

PERBANDINGAN MODEL *MACHINE LEARNING* UNTUK KLASTERISASI AREA RISIKO *WASTING* PADA BALITA BERBASIS CITRA SATELIT DAN *OFFICIAL STATISTIC* DI NTT

La Ode Muhammad Gazali*, Rani Nooraeni

Program Studi Komputasi Statistik, Politeknik Statistika STIS, Jakarta, Indonesia

*Penulis korespondensi: 222212696@stis.ac.id

ABSTRAK

Wasting merupakan bentuk malnutrisi akut yang masih menjadi tantangan serius di Indonesia, dengan Provinsi Nusa Tenggara Timur (NTT) mencatat prevalensi tertinggi secara nasional pada tahun 2023. Penelitian ini bertujuan untuk mengelompokkan area risiko *wasting* di NTT melalui perbandingan performa tiga metode klasterisasi, yaitu *K-Means*, *Fuzzy C-Means*, dan *Agglomerative Hierarchical*, yang memanfaatkan integrasi statistik resmi dan variabel lingkungan dari citra satelit. Analisis dilakukan pada seluruh kabupaten/kota di NTT dengan menggunakan 13 indikator sosial ekonomi, kesehatan, dan lingkungan. *Principal Component Analysis* (PCA) diterapkan untuk mengatasi multikolinearitas sehingga variabel direduksi menjadi empat komponen utama. Perbandingan kinerja model dilakukan menggunakan *Silhouette Score*, *Calinski Harabasz Index*, dan *Davies Bouldin Index*. Hasil evaluasi menunjukkan bahwa *K-Means* memberikan performa terbaik dibandingkan dua metode lainnya, ditunjukkan oleh *Silhouette Score* dan *Calinski Harabasz index* tertinggi serta *Davies Bouldin Index* terendah. Model ini menghasilkan tiga klaster risiko yang konsisten, yaitu klaster risiko tinggi yang terkonsentrasi di Pulau Sumba dan Timor, klaster risiko sedang yang mendominasi wilayah Flores, serta klaster risiko rendah yang hanya terdapat di Kota Kupang. Setiap klaster memiliki karakteristik multidimensional yang berbeda, terutama terkait kondisi sosial ekonomi, akses layanan dasar, dan faktor lingkungan.

Kata kunci: *wasting*, klasterisasi, *machine learning*, citra satelit, Nusa Tenggara Timur

1 PENDAHULUAN

Wasting merupakan salah satu bentuk malnutrisi akut yang berdampak serius terhadap morbiditas, perkembangan kognitif, dan risiko kematian pada balita (World Health Organization, 2023). Indonesia termasuk negara dengan beban *wasting* yang tinggi, dengan prevalensi nasional mencapai 8,5 persen pada tahun 2023 yang menunjukkan tren kenaikan dibandingkan tahun sebelumnya (Badan Kebijakan Pembangunan Kementerian Kesehatan, 2023). Provinsi Nusa Tenggara Timur (NTT) menempati prevalensi tertinggi secara nasional sebesar 12,3 persen, yang mencerminkan adanya tekanan multidimensional yang meliputi kondisi sosial ekonomi, pola asuh, layanan kesehatan, penyakit infeksi, serta keterbatasan sanitasi dan air bersih (Afriyani *et al.*, 2016; Suharmiati & Rochmansyah, 2018; Suratri *et al.*, 2023). Karakteristik geoklimatis NTT sebagai wilayah kepulauan dengan iklim kering turut memperberat tantangan penyediaan layanan dasar dan memperkuat kerentanan balita terhadap *wasting*.

Berbagai penelitian menunjukkan bahwa faktor sosial ekonomi, pendidikan ibu, cakupan imunisasi, sanitasi, serta kondisi lingkungan fisik berperan penting dalam mempengaruhi kejadian *wasting* (Afriyani *et al.*, 2016; Local Burden of Disease Child Growth Failure Collaborators, 2020). Namun, pemantauan faktor lingkungan fisik secara konvensional seringkali terbatas dalam cakupan wilayah dan keberlanjutannya. Perkembangan teknologi

penginderaan jauh membuka peluang untuk mengatasi keterbatasan tersebut melalui pemantauan kondisi lingkungan seperti suhu permukaan, vegetasi, dan ketersediaan air secara luas dan berkelanjutan melalui citra satelit. Informasi lingkungan tersebut dapat digunakan sebagai indikator tidak langsung yang berkaitan dengan kondisi suhu permukaan, vegetasi, serta ketersediaan air yang terbukti berkaitan dengan status gizi anak (Adha *et al.*, 2025; Maulani & Julianawati, 2022). Pendekatan ini memungkinkan integrasi data sosial ekonomi, kesehatan, dan lingkungan dalam kerangka analisis multidimensional untuk memahami variasi risiko wasting antarwilayah secara lebih komprehensif. Potensi pendekatan tersebut telah ditunjukkan dalam penelitian oleh *Local Burden of Disease Child Growth Failure Collaborators* (2020) yang memetakan kegagalan pertumbuhan anak, termasuk *wasting*, di seluruh Afrika melalui integrasi data survei kesehatan dengan variabel lingkungan yang diperoleh dari citra satelit. Walaupun demikian, riset mengenai *wasting* di Indonesia masih belum mengadopsi analisis multidimensional yang mempertimbangkan variasi sosial ekonomi, kesehatan, dan lingkungan berbasis citra satelit, sehingga pemetaan risiko wilayah masih belum optimal.

Sejumlah studi sebelumnya menerapkan berbagai metode klusterisasi untuk mengidentifikasi pola spasial risiko malnutrisi, seperti *K-Means*, *Fuzzy C-Means*, dan *hierarchical clustering* (Sanusi *et al.*, 2020; Yusuf, 2022). Namun, belum banyak penelitian yang secara khusus membandingkan performa metode-metode tersebut dalam konteks analisis *wasting* serta mengaitkannya dengan kondisi lingkungan fisik berbasis citra satelit. Kesenjangan ini menunjukkan perlunya pendekatan yang lebih komprehensif untuk mengungkap struktur risiko yang bersifat multidimensi dan berkaitan dengan konteks kewilayahan.

Dengan demikian, penelitian ini bertujuan untuk melakukan klusterisasi wilayah risiko *wasting* di Provinsi Nusa Tenggara Timur melalui perbandingan tiga metode *machine learning*, yaitu *K-Means*, *Fuzzy C-Means*, dan *Agglomerative Hierarchical method*, dengan memanfaatkan integrasi data statistik resmi dan variabel lingkungan yang diperoleh dari citra satelit. Analisis dilakukan pada seluruh wilayah administratif tingkat kabupaten/kota di Provinsi Nusa Tenggara Timur. Hasil penelitian diharapkan dapat memberikan gambaran spasial mengenai distribusi risiko yang lebih presisi serta menjadi dasar empiris dalam perancangan intervensi gizi yang lebih terarah dan berbasis wilayah.

2 METODE

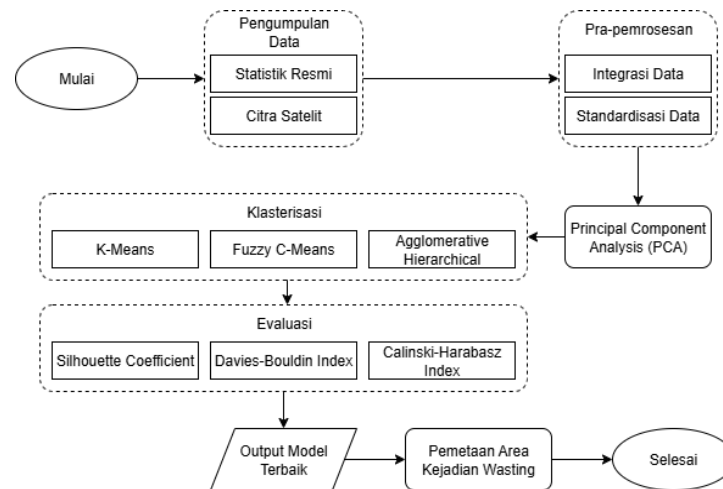
2.1 Waktu dan Tempat Penelitian

Penelitian ini menetapkan Provinsi Nusa Tenggara Timur sebagai wilayah kajian. Pemilihan wilayah ini didasarkan pada tingginya prevalensi *wasting* di provinsi tersebut serta karakteristik lingkungan yang didominasi oleh lahan kering dan kondisi sosial ekonomi yang kompleks. Unit analisis mencakup seluruh 22 kabupaten dan kota di Provinsi Nusa Tenggara Timur. Penelitian menggunakan data *cross sectional* yang merepresentasikan kondisi wilayah pada tahun 2023. Pemilihan tahun tersebut mempertimbangkan ketersediaan data yang paling lengkap dan mutakhir sehingga mampu memberikan gambaran situasi terkini.

2.2 Alur Penelitian

Penelitian ini dirancang secara sistematis yang terdiri dari beberapa tahapan utama. Proses dimulai dengan pengumpulan data yang terdiri dari statistik resmi dan citra satelit, yang kemudian melewati tahap pra-pemrosesan untuk integrasi, standardisasi, dan penanganan multikolinearitas dengan PCA. Data yang telah siap kemudian dianalisis menggunakan tiga algoritma klusterisasi yaitu *K-Means*, *Fuzzy C-Means*, dan *Agglomerative Hierarchical* untuk mengelompokkan wilayah berdasarkan faktor risiko. Hasil dari ketiga model tersebut selanjutnya dievaluasi secara kuantitatif menggunakan *Silhouette index*, *Davies-Bouldin index*,

dan *Calinski-Harabasz index* untuk memilih model terbaik. Pada tahap akhir, output dari model terbaik ini digunakan untuk melakukan pemetaan area kejadian *wasting* sebagai hasil akhir penelitian.



Gambar 1. Diagram alur penelitian

2.3 Data dan Sumber Data

Tabel 1. Ringkasan variabel

Variabel	Deskripsi	Satuan	Sumber Data
Imunisasi	Persentase penduduk usia 0-59 bulan menurut pemberian imunisasi lengkap.	Persen	BPS
ASI	Persentase penduduk usia 0-23 bulan yang pernah diberi asi	Persen	BPS
Sanitasi	Persentase rumah tangga dengan akses sanitasi layak	Persen	BPS
Akses air minum	Persentase rumah tangga dengan akses air minum layak	Persen	BPS
Prev Penyakit	Persentase penduduk yang mengalami penyakit menular	Persen	BPS
Pendidikan Ibu	Rata-rata lama sekolah (rls) wanita	Tahun	BPS
Akses Kesehatan	Rasio balita per posyandu	Rasio	BPS & Kemenkes
Kemiskinan	Persentase penduduk miskin	Persen	BPS
Buta huruf	Persentase buta huruf penduduk perempuan berumur 15 tahun ke atas	Persen	BPS
LST	<i>Land Surface Temperature</i> (Indeks suhu permukaan tanah)	Celcius	MODIS
NDVI	<i>Normalized Difference Vegetation Index</i> (Indeks vegetasi)	Tidak memiliki satuan	Landsat 8
NDWI	<i>Normalized Difference Water Index</i> (Indeks badan air)	Tidak memiliki satuan	Landsat 8

Variabel	Deskripsi	Satuan	Sumber Data
NDBI	<i>Normalized Difference Built up Index</i> (Indeks lahan terbangun)	Tidak memiliki satuan	Landsat 8

Penelitian ini mengintegrasikan dua kategori utama data sekunder, yaitu data statistik resmi dan data geospasial dari citra satelit. Variabel-variabel ini dipilih berdasarkan studi literatur mengenai faktor-faktor risiko *wasting*. Data statistik mencakup berbagai indikator sosio-ekonomi dan kesehatan, sementara data geospasial digunakan untuk menurunkan variabel-variabel yang merefleksikan kondisi lingkungan fisik.

Berdasarkan Tabel 1, terdapat 13 variabel penelitian yang digunakan, di mana sebagian besar variabel penelitian merupakan data statistik resmi yang bersumber dari Badan Pusat Statistik (BPS) dan Kementerian Kesehatan RI untuk menggambarkan kondisi sosial, ekonomi, dan kesehatan masyarakat yang menjadi faktor risiko mendasar terhadap masalah gizi.

Di sisi lain, variabel-variabel lingkungan diolah dari data citra satelit yang diakuisisi sepanjang periode 1 Januari 2023 hingga 31 Desember 2023. Data Suhu Permukaan Tanah (LST) diekstrak dari citra MODIS, sedangkan indeks vegetasi (NDVI), lahan terbangun (NDBI), dan badan air (NDWI) dihitung dari citra Landsat-8. Seluruh proses pengolahan data citra satelit ini dijalankan menggunakan skrip pada *platform Google Earth Engine* (GEE). Untuk mendefinisikan wilayah studi, batas administrasi dalam format *shapefile* (.shp) untuk setiap kabupaten/kota diunggah ke GEE. Batas ini kemudian digunakan untuk memotong citra dan memastikan analisis hanya dilakukan di dalam area yang relevan. Hasil akhir dari proses ini adalah nilai rata-rata untuk setiap variabel lingkungan pada masing-masing wilayah kabupaten/kota, yang kemudian diekspor sebagai statistik zonal untuk analisis kuantitatif lebih lanjut.

2.4 Pra-pemrosesan

Tahap pra-pemrosesan data terdiri dari tiga langkah utama, yaitu integrasi data dan standardisasi data, dan penanganan multikolinearitas. Langkah pertama adalah integrasi data, di mana seluruh data yang telah dikumpulkan dari berbagai sumber, baik data statistik resmi maupun data variabel lingkungan dari citra satelit digabungkan menjadi satu dataset tunggal. Dalam dataset ini, setiap baris merepresentasikan satu unit analisis (wilayah kabupaten/kota) dan setiap kolom mewakili satu variabel faktor risiko.

Selanjutnya, seluruh variabel dalam dataset distandardisasi menggunakan metode *Z-score normalization*. Transformasi ini mengubah setiap variabel sehingga memiliki rata-rata (mean) 0 dan standar deviasi 1. Standardisasi merupakan prasyarat penting untuk algoritma klasterisasi berbasis jarak seperti *K-Means*, karena dapat mencegah variabel dengan skala atau rentang nilai yang besar mendominasi proses perhitungan jarak secara tidak adil (Han *et al.*, 2022).

Selain itu, keterlibatan variabel yang cukup banyak dalam proses pengelompokan memungkinkan terjadinya keterkaitan yang cukup signifikan antar variabel tersebut yang disebut gejala multikolinieritas. Oleh karena itu untuk mengatasi hal tersebut maka diterapkan metode *Principal Component Analysis* (PCA), dengan cara membentuk komponen utama yang sudah saling bebas. Komponen utama yang dihasilkan akan menjadi variabel dasar pengelompokan wilayah kabupaten kota. Sebelum mengaplikasikan PCA, kelayakan data diuji menggunakan tes *Kaiser-Meyer-Olkin* (KMO) untuk mengukur kecukupan sampel dan *Bartlett's Test of Sphericity* untuk menguji keberadaan korelasi antar variabel. Komponen utama yang dihasilkan harus memenuhi minimal 80% kumulatif varians untuk digunakan sebagai input untuk model klasterisasi (Jolliffe, 2014). Penerapan PCA membawa konsekuensi

terhadap interpretasi hasil penelitian, yakni kluster yang terbentuk tidak lagi merepresentasikan pengelompokan berdasarkan variabel-variabel asli secara langsung, melainkan didasarkan pada komponen utama yang terbentuk dari variabel asal. Namun demikian, pendekatan ini tetap digunakan karena mampu mereduksi dimensi data dan mengurangi pengaruh gejala multikolinieritas.

2.5 Metode Klasterisasi

2.5.1 *K-Means*

K-Means adalah salah satu algoritma klasterisasi partisi yang paling populer dan fundamental, pertama kali diusulkan oleh MacQueen (MacQueen, 1967). Algoritma ini secara iteratif menugaskan setiap titik data ke salah satu dari K pusat kluster (*centroid*) berdasarkan jarak terdekat dengan *Euclidean Distance* dan kemudian menghitung ulang posisi *centroid* berdasarkan rata-rata dari semua titik data dalam kluster tersebut, hingga posisi *centroid* tidak lagi berubah secara signifikan. Model *K-Means* diterapkan untuk menghasilkan partisi kluster yang tegas (*hard clustering*). Penentuan jumlah kluster optimal (K) dilakukan melalui pendekatan kuantitatif dengan menganalisis *Elbow Method* pada kurva *Sum of Squared Errors* (SSE) dan memaksimalkan rata-rata *Silhouette Score* pada berbagai nilai K .

2.5.2 *Fuzzy C-Means*

Fuzzy C-Means (FCM) adalah perluasan dari *K-Means* yang dikembangkan oleh Dunn (1973) dan disempurnakan oleh Bezdek (1981) yang memungkinkan satu titik data menjadi anggota dari beberapa kluster dengan derajat keanggotaan yang berbeda. Algoritma ini bekerja dengan meminimalkan fungsi objektif yang tidak hanya bergantung pada jarak ke pusat kluster, tetapi juga pada bobot keanggotaan setiap titik data, yang diatur oleh parameter *fuzziness* (m). FCM diimplementasikan untuk memberikan perspektif *soft clustering*. Hasil dari FCM berupa matriks keanggotaan yang menunjukkan probabilitas setiap kabupaten/kota masuk ke dalam setiap kluster, sehingga dapat mengidentifikasi wilayah-wilayah dengan karakteristik risiko yang tumpang tindih.

2.5.3 *Agglomerative Hierarchical Clustering*

Agglomerative Hierarchical Clustering adalah salah satu metode klasterisasi paling fundamental dan mudah diinterpretasikan. Algoritma ini menggunakan pendekatan *bottom-up* atau dari bawah ke atas (MacQueen, 1967). Prosesnya dimulai dengan menganggap setiap titik data (dalam kasus ini, setiap kabupaten/kota) sebagai satu kluster individual. Selanjutnya, algoritma secara berulang, menghitung jarak antara semua kluster yang ada dan menggabungkan dua kluster yang paling dekat (paling mirip) menjadi satu kluster baru. Proses penggabungan ini terus berlanjut hingga semua titik data tergabung menjadi satu kluster besar tunggal.

Dalam penelitian ini, model *Agglomerative Hierarchical Clustering* digunakan untuk pendekatan berbasis hierarki. Analisis dendrogram digunakan sebagai alat utama untuk memahami bagaimana kabupaten/kota di NTT saling mengelompok secara alami. Metode pengukuran jarak antar kluster (*linkage*) yang digunakan adalah *Ward's linkage*, karena metode ini bertujuan untuk meminimalkan kenaikan total varians di dalam kluster setelah penggabungan, yang cenderung menghasilkan kluster-kluster dengan ukuran yang seimbang.

2.6 Evaluasi Hasil Klasterisasi

Kualitas partisi kluster yang dihasilkan oleh model dievaluasi menggunakan tiga metrik validasi yakni *Silhouette index*, *Davies-Bouldin index* (DBI), dan *Calinski-Harabasz* (CH) *index*. Secara kolektif, ketiga metode ini menilai performa *clustering* dengan mengukur dua aspek utama, yaitu seberapa padat (kompak) anggota di dalam suatu kluster dan seberapa jauh

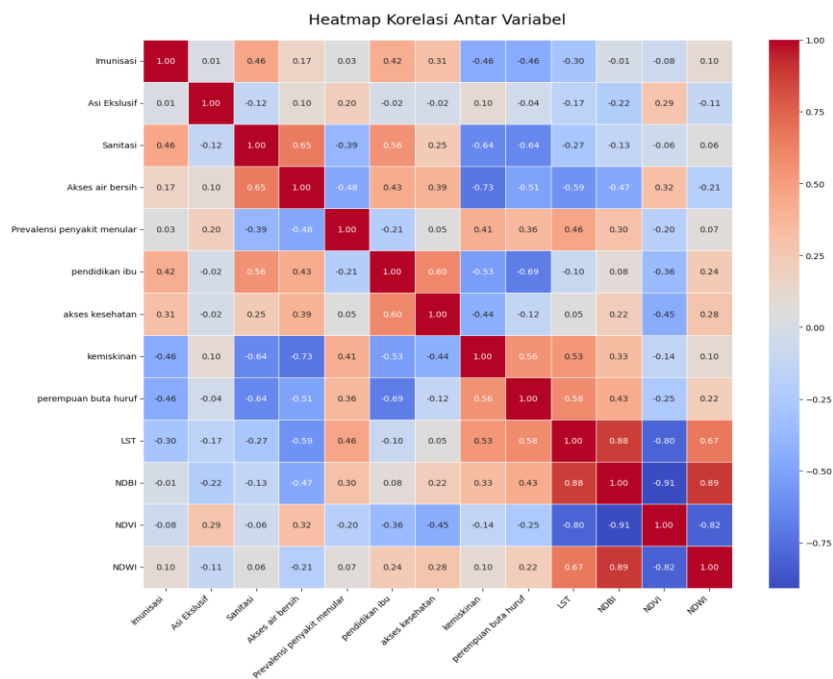
(terpisah) kluster tersebut dari kluster lainnya, tanpa memerlukan data label eksternal (Vaghela & Iyer, 2022).

Setiap metrik memiliki kriteria penilaian spesifik untuk menentukan model terbaik. Model yang optimal ditandai dengan nilai *Silhouette index* dan *Calinski-Harabasz index* yang tinggi, yang menunjukkan bahwa objek sangat cocok dengan klasternya dan kluster yang terbentuk terdefinisi dengan baik (Luchia *et al.*, 2022). Sebaliknya, untuk *Davies-Bouldin Index*, nilai yang lebih rendah dianggap lebih baik, karena ini menandakan bahwa kluster bersifat padat sekaligus terpisah jauh satu sama lain (Prasetyo & Sahala, 2014).

3 HASIL DAN PEMBAHASAN

3.1 Analisis Korelasi dan Reduksi Dimensi

Pemeriksaan awal dilakukan untuk melihat keterkaitan antar variabel sosial ekonomi, kesehatan, dan lingkungan yang digunakan dalam analisis.



Gambar 2. Heatmap korelasi antar variabel

Berdasarkan Gambar 2, hasil pemeriksaan korelasi menunjukkan adanya korelasi yang kuat ($r > 0,8$) pada beberapa variabel, terutama di antara indikator *remote sensing* dan beberapa indikator kesehatan. Kondisi ini mengindikasikan adanya potensi multikolinearitas, yang dapat menimbulkan redundansi informasi dan memengaruhi kualitas hasil klusterisasi.

Untuk mengatasi permasalahan multikolinearitas diterapkan teknik *Principal Component Analysis* (PCA). Sebelum implementasi PCA, dilakukan pengujian kelayakan data melalui dua prosedur statistik. Uji *Kaiser-Meyer-Olkin* (KMO) menghasilkan nilai yang berada di atas ambang batas 0,5, menunjukkan bahwa data memiliki kecukupan sampling yang layak untuk analisis faktor. Sementara itu, *Bartlett's Test of Sphericity* menghasilkan *p-value* yang signifikan ($p < 0,05$), mengkonfirmasi keberadaan korelasi yang memadai antar variabel untuk dilakukan pembentukan komponen utama. Berdasarkan hasil kedua uji tersebut, data dinyatakan memenuhi syarat sehingga PCA dapat diterapkan. Tabel 2 menunjukkan hasil penerapan PCA.

Tabel 2. Hasil *Principal Component Analysis* (PCA)

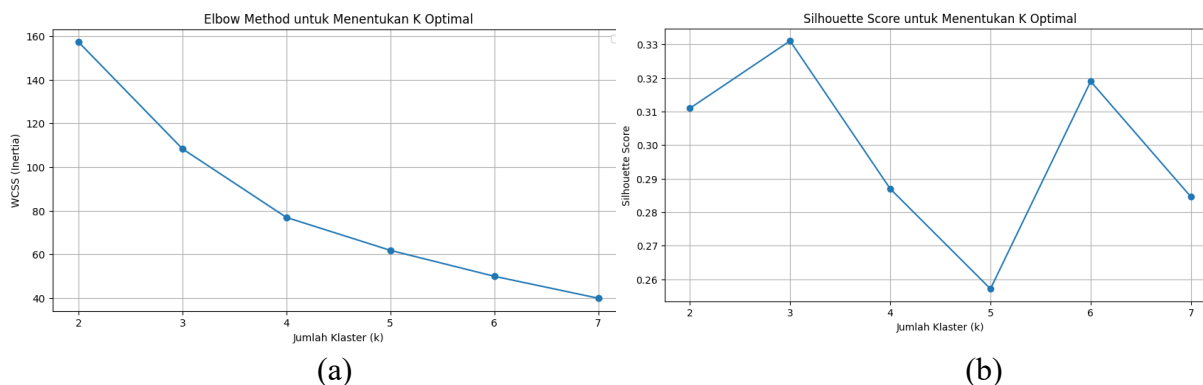
Komponen utama	Nilai Eigen	Varians (%)	Kumulatif varians (%)
1	5,0885	37,4403	37,4443
2	3,6368	26,7593	64,1996
3	1,3364	9,8329	74,0325
4	0,9778	7,1947	81,2272
5	0,8137	5,9872	87,2144
6	0,6187	4,5526	91,7670
7	0,5061	3,7237	95,4907
8	0,2499	1,8391	97,3298
9	0,1899	1,3971	98,7269
10	0,0882	0,6492	99,3762
11	0,0363	0,2673	99,6434
12	0,0303	0,2230	99,8664
13	0,0182	0,1336	100

Berdasarkan Tabel 2, hasil analisis PCA menunjukkan bahwa tiga komponen utama pertama memiliki nilai eigen di atas 1, namun varians kumulatifnya belum memenuhi ambang batas 80%, sehingga perlu adanya penambahan satu atau beberapa komponen lagi. Berdasarkan tabel 2 terlihat bahwa komponen utama sudah mencapai nilai proporsi kumulatif varians di atas 80% saat menggunakan empat komponen utama. Dengan empat komponen utama saja sudah cukup menjelaskan 81,23% dari total varians data asli. Dengan demikian, proses PCA berhasil mereduksi 13 variabel awal menjadi 4 komponen utama dan mengatasi masalah multikolinearitas untuk analisis selanjutnya.

3.2 Hasil Klasterisasi Wilayah

3.2.1 *K-Means*

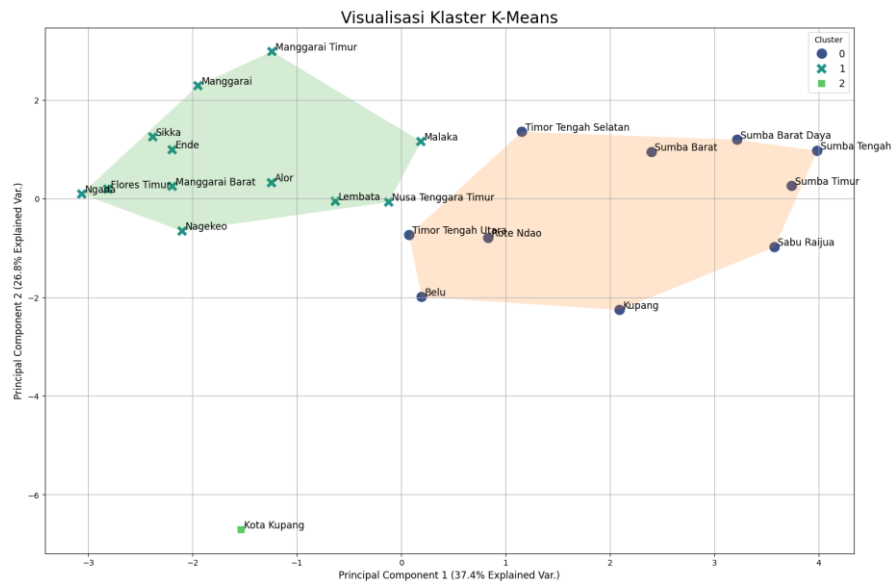
Penentuan jumlah kluster (k) optimal pada metode *K-Means* dievaluasi menggunakan *Elbow Method* dan *Silhouette Score*.



Gambar 3. Grafik *Elbow Method* (a) dan *Silhouette Score* (b) untuk model *K-Means*

Gambar 3a menunjukkan bahwa kurva inersia mengalami penurunan tajam hingga $k = 2$ dan mulai melandai pada $k = 3$. Titik ini menjadi indikasi adanya siku yang menandai jumlah kluster yang stabil. Hasil tersebut diperkuat oleh Gambar 3b yang menunjukkan nilai *Silhouette Score* tertinggi pada $k = 3$. Konsistensi kedua grafik ini menunjukkan bahwa tiga kluster merupakan konfigurasi yang paling sesuai untuk data. Hasil pengelompokan ini kemudian

divisualisasikan dalam sebuah *scatter plot* yang memetakan setiap daerah berdasarkan dua komponen utama pertama.

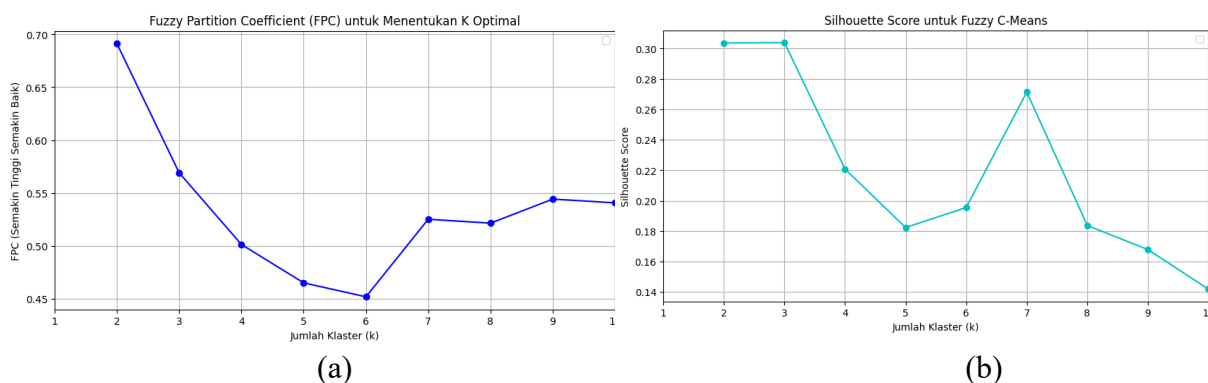


Gambar 4. Visualisasi Hasil Klasterisasi dengan Metode *K-Means* ($k=3$)

Gambar 4 menunjukkan visualisasi hasil klasterisasi menggunakan metode *K-Means* dengan tiga kluster yang diproyeksikan ke dalam dua komponen utama. Terlihat bahwa ketiga kluster terbentuk dengan pemisahan yang cukup jelas pada ruang komponen utama. Dua kluster membentuk kelompok wilayah yang relatif berdekatan, sedangkan satu kluster lainnya berada pada posisi yang lebih terpisah. Pemisahan yang tegas ini memberikan indikasi bahwa karakteristik wilayah dalam masing-masing kluster cukup berbeda.

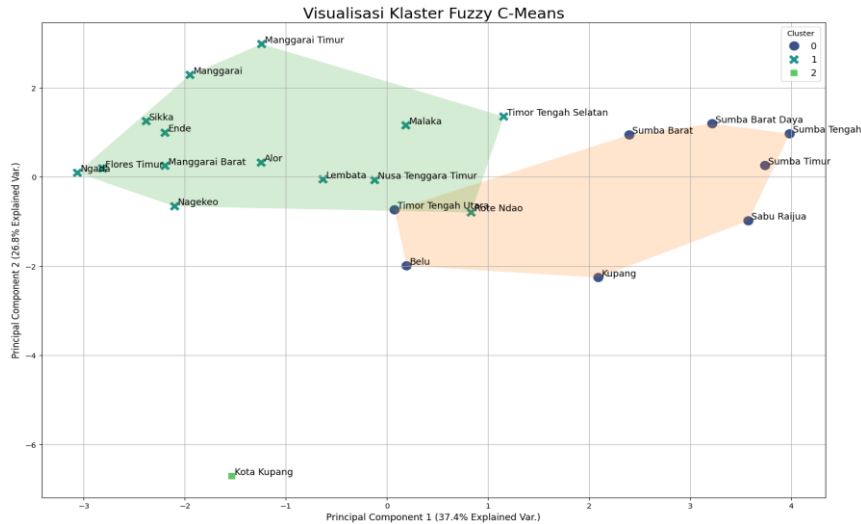
3.2.2 Fuzzy C-Means

Evaluasi jumlah kluster pada model *Fuzzy C-Means* dilakukan menggunakan *Partition Coefficient* dan *Silhouette Score*, sebagaimana ditunjukkan pada Gambar 5.



Gambar 5. Grafik *Partition Coefficient* (a) *Silhouette Score* (b) untuk model *Fuzzy C-Means*

Gambar 5a memperlihatkan nilai *Partition Coefficient* tertinggi pada $k=2$. Namun, Gambar 5b menunjukkan bahwa nilai *Silhouette Score* tertinggi berada pada $k=3$. Nilai *Silhouette* yang lebih tinggi mengindikasikan pemisahan kluster yang lebih baik pada konfigurasi tiga kluster. Oleh karena itu, jumlah kluster optimal ditetapkan sebanyak tiga. Pengelompokan yang diperoleh kemudian divisualisasikan melalui *scatter plot* yang menampilkan persebaran masing-masing daerah pada dua komponen utama pertama.

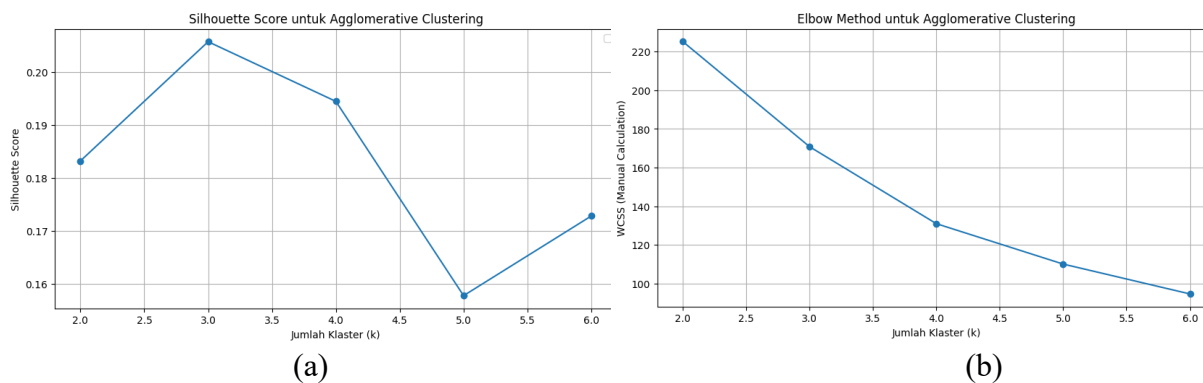


Gambar 6. Visualisasi Hasil Klasterisasi dengan Metode *Fuzzy C-Means* ($k=3$)

Berdasarkan Gambar 6, terlihat bahwa model berhasil menangkap sifat *fuzzy* pada data, di mana beberapa titik data terlihat berada di zona transisi antara dua kluster besar. Hal ini mengindikasikan adanya keanggotaan ganda pada beberapa daerah dan menunjukkan kemampuan model dalam memetakan tidak hanya kelompok yang berbeda, tetapi juga area abu-abu di antara keduanya. Selain itu, terdapat sebuah kluster tunggal yang posisinya sangat terisolasi, menandakan profilnya yang sangat berbeda.

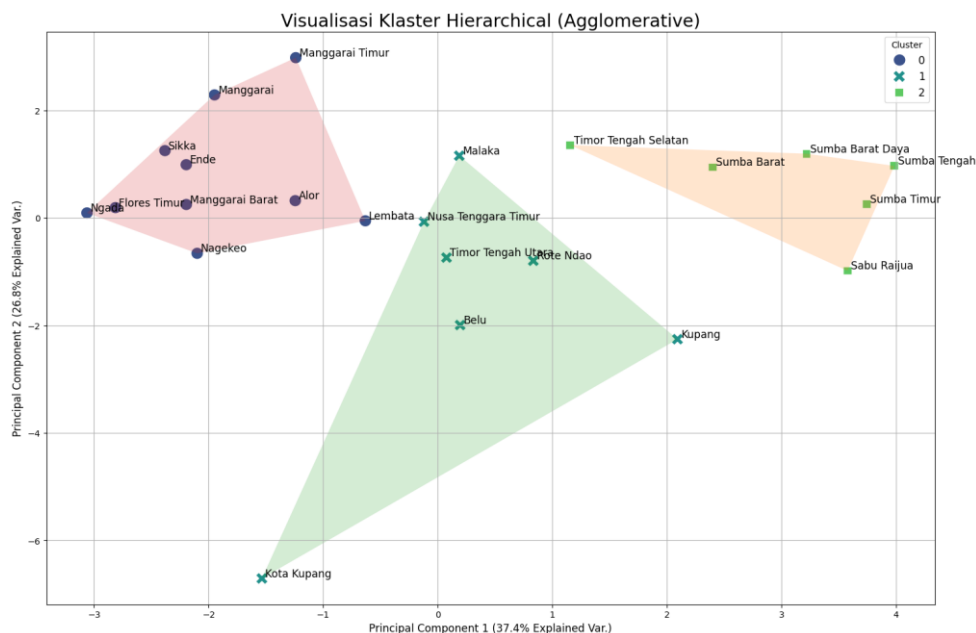
3.2.3 Agglomerative Hierarchical Clustering

Agglomerative Hierarchical Clustering diterapkan dengan menggunakan *Ward's linkage*. Pendekatan *Ward* digunakan karena mampu menggabungkan objek berdasarkan peningkatan varians yang paling minimal pada setiap tahap, sehingga menghasilkan kluster yang lebih kompak dan homogen. Evaluasi jumlah kluster untuk model ini ditunjukkan pada Gambar 7.



Gambar 7. Grafik *Elbow Method* (a) dan *Silhouette Score* (b) pada Model *Agglomerative Hierarchical*

Kurva *Elbow* pada Gambar 7a menunjukkan stabilisasi inersia pada $k=3$. Temuan ini diperkuat oleh grafik *Silhouette Score* pada Gambar 7b yang menunjukkan nilai tertinggi pada jumlah kluster yang sama. Dengan demikian, tiga kluster ditetapkan sebagai konfigurasi paling sesuai untuk model hierarki ini. Selanjutnya, hasil pengelompokan divisualisasikan dalam bentuk *scatter plot* yang memetakan daerah pada ruang dua komponen utama pertama.



Gambar 8. Visualisasi Hasil Klusterisasi dengan Metode *Agglomerative Hierarchical* ($k = 3$)

Hasil visualisasi menunjukkan tiga kelompok wilayah yang terpisah dengan baik berdasarkan dua komponen utama pertama. Struktur kluster yang terbentuk konsisten dengan model lainnya dan menunjukkan kelompok wilayah dengan kedekatan karakteristik yang serupa.

3.2.4 Perbandingan Hasil Evaluasi Metode Klusterisasi

Untuk memperoleh struktur kluster yang paling representatif, dilakukan perbandingan tiga metode pengelompokan yaitu *K-Means*, *Fuzzy C-Means*, dan *Agglomerative Hierarchical*. Ketiga metode tersebut diterapkan dengan jumlah kluster optimal masing-masing, dan kinerjanya dievaluasi menggunakan tiga metrik validasi internal guna menilai kualitas pemisahan dan kekompakan kluster.

Tabel 3. Perbandingan hasil evaluasi metode klusterisasi

Metode	<i>Silhouette Score</i>	<i>Calinski-Harabasz</i>	<i>Davies-Bouldin</i>
<i>K-Means</i>	0,331089	12,419988	0,840625
<i>Fuzzy C-Means</i>	0,303826	11,061020	1,243610
<i>Agglomerative</i>	0,315727	10,760694	0,925496

Berdasarkan Tabel 3, evaluasi menggunakan tiga metrik validasi internal menunjukkan bahwa *K-Means* memberikan hasil paling unggul dibandingkan dua metode lainnya. *K-Means* memperoleh nilai *Silhouette Score* tertinggi (0,331089), yang mengindikasikan bahwa struktur kluster memiliki pemisahan antar kelompok yang lebih baik. Nilai *Calinski Harabasz Index* yang paling tinggi (12,419988) juga menunjukkan bahwa kluster yang terbentuk memiliki rasio pemisahan dan kekompakan yang lebih optimal. Selain itu, nilai *Davies Bouldin Index* terendah (0,840625) semakin menguatkan bahwa kluster yang dihasilkan oleh *K-Means* lebih padat dan saling terpisah dengan baik.

Konsistensi performa *K-Means* pada ketiga metrik tersebut menunjukkan bahwa metode ini merupakan pendekatan klusterisasi yang paling sesuai untuk penelitian ini. Dengan demikian, *K-Means* dipilih sebagai metode utama untuk interpretasi hasil dan penentuan karakteristik masing-masing kluster wilayah.

3.3 Profiling Hasil Klasterisasi Risiko *Wasting*

Hasil klasterisasi menghasilkan tiga kelompok wilayah dengan profil risiko *wasting* yang berbeda. Struktur klaster yang terbentuk menunjukkan pemisahan yang jelas sehingga karakteristik masing-masing kelompok dapat dikenali dengan baik. Tiga klaster tersebut terdiri atas Klaster 0 (Risiko Tinggi), Klaster 1 (Risiko Sedang), dan Klaster 2 (Risiko Rendah). Meskipun klasterisasi dilakukan pada ruang komponen utama hasil PCA, karakterisasi profil setiap klaster dilakukan dengan memetakan kembali hasil klaster ke variabel asli untuk memudahkan interpretasi substantif. Rata-rata indikator pada setiap klaster disajikan pada Tabel 4.

Tabel 4. Nilai Rata-rata indikator pada setiap klaster hasil *K-Means*

Variabel	Klaster 0	Klaster 1	Klaster 2
Imunisasi	64,23	67,58	78,62
Asi	94,27	96,83	94,28
Sanitasi	68,77	78,63	89,47
Akses air minum	79,18	93,06	99,26
Prev Penyakit	0,002	0,01	0,01
Pendidikan Ibu	7,10	7,61	11,49
Akses Kesehatan	51,08	60,44	95,20
Kemiskinan	25,30	17,69	8,61
Buta huruf	9,22	4,28	1,00
LST	29,28	27,67	28,93
NDVI	0,61	0,74	0,40
NDWI	-0,58	-0,69	-0,44
NDBI	-0,08	-0,025	-0,04

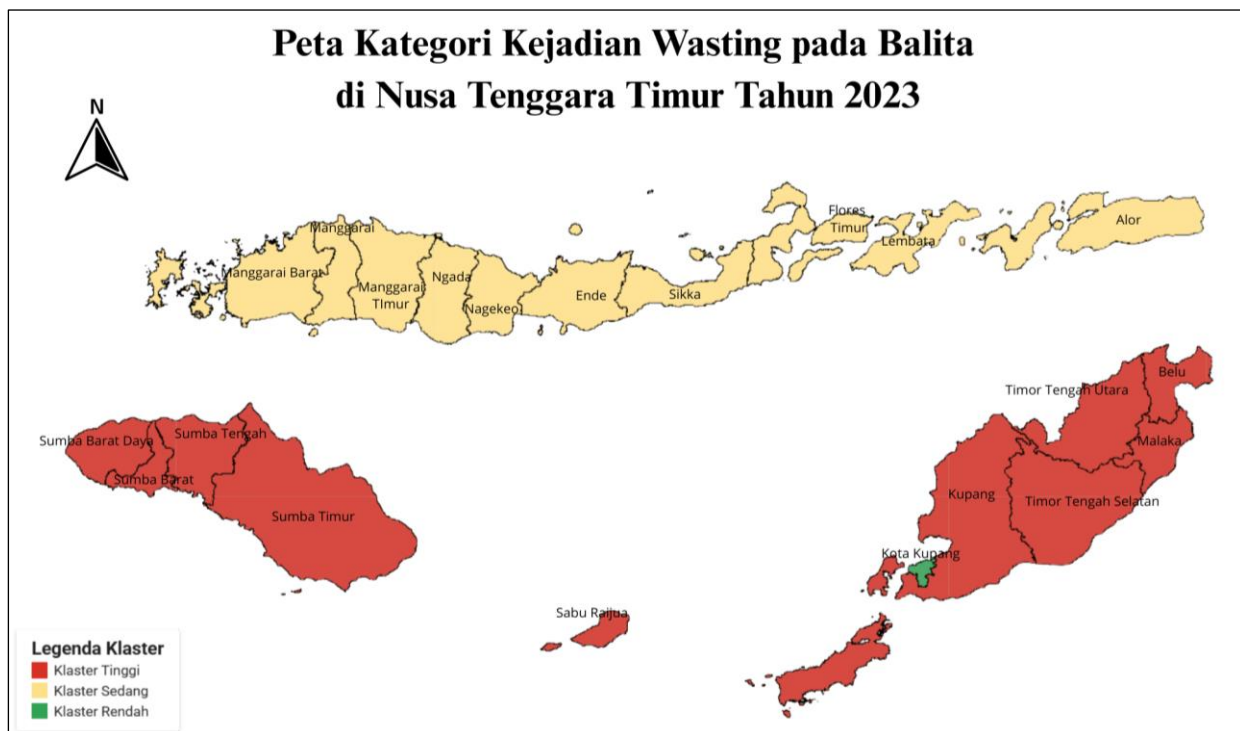
Tabel 4 memperlihatkan bahwa Klaster Risiko Tinggi dicirikan oleh kondisi sosial ekonomi dan kesehatan yang paling rendah. Wilayah dalam klaster ini memiliki tingkat kemiskinan yang tinggi, akses terendah terhadap imunisasi, sanitasi, dan layanan kesehatan, serta tingkat pendidikan ibu yang rendah. Faktor lingkungan turut memperburuk kerentanan, tercermin dari suhu permukaan tertinggi dan nilai indeks vegetasi yang tidak cukup kuat untuk mengimbangi kondisi sosial ekonomi yang lemah.

Sebaliknya, Klaster Risiko Rendah menunjukkan kondisi yang paling baik. Klaster ini memiliki kualitas sosial ekonomi tertinggi, termasuk tingkat kemiskinan yang rendah, pendidikan ibu yang lebih tinggi, serta akses terbaik terhadap sanitasi, air bersih, dan layanan kesehatan. Meskipun nilai vegetasinya relatif rendah, kapasitas sosial ekonomi yang kuat dan layanan dasar yang baik menjadi faktor protektif utama terhadap *wasting*.

Klaster Risiko Sedang berada di antara kedua kelompok tersebut dan menunjukkan karakteristik campuran. Klaster ini memiliki kondisi lingkungan yang relatif baik, khususnya nilai vegetasi yang tinggi dan suhu permukaan yang lebih rendah. Namun, faktor sosial ekonomi dan akses terhadap layanan dasar tidak sekuat Klaster Risiko Rendah, sehingga risiko *wasting* tetap berada pada tingkat menengah.

Perbedaan ketiga klaster ini memberikan implikasi langsung bagi strategi intervensi. Wilayah dalam Klaster Risiko Tinggi memerlukan peningkatan mendasar pada aspek sosial ekonomi dan layanan dasar. Klaster Risiko Sedang membutuhkan penguatan kapasitas sosial melalui peningkatan akses air, sanitasi, dan layanan kesehatan untuk memanfaatkan potensi lingkungan yang sudah baik. Klaster Risiko Rendah perlu mempertahankan capaian yang ada melalui kebijakan yang menjaga keberlanjutan layanan dasar dan adaptasi terhadap dinamika wilayah perkotaan.

3.4 Analisis Pola Sebaran Wilayah Risiko *Wasting*



Gambar 9. Peta Hasil Klaster dengan Metode Terbaik (*K-Means*)

Kombinasi faktor sosio-ekonomi dan lingkungan yang membentuk identitas unik setiap klaster memberikan wawasan tambahan ketika dipetakan secara spasial. Visualisasi spasial pada Gambar 9 memperlihatkan bahwa klaster-klaster ini tidak tersebar acak, melainkan membentuk konsentrasi geografis. Pengelompokan geografis yang begitu tegas ini memberikan indikasi kuat bahwa faktor-faktor risiko *wasting* tidak bekerja secara terisolasi di tiap kabupaten, melainkan dipengaruhi oleh konteks kewilayahan yang lebih luas. Misalnya, konsentrasi kejadian *wasting* yang tinggi terlihat menyebar hampir secara seragam di seluruh Pulau Sumba (meliputi Sumba Barat Daya, Sumba Barat, Sumba Tengah, dan Sumba Timur) dan sebagian besar daratan Pulau Timor (dari Kab. Kupang hingga Belu dan Malaka). Pola spasial ini menguatkan dugaan bahwa risiko tidak terdistribusi secara acak, melainkan menunjukkan adanya dependensi kewilayahan, di mana daerah yang berdekatan cenderung berbagi profil risiko yang serupa. Kemiripan ini kemungkinan besar didorong oleh faktor-faktor yang melintasi batas administratif, seperti kondisi agroekologis serta jejaring sosial-ekonomi dan budaya yang homogen pada tingkat regional.

Pola serupa, namun dengan karakteristik yang berbeda, terlihat di gugus Kepulauan Flores. Seluruh kabupaten di sepanjang Pulau Flores, dari Manggarai Barat di ujung barat hingga Flores Timur, beserta Lembata dan Alor, membentuk sebuah zona risiko sedang yang homogen. Keseragaman ini menunjukkan adanya kondisi-kondisi penentu yang khas untuk gugus kepulauan ini, yang membedakannya dari Sumba dan Timor. Hal ini bisa saja terkait dengan sifat ekologis vulkanik yang lebih subur, namun juga memiliki tantangan tersendiri terkait pola pembangunan ekonomi yang berada pada tahap transisi.

Kontras yang paling tajam muncul dari zona risiko rendah yang secara eksklusif ditemukan di Kota Kupang. Lokasinya yang sangat spesifik sebagai pusat urban dan administratif menandakan bahwa aglomerasi faktor-faktor positif, seperti konsentrasi fasilitas kesehatan, perputaran ekonomi yang lebih cepat, dan akses informasi yang lebih luas, menciptakan sebuah wilayah yang sangat berbeda dari wilayah di sekitarnya. Dengan

demikian, peta ini bukan sekadar menunjukkan sebaran kasus, melainkan melukiskan sebuah lanskap risiko kesehatan yang sangat terstruktur, di mana konteks geografis sebuah pulau atau wilayah tampaknya menjadi prediktor risiko yang lebih kuat daripada batas-batas administratif kabupaten semata. Namun demikian, interpretasi hasil pengelompokan wilayah di dalam penelitian ini masih bersifat deskriptif dan belum mempertimbangkan efek autokorelasi antar wilayahnya. Penelitian selanjutnya memiliki peluang untuk mengembangkan pendekatan yang lebih komprehensif guna meningkatkan ketepatan identifikasi wilayah prioritas penanganan *wasting*.

4 SIMPULAN

Penelitian ini menunjukkan bahwa integrasi data statistik resmi dan indikator lingkungan dari citra satelit dapat digunakan secara efektif untuk mengelompokkan wilayah berdasarkan tingkat risiko *wasting*. Melalui evaluasi komparatif tiga metode klasterisasi, ditemukan bahwa *K-Means* memberikan struktur klaster yang paling stabil dan informatif. Model ini menghasilkan tiga klaster risiko yang berbeda, masing-masing merepresentasikan variasi kondisi sosial ekonomi, kesehatan, dan lingkungan di Provinsi Nusa Tenggara Timur.

Pemetaan risiko yang dihasilkan memperlihatkan bahwa kerentanan *wasting* tidak tersebar secara acak, tetapi mengikuti pola kewilayahan yang konsisten. Temuan ini menegaskan bahwa konteks geografis dan karakteristik regional memainkan peran penting dalam membedakan tingkat risiko antarwilayah. Dengan demikian, hasil penelitian ini berkontribusi pada pemahaman yang lebih komprehensif mengenai distribusi risiko *wasting* serta menyediakan dasar empiris bagi perumusan intervensi yang lebih terarah.

UCAPAN TERIMAKASIH

Penulis menyampaikan terima kasih kepada Ibu Rani Nooraeni selaku dosen yang telah memberikan bimbingan, arahan, dan masukan yang sangat berarti selama proses penyusunan penelitian ini. Penulis turut berterima kasih kepada seluruh pihak yang membantu dalam penyediaan dan pengumpulan data penelitian ini. Dukungan tersebut sangat berperan dalam menyelesaikan penelitian ini dengan baik dan lancar.

DAFTAR PUSTAKA

- Adha, B. N., Rizqi, E. R., & Raiawan, A. (2025). Faktor determinan kejadian *wasting* pada balita di wilayah kerja Puskesmas Pusako Kabupaten Siak. *Jurnal Ilmu Gizi Dan Dietetik*, 4(1), 1–8. <https://doi.org/10.25182/jigd.2025.4.1.1-8>
- Afriyani, R., Malahayati, N., & Hartati, H. (2016). Faktor-faktor yang mempengaruhi kejadian *wasting* pada balita usia 1-5 tahun di Puskesmas Talang Betutu Kota Palembang. *Jurnal Kesehatan*, 7(1), 66. <https://doi.org/10.26630/jk.v7i1.120>
- Badan Kebijakan Pembangunan Kementerian Kesehatan. (2023). *SKI Dalam Angka 2023*.
- Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. Springer US. <https://doi.org/10.1007/978-1-4757-0450-1>
- Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3), 32–57. <https://doi.org/10.1080/01969727308546046>
- Han, J., Pei, J., & Tong, H. (2022). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.
- Jolliffe, I. (2014). Principal component analysis. In *Wiley StatsRef: Statistics reference online*. Wiley. <https://doi.org/10.1002/9781118445112.stat06472>
- Local burden of disease child growth failure collaborators. (2020). Mapping child growth failure across low- and middle-income countries. *Nature*, 577(7789), 231–234. <https://doi.org/10.1038/s41586-019-1878-8>
- Luchia, N. T., Handayani, H., Hamdi, F. S., Erlangga, D., & Octavia, S. F. (2022). Perbandingan

- K-Means dan K-Medoids pada pengelompokan data miskin di Indonesia. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 2(2), 35–41. <https://doi.org/10.57152/malcom.v2i2.422>
- MacQueen, J. (1967). *Some methods for classification and analysis of multivariate observations*.
- Maulani, G. R., & Julianawati, T. (2022). Pengaruh pemberian MP-ASI dan penyakit infeksi terhadap kejadian wasting pada balita usia 0-59 bulan di Kota Solok dan Kota Pariaman. *Jurnal Promotif Preventif*, 4(2), 88–93. <https://doi.org/10.47650/jpp.v4i2.363>
- Prasetyo, E., & Sahala, A. (2014). *Data Mining: mengolah data menjadi informasi menggunakan Matlab*. Andi.
- Sanusi, W., Zaky, A., & Afni, B. N. (2020). Analisis Fuzzy C-Means dan penerapannya dalam pengelompokan kabupaten/kota di Provinsi Sulawesi Selatan berdasarkan faktor-faktor penyebab gizi buruk. *Journal of Mathematics, Computations, and Statistics*, 2(1), 47. <https://doi.org/10.35580/jmathcos.v2i1.12458>
- Suharmiati, S., & Rochmansyah, R. (2018). Mengungkap kejadian infeksi kecacingan pada anak sekolah dasar (studi etnografi di Desa Taramanu Kabupaten Sumba Barat). *Buletin Penelitian Sistem Kesehatan*, 21(3), 211–217. <https://doi.org/10.22435/hsr.v21i3.420>
- Suratri, M. A. L., Putro, G., Rachmat, B., Nurhayati, Ristrini, Pracoyo, N. E., Yulianto, A., Suryatma, A., Samsudin, M., & Raharni. (2023). Risk factors for stunting among children under five years in the Province of East Nusa Tenggara (NTT), Indonesia. *International Journal of Environmental Research and Public Health*, 20(2), 1640. <https://doi.org/10.3390/ijerph20021640>
- Vaghela, R. D., & Iyer, S. S. (2022). A comparative analysis of clustering algorithm. *ECS Transactions*, 107(1), 2435–2443. <https://doi.org/10.1149/10701.2435ecst>
- World Health Organization. (2023). *Levels and trends in child malnutrition: Joint child malnutrition estimates 2023*.
- Yusuf, A. (2022). K-Means clustering based on distance measures: Stunting prevalence clustering in South Kalimantan. *2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 706–710. <https://doi.org/10.1109/ISRITI56927.2022.10052925>