. *IEEE Transactions on Fuzzy System*, 6706(c), 1–17. <https://doi.org/10.1109/TFUZZ.2014.2322495>
- Wu, K. L, & Yang, M. (2005). A Cluster Validity Index for Fuzzy Clustering. *Pattern Recognition Letters*, 26(9), 1275–1291. <https://doi.org/10.1016/j.patrec.2004.11.022>