



# Pemetaan Kejadian Balita *Stunting* Melalui Integrasi Citra Satelit Multisumber dan *Official Statistics* di Provinsi Nusa Tenggara Timur Menggunakan *Machine Learning*

Suhendra Widi Prayoga<sup>1</sup>, Setia Pramana<sup>2</sup>

<sup>1,2</sup>D-IV Komputasi Statistik, Politeknik Statistika STIS

<sup>1</sup>[222112382@stis.ac.id](mailto:222112382@stis.ac.id)

<sup>2</sup>[setia.pramana@stis.ac.id](mailto:setia.pramana@stis.ac.id)

Corresponding author email: [setia.pramana@stis.ac.id](mailto:setia.pramana@stis.ac.id)

**Abstract:** *Stunting is a serious problem for the health of children under five. The lack of quality of life for toddlers is one of the causes of the current high prevalence of stunting. East Nusa Tenggara ranks second with the highest prevalence of stunting in Indonesia in 2023. This high rate is triggered by several factors, such as health, socio-economics and environment. In terms of the environment, remote sensing technology can be utilised as a supporting tool in monitoring the incidence of stunting in the region. This study aims to map districts/cities in East Nusa Tenggara based on the incidence of stunting in children under five years old through the integration of multi-source satellite imagery and official statistics using machine learning algorithms. Researchers used non-hierarchical clustering methods such as K-Means, K-Medoids, and Fuzzy C-Means which will then be compared to find the best method. The best method was obtained based on internal validity, such as connectivity, Dunn index, and Silhouette coefficient. The results of this study show that K-Means is the best clustering method based on internal validity criteria. The optimal number of clusters formed is two with connectivity, Dunn index, and Silhouette coefficient values of 2.9290, 0.6931, and 0.4509, respectively. Areas included in cluster two are still very vulnerable to stunting problems. The cluster results obtained are expected to be the basis for the government in overcoming stunting problems in East Nusa Tenggara.*

**Keywords:** *machine learning, East Nusa Tenggara, official statistics, remote sensing, stunting*

**Abstrak:** *Stunting menjadi masalah yang serius bagi kesehatan balita. Kurangnya kualitas hidup balita menjadi salah satu penyebab tingginya prevalensi stunting saat ini. Nusa Tenggara Timur menempati peringkat kedua dengan prevalensi stunting tertinggi di Indonesia tahun 2023. Tingginya angka ini dipicu oleh beberapa faktor, seperti kesehatan, sosial-ekonomi, dan lingkungan. Dari segi lingkungan, teknologi remote sensing dapat dimanfaatkan sebagai alat pendukung dalam memantau kejadian stunting di wilayah tersebut. Penelitian ini bertujuan untuk memetakan wilayah kabupaten/kota di Nusa Tenggara Timur berdasarkan indikator kejadian balita stunting melalui integrasi citra satelit multisumber dan official statistics menggunakan algoritma machine learning. Peneliti menggunakan metode non-hierarchical clustering berupa K-Means, K-Medoids, dan Fuzzy C-Means yang kemudian akan dibandingkan untuk menemukan metode terbaik. Metode terbaik diperoleh berdasarkan validitas internal, berupa connectivity, indeks Dunn, dan koefisien Silhouette. Hasil penelitian ini menunjukkan bahwa K-Means menjadi metode clustering terbaik berdasarkan kriteria validitas internal. Jumlah klaster optimal yang terbentuk sebanyak dua dengan nilai connectivity, indeks Dunn, dan koefisien Silhouette berturut-turut sebesar 2.9290, 0.6931, dan 0.4509. Wilayah yang termasuk ke dalam klaster dua termasuk masih sangat rentan terhadap permasalahan stunting. Hasil klaster yang diperoleh diharapkan dapat menjadi dasar bagi pemerintah dalam mengatasi permasalahan stunting di Nusa Tenggara Timur.*

**Kata kunci:** *machine learning, Nusa Tenggara Timur, official statistics, remote sensing, stunting*

## I. PENDAHULUAN

Malnutrisi menjadi salah satu ancaman bagi kehidupan seseorang dan pembangunan suatu negara sehingga memerlukan perhatian penting untuk mengatasi masalah tersebut [1]. *Stunting*, salah satu dampak malnutrisi, merupakan masalah kesehatan yang serius bagi balita karena dapat mengganggu perkembangan fisik dan kognitif mereka serta berdampak signifikan bagi kualitas hidup. Menurut WHO, *stunting* adalah masalah pertumbuhan pada anak yang diakibatkan oleh malnutrisi kronis dan infeksi berulang, yang ditandai dengan tinggi badan yang lebih rendah dari rata-rata [2]. Hal ini menjadi salah satu fokus utama dalam mencapai *Sustainable Development Goals* (SDGs) tujuan 2, target 2.2, yaitu



segala bentuk malnutrisi harus berakhir pada tahun 2030, termasuk mencapai target terkait *stunting* dan *wasting* pada balita.

Berdasarkan data WHO dan UNICEF, Indonesia menempati posisi ke-27 dari 154 negara di dunia terkait angka prevalensi *stunting* dan posisi ke-5 di Asia [3], [4]. Hal ini menunjukkan betapa mendesaknya kebutuhan untuk mengambil tindakan efektif dalam menangani masalah ini, terutama di daerah-daerah dengan prevalensi *stunting* yang sangat tinggi. Hasil Survei Kesehatan Indonesia 2023 melaporkan bahwa Provinsi Nusa Tenggara Timur memiliki prevalensi *stunting* tertinggi kedua di Indonesia, yang dapat dipicu oleh faktor-faktor, seperti kesehatan, sosial-ekonomi, bahkan lingkungan [5]. Kondisi lingkungan menjadi faktor penting untuk mengukur seberapa layak wilayah yang ditempati masyarakat. Lingkungan yang sehat dan memadai dapat membantu pertumbuhan dan perkembangan anak secara optimal. Sebaliknya, lingkungan yang buruk dapat menghambat akses terhadap air bersih dan sanitasi yang memadai, serta meningkatkan risiko infeksi dan penyakit yang dapat memperburuk malnutrisi [6].

Pemanfaatan teknologi *remote sensing* sebagai alternatif pemantauan kondisi lingkungan secara tidak langsung melalui citra satelit multisumber menawarkan sejumlah keuntungan, seperti kemampuan untuk mengumpulkan data dalam skala besar (*big data*) dan detail tinggi. Penelitian [7] memanfaatkan citra satelit Landsat-8 untuk menganalisis kejadian *stunting* di Lampung menggunakan *Artificial Neural Network*. Pemetaan dilakukan berdasarkan prediksi *stunting* dengan algoritma tersebut. Selain itu, pemantauan kondisi juga memerlukan data statistik yang berasal dari lembaga resmi pemerintah, seperti Badan Pusat Statistik dan Kementerian Kesehatan. Penelitian [1] menggunakan data *official statistics* untuk memetakan provinsi berdasarkan indeks khusus penanganan *stunting* di Indonesia melalui metode *hierarchical* dan *nonhierarchical clustering*. Integrasi data citra satelit multisumber dan *official statistics* memungkinkan pemetaan yang lebih akurat dan komprehensif tentang kejadian *stunting* di Nusa Tenggara Timur.

Penelitian ini melakukan pemetaan kabupaten/kota berdasarkan indikator kejadian *stunting* di Provinsi Nusa Tenggara Timur melalui integrasi citra satelit multisumber dan *official statistics*. Metode *machine learning*, khususnya algoritma *clustering* seperti *K-Means*, *K-Medoids*, dan *Fuzzy C-Means*, digunakan untuk menganalisis data ini karena kemampuannya dalam mengidentifikasi pola dan mengelompokkan secara efektif sehingga memberikan pendekatan holistik dan inovatif dalam pemantauan dan analisis. Hasil penelitian diharapkan dapat menjadi dasar bagi pemerintah dalam merumuskan kebijakan dan strategi yang lebih efektif untuk mengatasi masalah *stunting*, khususnya di Nusa Tenggara Timur, serta mendukung pencapaian SDGs tujuan 2, target 2.2.

## II. METODE PENELITIAN

### 2.1. Lokasi Penelitian

Penelitian ini mengambil lokasi studi di Provinsi Nusa Tenggara Timur, Indonesia pada tahun 2023. Provinsi Nusa Tenggara Timur merupakan sebuah provinsi yang berada bagian timur Indonesia, berdekatan dengan Pulau Maluku dan Papua. Provinsi ini dipilih sebagai wilayah studi pada penelitian ini karena memiliki angka prevalensi *stunting* yang tinggi tiap tahunnya.

### 2.2. Data dan Variabel

Penelitian ini menggunakan data *official statistics* yang diperoleh dari Badan Pusat Statistik dan Kementerian Kesehatan serta citra satelit berupa *Normalized Difference Vegetation Index* (NDVI), *Soil Adjusted Vegetation Index* (SAVI), *Normalized Difference Built-up Index* (NDBI), *Normalized*



*Difference Water Index* (NDWI) dari Landsat-8 dan *Land Surface Temperature* (LST) dari MODIS yang diperoleh melalui pemrosesan *Google Earth Engine*. Berikut rincian variabel pada penelitian ini.

Tabel 1. Deskripsi data

Variabel	Deskripsi	Satuan	Sumber	Referensi
wasting	Balita yang mengalami <i>wasting</i>	Persen	Kemenkes	[8]
imunisasi	Penduduk Usia 0-59 bulan Menurut Pemberian Imunisasi Lengkap	Persen		[9]
sanitasi	Rumah Tangga dengan Akses Sanitasi Layak	Persen		[8]
asi_eks	Penduduk Usia 0-23 Bulan yang Pernah Diberi ASI	Persen		[8]
air_minum	Rumah Tangga dengan Sumber Air Minum Layak	Persen		[10]
miskin	Penduduk Miskin	Persen	BPS	[9]
rls_p	Rata-Rata Lama Sekolah Penduduk Perempuan	Tahun		[11]
sma_p	Penduduk Perempuan Berusia 15 Tahun ke Atas Menurut Ijazah/STTB SMA/Sederajat yang Dimiliki	Persen		[9]
huruf_p	Buta Huruf Penduduk Perempuan	Persen		[12]
konsumsi	Rata-rata Konsumsi Kalori Perkapita Sehari	kcal		[10]
NDVI	<i>Normalized Difference Vegetation Index</i>	Indeks		
NDWI	<i>Normalized Difference Water Index</i>	Indeks	Landsat-8	[7]
SAVI	<i>Soil Adjusted Vegetation Index</i>	Indeks		
NDBI	<i>Normalized Difference Built-up Index</i>	Indeks		
LST	Suhu Permukaan Lahan	Celcius	MODIS	[7]

### 2.3. Machine Learning

Pendekatan *machine learning* terbagi menjadi dua kategori, yaitu *supervised* dan *unsupervised*. *Unsupervised learning* merupakan algoritma *machine learning* di mana algoritma dilatih menggunakan data yang tidak berlabel [13][14]. Penelitian ini menggunakan algoritma *unsupervised* berupa *K-Means*, *K-Medoids*, dan *Fuzzy C-Means clustering*. Ketiga algoritma tersebut termasuk *non-hierarchical clustering* yang dirancang untuk mengelompokkan objek ke dalam sejumlah *k* klaster yang ditentukan terlebih dahulu [1].

#### 2.3.1. Principal Component Analysis (PCA)

Sebelum melakukan *clustering*, pemeriksaan multikolinearitas perlu dilakukan pada setiap variabel dalam data dengan melihat korelasi antar variabel. Kriteria korelasi dapat dikatakan ada multikolinearitas apabila  $\geq 0.8$  sehingga data perlu diatasi. *Principal Component Analysis* (PCA) merupakan salah satu metode untuk mengatasi multikolinearitas dalam data dengan mereduksi kompleksitas dimensi data menjadi lebih sederhana [15]. Proses reduksi variabel membentuk komponen baru yang mempertahankan variasi total sehingga tidak menghilangkan keseluruhan informasi awal. Komponen baru yang dihasilkan bersifat independen satu sama lain, dan setiap komponen dapat mewakili atau menggantikan informasi dari variabel-variabel dalam data [1]. Jumlah komponen utama dapat ditentukan menggunakan *scree plot* yang melandai, kumulatif proporsi varians total minimal 80%, dan nilai *eigen*  $> 1$ [16].

#### 2.3.2. K-Means Clustering

Metode *K-Means clustering* telah diperkenalkan oleh Macqueen sejak tahun 1967. *K-Means* merupakan algoritma *unsupervised learning* berbasis jarak yang kemampuannya cocok untuk dataset besar secara cepat dan efisien. Perbedaan jarak yang kecil akan memberikan kesamaan secara signifikan antar objek. Algoritma *K-Means clustering* dimulai dengan menentukan jumlah klaster *k* yang diinginkan, biasanya dengan *Elbow* dan *Silhouette plot*. Setelah itu, titik-titik data dipilih secara acak untuk menjadi pusat awal dari setiap klaster. Selanjutnya, jarak antara setiap titik data dan pusat klaster dihitung menggunakan *euclidean distance*. Berdasarkan jarak yang



dihitung ini, setiap titik data kemudian ditempatkan ke dalam kluster yang pusatnya paling dekat. Setelah semua titik data dikelompokkan ke dalam kluster, pusat kluster diperbarui dengan menghitung rata-rata posisi dari seluruh titik data dalam kluster tersebut. Proses penempatan titik data ke dalam kluster dan pembaruan pusat kluster ini diulangi terus-menerus hingga posisi pusat kluster tidak lagi berubah, yang menandakan bahwa kluster telah mencapai kestabilan [17].

### 2.3.3. *K-Medoids Clustering*

Salah satu pengembangan dari *K-Means clustering* adalah *K-Medoids* yang dirancang lebih tahan terhadap pencilan dan *noise*. *K-Medoids* menggunakan nilai median yang tidak terpengaruh oleh pencilan dan *noise*, sehingga lebih *robust* dibandingkan *K-Means* [18]. Dalam metode *K-Medoids*, setiap kluster diwakili oleh satu objek yang terletak di dekat pusat kluster, dengan tujuan mengurangi jumlah ketidaksamaan antara setiap objek dengan objek perwakilannya. Proses pembentukan kluster dengan *K-Medoids* dimulai dengan menentukan jumlah kluster awal  $k$  yang diinginkan. Selanjutnya, setiap objek dikelompokkan ke dalam kluster berdasarkan jarak terdekatnya dengan objek representatif. Objek baru kemudian dipilih secara acak untuk berfungsi sebagai *medoid*, dan set *medoid* baru ini digunakan untuk menghitung ulang biaya. Jika biaya yang dihasilkan kurang dari nol, *medoid* ke- $j$  diganti dengan *medoid* acak baru dari dataset. Langkah-langkah ini diulangi hingga tidak ada perubahan lebih lanjut dalam *medoid*, yang menandakan bahwa tidak ada objek yang berpindah lagi [1].

### 2.3.4. *Fuzzy C-Means Clustering*

Metode *Fuzzy C-Means* merupakan versi *fuzzy* dari algoritma *hard clustering*. *Fuzzy C-Means* termasuk metode *clustering* yang sering digunakan karena mudah dan efisien [19]. Algoritma ini melibatkan keanggotaan *fuzzy*  $\mu_{ij}$  dan *centroid* kluster  $V_i$ , dengan eksponen  $m$  yang digunakan untuk meminimalkan fungsi tujuan dan mengatur tingkat *fuzziness*. Nilai  $m = 1$  menghasilkan hasil yang jelas, sedangkan nilai  $m$  yang mendekati  $\infty$  menghasilkan ketidakjelasan yang sempurna. Biasanya nilai  $m$  terbaik antara 1 dan 3 digunakan untuk hasil yang lebih akurat. Persamaan (1) merupakan fungsi objektif pada *Fuzzy C-Means clustering* [1].

$$J_m = \sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^m d(X_j, V_i)^2 \quad (1)$$

di mana  $\mu_{ij}$ : nilai keanggotaan objek ke- $j$  pada kluster ke- $i$ ,  $d(X_j, V_i)$ : jarak antara objek ke- $j$  ( $X_j$ ) dengan *centroid* kluster ke- $i$  ( $V_i$ ),  $n$ : jumlah observasi, dan  $m$ : parameter *fuzziness* hasil *clustering* ( $m > 1$ ). Algoritma *Fuzzy C-Means* dimulai dengan menginisialisasi nilai keanggotaan  $\mu_{ij}$  dari objek ( $X_j$ ) pada kluster ke- $i$ , dengan ketentuan jumlah keanggotaan setiap objek sama dengan 1. Selanjutnya, *centroid fuzzy* ( $V_i$ ) dihitung dan dilanjutkan dengan memperbarui nilai keanggotaan *fuzzy*. Langkah tersebut diulangi hingga tidak ada nilai  $J_m$  yang menurun [1].

## 2.4. Evaluasi Hasil

Evaluasi hasil kluster dilakukan untuk menentukan metode terbaik berdasarkan kriteria validasi internal berikut:

### 2.4.1. Koefisien *Connectivity*

Semakin rendah nilai koefisien *connectivity*, semakin baik kluster yang terbentuk. Koefisien *connectivity* memiliki rentang dari nol (0) hingga tak hingga [1]. Persamaan (2) merupakan rumus koefisien *connectivity*.

$$Conn(C) = \sum_{i=1}^N \sum_{j=1}^L x_{i,nn_{i(j)}} \quad (2)$$



di mana  $Conn(C)$ : koefisien *connectivity*,  $x_{i,nn_{i(j)}}$ : bernilai nol jika  $i$  dan  $j$  berada dalam kluster yang sama dan  $1/j$  jika tidak,  $nn_{i(j)}$ : tetangga terdekat objek ke- $j$  dari objek ke- $i$ ,  $N$ : jumlah data observasi, dan  $L$ : parameter jumlah tetangga terdekat yang dipertimbangkan.

#### 2.4.2. Indeks *Dunn*

Indeks *Dunn* adalah perbandingan antara jarak minimum antara observasi dalam kluster yang memiliki perbedaan dengan jarak intra-kluster terbesar. Indeks *Dunn* memiliki rentang nilai dari nol sampai tak hingga. Nilai indeks *Dunn* yang semakin tinggi pada suatu kluster menunjukkan semakin bagus hasil klasternya [1]. Persamaan (3) untuk menghitung indeks *Dunn*.

$$D(C) = \frac{\min_{C_k, C_l \in C, C_k \neq C_l} \left( \min_{i \in C_k, j \in C_l} dist(i, j) \right)}{\max_{C_m \in C} (diam(C_m))} \quad (3)$$

di mana  $D(C)$ : indeks *Dunn* dan  $diam(C_m)$ : jarak maksimum antar observasi dalam kluster  $C_m$ .

#### 2.4.3. Koefisien *Silhouette*

Koefisien *Silhouette* digunakan sebagai metrik untuk mengevaluasi keandalan dari proses *clustering* suatu observasi. Nilai koefisien *Silhouette* berkisar antara -1 hingga 1. Nilai koefisien *Silhouette* yang mendekati 1 menunjukkan kluster yang baik, sedangkan yang mendekati -1 dianggap buruk [1]. Persamaan (4) merupakan rumus koefisien *Silhouette*.

$$S(i) = \frac{b_i - a_i}{\max(b_i, a_i)} \quad (4)$$

di mana  $S(i)$ : koefisien *Silhouette*,  $a_i$ : jarak rata-rata antara objek ke- $i$  dengan semua objek dalam kluster yang sama, dan  $b_i$ : jarak rata-rata antara objek ke- $i$  dengan objek di kluster tetangga terdekat.

### III. HASIL DAN PEMBAHASAN

Asumsi yang harus dipenuhi sebelum analisis kluster, yaitu pengecekan non-multikolinearitas. Berdasarkan korelasi setiap variabel, diketahui semua variabel indeks *remote sensing* memiliki korelasi lebih dari 0.80. Hal ini menunjukkan adanya multikolinearitas dalam data tersebut, sehingga perlu diatasi agar dapat dilakukan analisis lanjutan berupa *clustering*. Selanjutnya, analisis menggunakan PCA dilakukan dengan membentuk komponen utama yang merupakan kombinasi linear dari variabel dalam data.

Tabel 2. Penentuan jumlah komponen utama

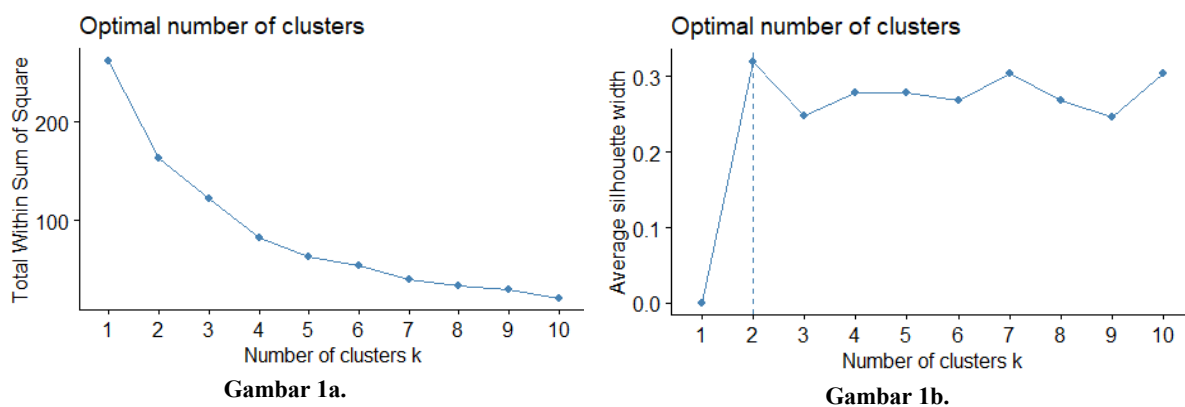
Komponen Utama	Nilai Eigen	Varians (%)	Kumulatif Varians (%)
1	5.8593	0.3906	39.0619
2	4.0405	0.2694	65.9984
3	1.4295	0.0953	75.5284
4	1.1618	0.0775	83.2737
5	0.9156	0.0610	89.3778
6	0.5710	0.0381	93.1845
7	0.3497	0.0233	95.5157
8	0.3169	0.0211	97.6281
9	0.1597	0.0106	98.6928
10	0.0735	0.0049	99.1826
11	0.0554	0.0037	99.5517
12	0.0326	0.0022	99.7690
13	0.0277	0.0018	99.9539
14	0.0052	0.0003	99.9887
15	0.0017	0.0001	100



Berdasarkan Tabel 2, terlihat bahwa jumlah komponen utama yang terbentuk sebanyak empat karena nilai  $eigen > 1$  dan kumulatif varians sebesar 83.27%. Keragaman data yang dapat dijelaskan sebesar 83.27% cukup menggambarkan struktur data. Dengan demikian, penelitian ini menggunakan empat komponen utama untuk analisis selanjutnya berupa *clustering*.

### 3.1. K-Means Clustering

Metode *K-Means* menerapkan pengelompokan setiap observasi (objek) berdasarkan titik pusat yang paling dekat. Pertama, jumlah kluster optimal ditentukan menggunakan *Elbow* dan *Silhouette plot*. Gambar (1) merupakan jumlah kluster optimal dari metode *K-Means*.

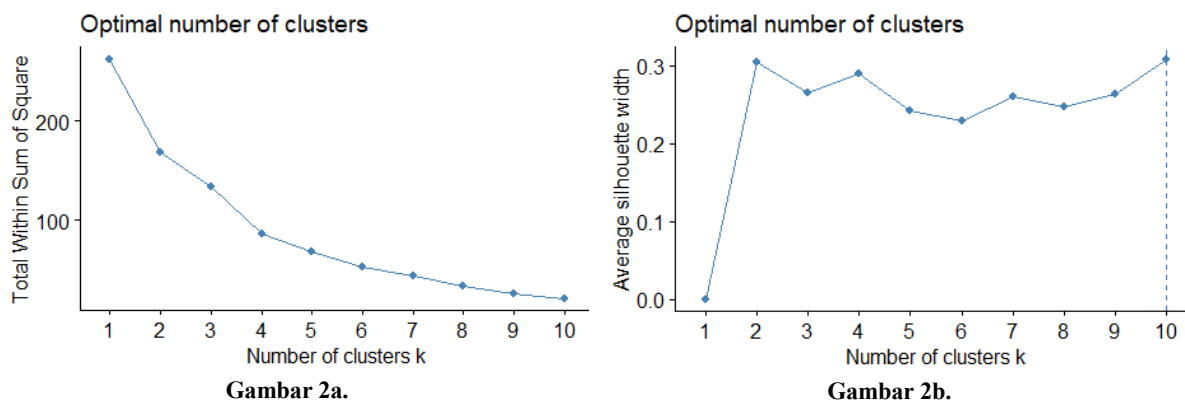


**Gambar 1.** Penentuan jumlah kluster optimal *K-Means* (1a) *Elbow plot* (1b) *Silhouette plot*

Berdasarkan Gambar (1), jumlah kluster optimal *K-Means* yang terbentuk yaitu sebanyak dua kluster. Oleh karena itu, jumlah kluster sebanyak dua akan digunakan sebagai analisis selanjutnya Hasil evaluasi berupa validasi internal diperoleh koefisien *connectivity*, indeks *Dunn*, dan koefisien *Silhouette* berturut-turut sebesar 2.9290, 0.6931, dan 0.4509.

### 3.2. K-Medoids Clustering

Metode *K-Medoids* lebih *robust* daripada *K-Means* yang menggunakan median karena tidak terpengaruh *outlier* dan *noise*. Tahap pertama pada *K-Medoids* sama halnya dengan *K-Means*, yaitu menentukan jumlah kluster optimal dengan metode *Elbow* dan *Silhouette*. Gambar (2) menunjukkan grafik jumlah kluster optimal dengan metode *K-Medoids*.

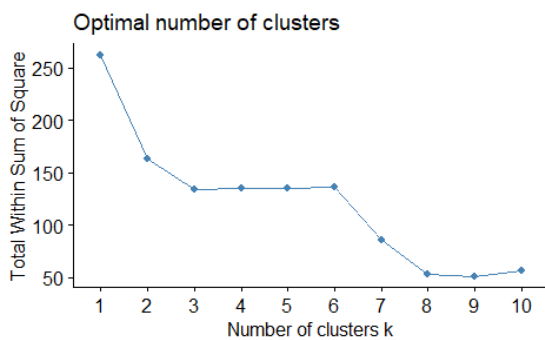


**Gambar 2.** Penentuan jumlah kluster optimal *K-Medoids* (2a) *Elbow plot* (2b) *Silhouette plot*

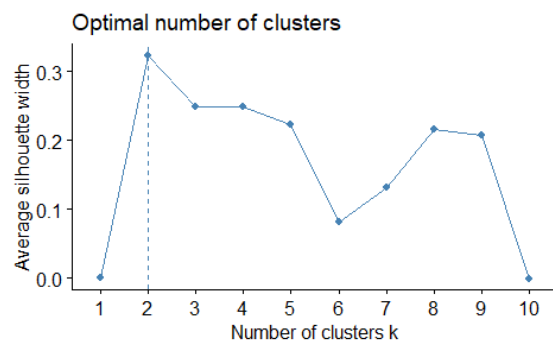
Berdasarkan Gambar (2), jumlah kluster optimal yang terbentuk yaitu sebanyak 10 kluster dengan metode *Silhouette*. Hasil ini menghasilkan karakteristik kluster yang beragam, sehingga sulit untuk diinterpretasikan. Namun, jumlah kluster sebanyak dua menjadi rekomendasi terbanyak kedua. Oleh karena itu, jumlah kluster yang digunakan untuk analisis selanjutnya adalah sebanyak dua kluster. Hasil evaluasi berupa validasi internal diperoleh koefisien *connectivity*, indeks *Dunn*, dan koefisien *Silhouette* berturut-turut sebesar 7.7012, 0.2126, dan 0.3047.

### 3.3. Fuzzy C-Means Clustering

Metode *Fuzzy C-Means* menerapkan nilai keanggotaan *fuzzy* untuk seberapa besar suatu objek termasuk dalam setiap kluster yang ada, berbeda dengan metode *hard clustering* yang menetapkan objek secara tegas ke dalam satu kluster. Algoritma ini mempertimbangkan setiap objek sebagai anggota dari setiap kluster dengan derajat keanggotaan yang bervariasi antara 0 dan 1. Penentuan jumlah kluster optimal pada metode ini masih sama dengan metode sebelumnya, yaitu *Elbow* dan *Silhouette*. Gambar (3) merupakan grafik jumlah kluster optimal pada *Fuzzy C-Means*.



Gambar 3a.



Gambar 3b.

Gambar 3. Penentuan jumlah kluster optimal FCM (3a) *Elbow plot* (3b) *Silhouette plot*

Berdasarkan Gambar (3), jumlah kluster optimal yang terbentuk sebanyak dua kluster. Oleh karena itu, jumlah kluster sebanyak dua akan digunakan sebagai analisis selanjutnya. Hasil evaluasi berupa validasi internal menunjukkan nilai koefisien *connectivity*, indeks *Dunn*, dan koefisien *Silhouette* berturut-turut sebesar 8.9230, 0.1075, dan 0.3226.

### 3.4. Evaluasi Hasil Kluster

Evaluasi hasil berupa validitas internal digunakan untuk menentukan metode analisis kluster terbaik berdasarkan koefisien *connectivity*, indeks *Dunn*, dan koefisien *Silhouette*. Tabel (3) menampilkan nilai validitas internal dari ketiga metode analisis kluster.

Tabel 3. Validitas internal

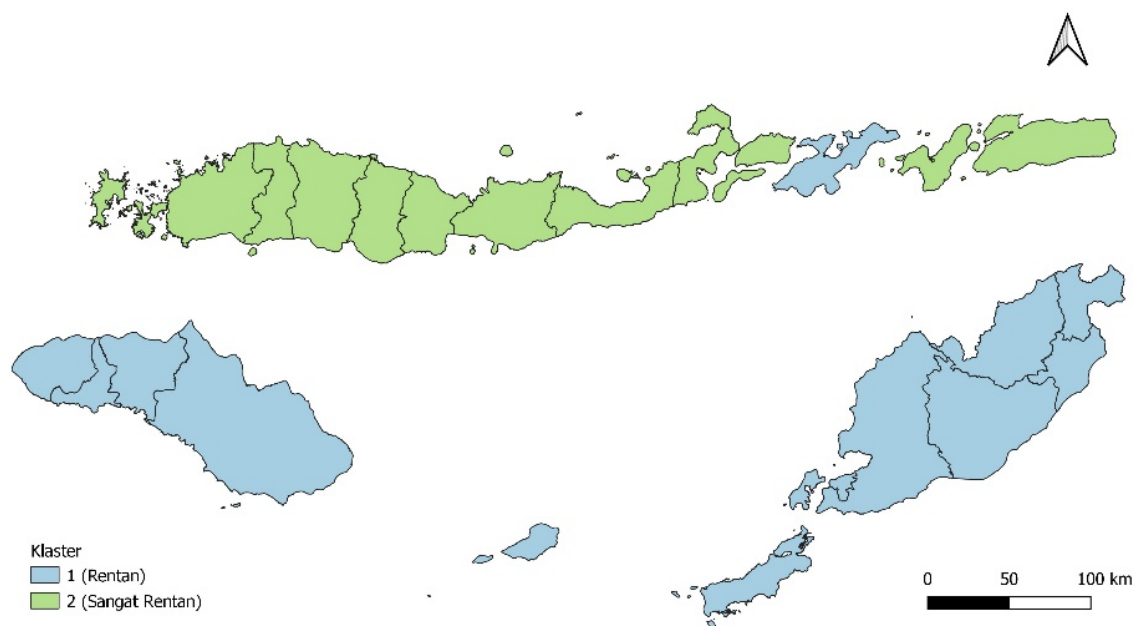
Metode	<i>Connectivity</i>	<i>Dunn</i>	<i>Silhouette</i>
<i>K-Means</i>	2.9290	0.6931	0.4509
<i>K-Medoids</i> (PAM)	7.7012	0.2126	0.3047
<i>Fuzzy C-Means</i> (FCM)	8.9230	0.1075	0.3226

Berdasarkan Tabel 3 yang menunjukkan validitas internal dari metode *K-Means*, *K-Medoids* (PAM), dan *Fuzzy C-Means* (FCM), dapat disimpulkan bahwa metode *K-Means* memberikan hasil kluster yang terbaik. Hal ini terlihat dari koefisien *connectivity* yang paling rendah (2.9290), menunjukkan kluster yang lebih terhubung dan kompak. Selain itu, nilai indeks *Dunn* tertinggi (0.6931) menunjukkan pemisahan kluster yang lebih baik, dan koefisien *Silhouette* tertinggi (0.4509) menunjukkan bahwa objek dalam kluster lebih mirip dengan klasternya sendiri dibandingkan dengan

kluster lain. Sebaliknya, metode *K-Medoids* dan *Fuzzy C-Means* menunjukkan kinerja yang kurang baik dengan nilai-nilai yang lebih tinggi pada *Connectivity* dan lebih rendah pada indeks *Dunn* serta koefisien *Silhouette*. Oleh karena itu, metode *K-Means* dapat dianggap lebih efektif dalam menghasilkan kluster yang baik dalam dataset yang dievaluasi.

### 3.5 Pemetaan Kejadian Balita *Stunting* di Provinsi Nusa Tenggara Timur

Setelah diperoleh metode *clustering* terbaik berupa *K-Means* sebanyak dua kluster, berikutnya dilakukan *profiling* kluster berupa pemetaan berdasarkan indikator balita *stunting* di Provinsi Nusa Tenggara Timur. Gambar (4) merupakan hasil pemetaan menggunakan metode *K-Means clustering*.



**Gambar 4.** Pemetaan kejadian balita *stunting* di Nusa Tenggara Timur tahun 2023

Gambar (4) menunjukkan bahwa wilayah kabupaten/kota bagian selatan di Provinsi Nusa Tenggara Timur merupakan kluster yang tergolong rentan terhadap masalah *stunting*. Kabupaten/kota yang tergolong ke dalam kluster ini, meliputi Belu, Kota Kupang, Kupang, Lembata, Malaka, Sabu Raijua, Rote Ndao, Sumba Barat Daya, Sumba Barat, Sumba Timur, Sumba Tengah, Timor Tengah Utara, dan Timor Tengah Selatan. Hal ini karena wilayah yang termasuk ke dalam kluster satu memiliki rata-rata persentase balita *wasting*, penduduk berusia 0 hingga 23 bulan yang pernah diberi ASI, dan rata-rata lama sekolah penduduk perempuan yang tidak berbeda dengan kluster dua yang tergolong sangat rentan. Sementara itu, untuk kluster dua yang meliputi Flores Timur, Ende, Alor, Manggarai, Ngada, Nagekeo, Manggarai Timur, Manggarai Barat, dan Sikka masih memiliki persentase *wasting* cukup tinggi, imunisasi lengkap rendah, sanitasi dan air minum layak sangat rendah, kemiskinan sangat tinggi, dan lain sebagainya.

**Tabel 4.** Rata-rata indikator kejadian balita *stunting* menggunakan metode *K-Means*

Indikator	Kluster 1	Kluster 2
wasting	7.93	9.25
imunisasi	69.57	64.14
sanitasi	79.66	70.71
asi_eks	96.90	94.53
air_minum	93.58	81.97
miskin	17.79	23.01





<b>Indikator</b>	<b>Klaster 1</b>	<b>Klaster 2</b>
rls_p	7.68	7.45
sma_p	14.81	17.43
huruf_p	3.48	8.65
konsumsi	1995.00	1892.75
NDVI	0.76	0.63
NDWI	-0.84	-0.77
SAVI	0.46	0.37
NDBI	-0.28	-0.13
LST	27.99	31.07

Selain ditinjau dari segi *official statistics*, *stunting* pada balita seringkali dipengaruhi oleh berbagai faktor lingkungan, termasuk pemantauan melalui peninderaan jauh. Seperti pada Tabel (4), kerapatan vegetasi yang lebih rendah pada klaster dua, ditunjukkan oleh nilai NDVI dan SAVI yang lebih rendah, dapat menunjukkan bahwa wilayah tersebut memiliki akses yang lebih terbatas terhadap sumber daya alam yang penting untuk kesehatan dan nutrisi, seperti buah-buahan dan sayuran segar. Suhu permukaan (LST) yang lebih tinggi pada klaster dua dapat mengindikasikan kondisi lingkungan yang lebih keras yang mungkin mempengaruhi ketersediaan air bersih dan kesehatan lingkungan secara keseluruhan. Kondisi lingkungan yang lebih keras ini dapat berkontribusi pada meningkatnya risiko penyakit yang dapat berdampak pada pertumbuhan anak-anak. Selain itu, nilai NDBI yang lebih tinggi pada klaster dua menunjukkan adanya lebih banyak area terbangun. Hal ini mengindikasikan urbanisasi yang lebih tinggi, yang sering kali dikaitkan dengan masalah kepadatan penduduk dan sanitasi yang buruk, yang juga dapat berkontribusi terhadap peningkatan angka *stunting*.

Oleh karena itu, wilayah kabupaten/kota yang termasuk ke dalam klaster satu perlu mengatasi permasalahan balita yang mengalami *wasting* dengan memberikan bantuan tambahan MPASI dan meningkatkan kesejahteraan pendidikan bagi perempuan. Sedangkan untuk kabupaten/kota yang termasuk dalam klaster dua wajib memperhatikan kesejahteraan penduduk secara keseluruhan, seperti mengentaskan kemiskinan, meningkatkan akses sanitasi dan air minum yang layak, mengatasi angka buta huruf penduduk perempuan, dan lain sebagainya. Dari segi lingkungan, wilayah pada klaster dua perlu mengatur tata ruang dan menghindari urbanisasi yang tidak terkontrol, yang sering kali berdampak negatif terhadap kesehatan lingkungan. Langkah-langkah yang perlu diambil termasuk meningkatkan ruang terbuka hijau (RTH) untuk menyediakan udara bersih dan mengurangi efek panas kota, menggunakan bahan bangunan yang ramah lingkungan untuk menurunkan suhu permukaan, serta menggalakkan kegiatan penghijauan dengan menanam pohon untuk menambah kerapatan vegetasi. Dengan demikian, lingkungan yang lebih sehat dan berkualitas dapat tercipta, yang akan mendukung kesehatan dan pertumbuhan balita, serta berkontribusi pada pengurangan angka *stunting* melalui peningkatan kondisi lingkungan dan akses terhadap sumber daya yang mendukung perkembangan yang sehat.

#### IV. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan dapat disimpulkan bahwa metode terbaik dalam pemetaan kabupaten/kota berdasarkan indikator kejadian balita *stunting* di Provinsi Nusa Tenggara Timur pada tahun 2023, yaitu metode *K-Means*. Metode terbaik ini terpilih setelah melalui perbandingan dengan kedua metode lainnya berdasarkan validitas internal berupa koefisien *connectivity*, indeks *Dunn*, dan koefisien *Silhouette*. Jumlah klaster dengan metode *K-Means* diperoleh sebanyak dua klaster, dengan kategori klaster satu termasuk rentan terhadap *stunting* yang terdiri dari 13 kabupaten/kota, sedangkan klaster dua termasuk sangat rentan terhadap permasalahan *stunting* yang



terdiri dari 9 kabupaten/kota. Dari pengkategorian tersebut, Provinsi Nusa Tenggara Timur masih memiliki permasalahan *stunting* yang sangat tinggi dibanding provinsi lainnya di Indonesia. Perlu penanganan khusus untuk mengatasi permasalahan *stunting* ini, khususnya klaster dua, seperti meningkatkan kesejahteraan masyarakat dengan memberikan bantuan kepada keluarga miskin, memperhatikan akses sanitasi dan air minum layak, meningkatkan pendidikan penduduk perempuan, dan lain sebagainya.

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Politeknik Statistika STIS yang telah memberikan dukungan dana penelitian dalam menyelesaikan kajian ini.

## REFERENSI

1. G. R. Suraya and A. W. Wijayanto, “Comparison of Hierarchical Clustering, K-Means, K-Medoids, and Fuzzy C-Means Methods in Grouping Provinces in Indonesia according to the Special Index for Handling Stunting,” *Indones. J. Stat. Its Appl.*, vol. 6, no. 2, pp. 180–201, 2022, doi: 10.29244/ijsa.v6i2p180-201.
2. D. T. Saputra, “Comparison Of Stunting Clusters For Each Province In Indonesia In 2019 And 2020 With 2021 And 2022 Using The K-Means Method,” *J. Indones. Sos. Teknol.*, vol. 5, no. 2, pp. 384–393, 2024, doi: 10.59141/jist.v5i2.920.
3. M. Alam, “Perlu Terobosan dan Intervensi Tepat Sasaran Lintas Sektor untuk Atasi Stunting,” *Kemenko PMK*, 2023. [https://www.kemenkopmk.go.id/perlu-terobosan-dan-intervensi-tepat-sasaran-lintas-sektor-untuk-atasi-stunting#:~:text=Secara global%2C berdasarkan data UNICEF,diantara negara-negara di Asia. \(accessed Jun. 08, 2024\).](https://www.kemenkopmk.go.id/perlu-terobosan-dan-intervensi-tepat-sasaran-lintas-sektor-untuk-atasi-stunting#:~:text=Secara%20berdasarkan%20data%20UNICEF,diantara%20negara-negara%20di%20Asia.&context=secara%20global%20berdasarkan%20data%20UNICEF,diantara%20negara-negara%20di%20Asia)
4. S. Sufri, Nurhasanah, M. Jannah, T. P. Dewi, F. Sirasa, and S. Bakri, “Child Stunting Reduction in Aceh Province: Challenges and a Way Ahead,” *Matern. Child Health J.*, vol. 27, no. 5, pp. 888–901, 2023, doi: 10.1007/s10995-023-03601-y.
5. K. Kesehatan, *Survei Kesehatan Indonesia dalam Angka 2023*. Jakarta: Kementerian Kesehatan RI, 2023.
6. I. Sasongko, *Pengembangan Berkelanjutan: Penyediaan Infrastruktur Pada Kawasan Pemukiman Secara Berkelanjutan*. Surabaya: PT Muara Karya (Anggota IKAPI), 2023.
7. M. F. Ghazali, A. Aqzela, C. Gracia, R. S. Febriningtyas, and D. Wijayanti, “Analisis Geospasial Kasus Stunting menggunakan Artificial Neural Network (ANN) di Kecamatan Gadingrejo, Pringsewu-Lampung,” *Maj. Geogr. Indones.*, vol. 37, no. 1, p. 1, 2022, doi: 10.22146/mgi.70474.
8. M. F. Lais, A. Atti, R. M. Pangaribuan, and R. D. Guntur, “Model Generalized Poisson Regression (GPR) Pada Kasus Stunting Di Provinsi Nusa Tenggara Timur,” *J. Difer.*, vol. 5, no. 2, pp. 68–75, 2023, doi: 10.35508/jd.v5i2.11562.
9. M. G. L. Bele, E. M. P. Hermanto, and F. Fitriani, “Pemodelan Geographically Weighted Regression pada Kasus Stunting di Provinsi Nusa Tenggara Timur Tahun 2020,” *J. Stat. dan Apl.*, vol. 6, no. 2, pp. 179–191, 2022, doi: 10.21009/jsa.06204.
10. I. K. Hasan, Nurwan, Nur Falaq, and Muhammad Rezky Friesta Payu, “Optimization Fuzzy Geographically Weighted Clustering with Gravitational Search Algorithm for Factors Analysis Associated with Stunting,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 1, pp. 120–128, 2023, doi: 10.29207/resti.v7i1.4508.
11. L. M. Fauziah, F. N. Ihsan, R. Sighthannisa, L. V. Pangesti, T. S. Nugroho, and R. F. Putri, “Factors Affecting Stunting among Toddlers: A Case Study in West Nusa Tenggara Province,” *E3S Web Conf.*, vol. 468, pp. 1–6, 2023, doi: 10.1051/e3sconf/202346806009.
12. S. Candra Puspitasari, Sunarti, “Systematic Literature Review: Perlukah Pendekatan Spasial Dalam Penentuan Program Penanggulangan Stunting?,” *J. Kesehat. Tambusai*, vol. 4, no. 3, pp. 3378–3393, 2023, [Online]. Available: <http://journal.universitaspahlawan.ac.id/index.php/jkt/article/view/17911%0Ahttp://journal.universitaspahlawan.ac.id/index.php/jkt/article/download/17911/14061>
13. S. Pramana, R. Yordani, R. Kurniawan, and B. Yuniarto, *Dasar-Dasar Statistika Dengan Software R Konsep dan Aplikasi*. Bogor: IN MEDIA, 2016.
14. S. Ndagijimana, I. H. Kabano, E. Masabo, and J. M. Ntaganda, “Prediction of Stunting among Under-5 Children in Rwanda Using Machine Learning Techniques,” *J. Prev. Med. Public Heal.*, vol. 56, no. 1, pp.



- 41–49, 2023, doi: 10.3961/jpmph.22.388.
15. A. Asra, A. P. Utomo, M. Asikin, and N. H. Puspongoro, *Analisis Multivariabel Suatu Pengantar*. Bogor: In Media, 2017.
  16. S. Pramana, B. Yuniarto, I. Santoso, R. Nooraeni, and L. H. Suadaa, *Data Mining dengan R: Konsep dan Implementasi*, Edisi 2. Bogor: IN MEDIA, 2023.
  17. A. Yusuf, “K-Means Clustering Based on Distance Measures: Stunting Prevalence Clustering in South Kalimantan,” *2022 5th Int. Semin. Res. Inf. Technol. Intell. Syst. ISRITI 2022*, pp. 706–710, 2022, doi: 10.1109/ISRITI56927.2022.10052925.
  18. J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, 2nd Editio. San Fransisco: Elsevier, Inc, 2006.
  19. I. H. Pamungkas and S. Pramana, “Improvement Method Of Fuzzy Geographically Weighted Clustering Using Gravitational Search Algorithm,” *J. Ilmu Komput. dan Inf.*, vol. 11, no. 1, pp. 10–16, 2018.