MATHunesa
Jurnal Ilmiah Matematika
e-ISSN : 2716-506X | p-ISSN : 2301-9115

Volume 14 No 02
Tahun 2026

IMPLEMENTASI *K-MEANS* DAN *K-NEAREST NEIGBOR* UNTUK KLASIFIKASI STATUS GIZI BALITA (STUDI KASUS: GIZI BALITA DESA SUKA DAMAI 2024)

Ayu Risma Agustina
Program Studi Matematika , Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mulawarman, Samarinda, Indonesia, E-mail: ayurisma020804@gmail.com

Andri Azmul Fauzi
Program Studi Matematika , Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mulawarman, Samarinda, Indonesia, E-mail: andriazmul161022@fmipa.unmul.ac.id *

Syaripuddin
Program Studi Matematika , Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mulawarman, Samarinda, Indonesia, E-mail: syarifrahman2014@gmail.com

Abstrak

K-Means dan *K-Nearest Neighbor* merupakan metode kuantitatif yang digunakan dalam analisis data untuk pengelompokan dan klasifikasi. *K-Means* digunakan untuk mengelompokkan data berdasarkan kedekatan terhadap *centroid*, sedangkan *K-Nearest Neighbor* digunakan untuk mengklasifikasikan data berdasarkan tetangga terdekat. Penelitian ini bertujuan menganalisis status gizi balita di Desa Suka Damai tahun 2024 berdasarkan indikator pertumbuhan. Data yang digunakan sebanyak 149 balita meliputi umur, berat badan, tinggi badan, dan lingkar kepala. Jumlah *cluster* ditentukan sebanyak 5 kategori sesuai standar *World Health Organization*, yaitu gizi buruk, gizi kurang, gizi normal, gizi lebih, dan obesitas. Hasil pengelompokan menggunakan *K-Means* menunjukkan 23 balita termasuk kategori gizi buruk, 29 gizi kurang, 49 gizi normal, 35 gizi lebih, dan 13 obesitas. Evaluasi *cluster* menggunakan *Silhouette Coefficient* menghasilkan nilai rata-rata 0,445 yang menunjukkan kualitas *cluster* masih tergolong lemah. Selanjutnya, klasifikasi menggunakan metode *K-Nearest Neighbor* dilakukan dengan pembagian data training 80% (119 data) dan data testing 20% (30 data). Hasil evaluasi *confusion matrix* menunjukkan akurasi sebesar 100% dengan nilai *Precision*, *Recall*, dan *F1-Score* masing-masing sebesar 1. Hasil tersebut menunjukkan bahwa metode *K-Nearest Neighbor* memiliki performa sangat baik dalam mengklasifikasikan status gizi balita, sedangkan *K-Means* mampu memberikan gambaran distribusi status gizi meskipun kualitas *cluster* belum optimal.
Kata Kunci: *K-Means*, *K-Nearest Neighbor*, *Clustering*, Klasifikasi, Status Gizi Balita.

Abstract

K-Means and K-Nearest Neighbor are quantitative methods used in data analysis for clustering and classification. K-Means is used to group data based on proximity to the centroid, while K-Nearest Neighbor is used to classify data based on the nearest neighbors. This study aims to analyze the nutritional status of toddlers in Suka Damai Village in 2024 based on growth indicators. The data used consisted of 149 toddlers, including age, weight, height, and head circumference. The number of clusters was determined as five categories according to World Health Organization standards, namely severe malnutrition, undernutrition, normal nutrition, overnutrition, and obesity. The clustering results using K-Means showed that 23 toddlers were categorized as severely malnourished, 29 as undernourished, 49 as normal, 35 as overnourished, and 13 as obese. Cluster evaluation using the Silhouette Coefficient produced an average value of 0.445, indicating that the cluster quality was still relatively weak. Furthermore, classification using the K-Nearest Neighbor method was carried out by dividing the data into 80% training data (119 data points) and 20% testing data (30 data points). The confusion matrix evaluation results showed an accuracy of 100%, with Precision, Recall, and F1-Score values each reaching 1. These results indicate that the K-Nearest Neighbor method has excellent performance in classifying toddlers' nutritional status, while K-Means is capable of providing an overview of nutritional status distribution although the cluster quality is not yet optimal.
Keywords: *Clustering, Classification, K-Means, K-Nearest Neighbor, Toddler Nutritional Status.*

PENDAHULUAN

Anak-anak di bawah usia lima tahun (balita) merupakan salah satu kelompok paling rentan dalam hal kesehatan dan gizi. Masa balita adalah periode kritis sebab pertumbuhan fisik dan perkembangan kognitif berlangsung sangat cepat, sehingga kekurangan gizi pada fase ini dapa berdampak jangka panjang terhadap kualitas hidup seseorang. Gizi merupakan kebutuhan mendasar dalam proses tumbuh kembang balita, apabila asupan gizi tidak memadai, anak dapat mengalami kelainan gizi seperti stunting, gizi kurang, hingga obesitas, yang mungkin tidak segera terlihat oleh orang tua maupun masyarakat karena gejala awalnya tidak selalu jelas.

Status gizi balita bukan hanya urusan kesehatan individu, melainkan juga indikator pembangunan sumber daya manusia. Organisasi Kesehatan Dunia (*World Health Organization*/ WHO) menekankan bahwa periode 0–59 bulan (*golden age*) sangat menentukan potensi pertumbuhan masa depan. Gangguan gizi pada usia dini dapat menghambat kemampuan kognitif, menurunkan produktivitas, dan meningkatkan risiko penyakit kronis saat dewasa. Di Indonesia, data terbaru dari Survei Status Gizi Indonesia (SSGI) mengindikasikan perlunya perhatian serius: menurut laporan Kementerian Kesehatan, prevalensi stunting nasional turun menjadi 19,8% pada 2024 dari 21,5% pada 2023. Penurunan ini merupakan kemajuan, tetapi tantangan besar masih tetap ada. Pemerintah menargetkan penurunan lebih lanjut hingga 14,2% pada tahun 2029 (Kementrian Kesehatan RI, 2025).

Pemantauan pertumbuhan balita di tingkat desa melalui Pos Pelayanan Terpadu (Posyandu) umumnya dilakukan dengan pengukuran antropometri seperti berat badan, tinggi badan, dan usia. Meskipun data tersebut dikumpulkan secara rutin, proses pengolahannya masih kerap terlambat sehingga belum dimanfaatkan secara optimal untuk mendeteksi risiko gizi buruk, stunting, maupun obesitas. Di tengah keterbatasan tenaga kesehatan dan kader Posyandu, diperlukan metode analisis data yang mampu menyederhanakan proses identifikasi status gizi serta memetakan kelompok balita berdasarkan kesamaan karakteristik gizinya.

Perkembangan teknologi komputer yang semakin pesat memungkinkan berbagai kebutuhan diintegrasikan dalam suatu sistem untuk mempermudah aktivitas pengguna. Dalam konteks penilaian status gizi, pemanfaatan sistem berbasis teknologi dapat membantu proses prediksi secara lebih cepat dan akurat. Oleh karena itu, diperlukan metode analisis yang tepat untuk mendukung penentuan status gizi balita secara efisien. Salah satu metode yang sangat relevan adalah *K-Means Clustering*, sebuah algoritma pengelompokan non-hierarki yang membagi data menjadi dua kelompok atau lebih berdasarkan tingkat kesamaan karakteristiknya. Algoritma ini mengelompokkan data secara iteratif dengan meminimalkan jarak antara setiap titik data dan titik pusat *cluster*. Hasil proses pengelompokan ini kemudian dapat digunakan untuk memprediksi status gizi balita berdasarkan pola distribusi dan kedekatan karakteristik gizi dari data historis (Napitu & Hutabarat, 2024). Setelah pembentukan *cluster*, metode *K-Nearest Neighbor* (K-NN) dapat digunakan untuk klasifikasi status gizi balita baru. Prinsip kerja K-NN adalah mencari jarak terdekat antara data yang akan dievaluasi dengan tetangga (*neighbor*) terdekatnya dalam data pelatihan. Dekat atau jauhnya tetangga (*neighbor*) biasanya dihitung berdasarkan jarak *Euclidean (Euclidean Distance*)(Loka & Marsal, 2023)

Penelitian ini menerapkan teknik data *mining* dengan menggabungkan metode *K-Means* untuk mengelompokkan populasi balita berdasarkan status gizi sebagai data *training* dan K-NN untuk menentukan status individual data *testing* dengan mempertimbangkan perbandingan perhitungan jarak yang sesuai . Belum banyak studi kasus berbasis data lokal yang mengevaluasi implementasi teknis algoritma sekaligus potensi penerapannya dalam mendukung keputusan di lapangan. Sehingga penelitian ini diberi judul "Implementasi Algoritma *K-Means* dan *K-Nearest Neighbor* untuk Klasifikasi Status Gizi Balita (Studi Kasus: Gizi Balita Desa Suka Damai 2024)". Proses analisis data dilakukan melalui tahapan pra-pemrosesan, normalisasi, pembentukan *cluster* dan evaluasi hasil *cluster* data *training*, hingga klasifikasi terhadap data *testing* sehingga menghasilkan keluaran yang lebih akurat dan terstruktur. Implementasi *algoritma K-Means* dan K-NN diharapkan mampu membantu

tenaga kesehatan maupun kader Posyandu dalam mengidentifikasi kelompok balita berdasarkan kondisi gizi serta mengklasifikasikan status gizi secara lebih akurat.

KAJIAN TEORI

DATA *MINING*

Data *mining* adalah proses penggalian informasi atau pola yang berguna dari kumpulan data yang besar menggunakan teknik statistik, matematika, dan algoritma komputer. Proses ini bertujuan untuk mengidentifikasi hubungan tersembunyi, tren, atau pola yang dapat digunakan untuk pengambilan Keputusan atau prediksi di berbagai bidang, termasuk bisnis, kesehatan, dan ilmu sosial(Ananda et al., 2023). Data *mining* juga bisa dilakukan pada berbagai macam tipe database serta pemyimpanan informasi, tetapi jenis pola yang hendak ditemui ditetapkan oleh berbagai macam fungsi data *mining* seperti fungsi deskripsi, fungsi estimasi, fungsi prediksi, fungsi klasifikasi, fungi pengelompokan, dan fungsi asosiasi (Susanto & Suryadi, 2010).

Menurut (Amna et al., 2023) Data *mining* merupakan salah satu dari rangkaian *Knowledge Discovery in Database* (KDD). KDD berhubungan dengan Teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah data. Serangkaian proses tersebut memiliki proses sebagai berikut:

1. Pembersihan data untuk membuang data yang tidak konsisten dan *noise*
2. Integrasi data yang menggabungkan data dari beberapa sumber
3. Tranformasi data untuk mengubah data menjadi bentuk yang lebih sesuai
4. Aplikasi Teknik data *mining*, proses mengekstrasi pola dari data yang ada
5. Evaluasi pola yang ditemukan menjadi pengetahuan yang dapat digunakan untuk mengambil sebuah keputusan
6. Merepresentasikan hasil menjadi sebuah pengetahuan

Tahap ini merupakan bagian dari proses pencarian pengetahuan yang mencakup pemeriksaan pola dan informasi yang didapatkan bertentangan dengan fakta atau hipotesa yang sudah ada.

Pada umumnya kegunaan data *mining* dibagi menjadi dua yaitu prediktif dan deskriptif. Prediktif berarti data *mining* digunakan untuk membentuk suatu model pengetahuan yang digunakan untuk melakukan prediksi. Deskriptif berarti data *mining* digunakan utuk mencari pola-pola yang bisa dimengerti manusia yang menerangkan ciri data. Berdasarkan fungsionalitasnya, data *mining* dapat dikelompokkan kedalam enam kelompok pembelajaran yaitu aturan asosiasi, klasifikasi, *clustering*, deteksi anomali, regresi dan perangkuman (Susanto & Suryadi, 2010).

MIN-MAX NORMALIZATION

Min-max Normalization adalah proses tranformasi pada data awal menjadi data dengan interval tertentu. *Min-max normalization* bertujuan untuk menskalakan objek data pada rentang 0 sampai 1 agar tidak ada rentang jarak yang terlalu besar antar data (Iqbal et al., 2023). *Min-max normalization* dapat dirumuskan sebagai berikut:

$$x' = \left(\frac{x - X_{min}}{X_{max} - X_{min}}\right) \quad (1)$$

Dengan:

x = Nilai data
x′ = Nilai baru data x hasil normalisasi
X_{min} = Nilai minimum variabel x
X_{max} = Nilai Maksimum variabel x

K-MEANS CLUSTERING

K-Means adalah algoritma *Clustering* yang membagi data ke dalam *k* kelompok berdasarkan jarak terdekat antara data dengan pusat *cluster* (*centroid*). Algoritma ini bekerja dengan cara memilih sejumlah *k* centroid awal, kemudian memperbaruinya secara iteratif hingga posisi centroid stabil. *K-Means* popular karena kemudahan implementasi, efisiensi komputasi, serta kemampuannya menangani dataset berukuran besar. Metode ini sudah cukup banyak digunakan dalam analisis pola, segmentasi pelanggan, *image processing*, dan berbagai aplikasi data *mining* lainnya(Permana et al., 2023).

Tahapan *clustering* dengan menggunakan algoritma *K-Means* adalah sebagai berikut:

1. Menentukan nilai *k*
   Algoritma *K-Means* memulai proses *Clustering* dengan menentukan jumlah *cluster* yang ingin di bentuk. Dalam penelitian ini nilai *k* di tentukan berdasarkan standar WHO, yaitu *k* = 5 dengan masing-masing *cluster*nya yaitu: gizi buruk, gizi kurang, gizi normal, gizi lebih dan obesitas.

2. Inisialisasi untuk pusat cluster awal (*centroid*) sebanyak *k*.
3. Menghitung jarak setiap data ke *centroid* menggunakan rumus *Euclidean Distance* berikut:

$$d(y_j, x_k) = \sqrt{\sum_{i=1}^{n} (y_{ji} - x_{ki})^2} \quad (2)$$

Dengan:

$d(y_j, x_k)$ = Jarak *Euclid* data ke pusat *cluster*
n = Banyak data
x_{ki} = Nilai pusat cluster
y_{ji} = Nilai data ke-*i*

4. Mengelompokkan data ke dalam cluster sesuai *centroid* terdekat.
5. Memperbarui tiap pusat cluster dengan menggunakan rumus rata-rata sebagai berikut:

$$c_{jk} = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ik} \quad (3)$$

Dengan:

c_{jk} = Pusat *cluster* baru
x_{ik} = Data pada cluster ke-*i*
n_j = Banyak data *cluster* ke-*i*

6. Mengulangi proses hingga posisi centroid tidak berubah signifikan atau algoritma mencapai konvergensi.

(Permana et al., 2023)

K-NEAREST NEIGHBOR

K-Nearest Neighbor (K-NN) adalah metode klasifikasi non-parametrik dan *instance-based* dalam *supervised learning*, yang menentukan kelas sebuah data baru berdasarkan kelas mayoritas dari *k* tetangga terdekatnya dalam ruang fitur. Secara teknis, data latih direpresentasikan sebagai vektor berdimensi-*n*, dan ketika sebuah data uji dimasukkan, algoritma menghitung jarak antara data uji dan seluruh data latih. Umumnya menggunakan metrik seperti jarak *Euclidean* lalu memilih *k* data dengan jarak terkecil sebagai "tetangga terdekat". Label kelas data uji kemudian diputuskan berdasarkan suara mayoritas dari kelas pada tetangga tersebut. Kelebihan signifikan dari K-NN adalah kesederhanaannya yang tidak memerlukan fase "pelatihan" eksplisit karena semua komputasi dilakukan saat prediksi, menjadikannya mudah diimplementasikan dan fleksibel diterapkan pada berbagai jenis data serta problem klasifikasi atau regresi (Siregar et al., 2023).

Tahapan klasifikasi dengan menggunakan algoritma K-NN adalah sebagai berikut::

1. Pengacakan Data: Pengacakan data dilakukan pada data training agar seluruh data memiliki peluang yang sama untuk menjadi data testing.
2. Pembagian data training dan data testing
   Proporsi yang digunakan adalah 80:20 (80% data *training* dan 20% data *testing*). Data yang berada di urutan pertama akan menjadi data *training* dan sisanya menjadi data *testing*.
3. Menghitung Jarak Antar Data
   Algoritma K-NN memulai proses klasifikasi dengan menghitung jarak antara data baru dan data yang telah ada dalam dataset. Metode perhitungan jarak yang digunakan adalah jarak Euclidean, dengan rumus sebagai berikut::

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (4)$$

   Dengan:

   $d(x, y)$ = Jarak *Euclid* data *training* ke data *testing*
   n = Banyak data
   x_i = Data *training*
   y_i = Data *testing*

4. Menentukan Tetangga Terdekat (*k*)
   Setelah jarak dihitung, langkah berikutnya adalah menentukan *k* tetangga terdekat berdasarkan nilai jarak terdekat yang telah diperoleh.
5. Setelah tetangga terdekat ditentukan, data baru akan diklasifikasikan ke dalam kelas yang paling umum di antara tetangga terdekatnya. Misalnya, jika mayoritas tetangga terdekat tergolong dalam kelas "gizi buruk", maka data baru akan diklasifikasikan ke dalam kelas tersebut. Dengan demikian, K-NN memberikan hasil klasifikasi yang didasarkan pada kemiripan pola antar data, sehingga memungkinkan algoritma ini untuk menghasilkan prediksi yang akurat dalam kondisi data yang besar dan seragam.

(Yudhana, Hartanta, et al., 2020)

STATUS GIZI BALITA

Status kesehatan balita merupakan indikator penting pembangunan Masyarakat dan nasional. Asupan nutrisi yang optimal sangat memengaruhi pertumbuhan fisik dan perkembangan kognitif anak, yang pada akhirnya membentuk kualitas sumber daya manusia untuk generasi mendatang. Di Indonesia, Pos Pelayanan Terpadu (Posyandu) memainkan peran sentral dalam memantau status gizi balita melalui pemeriksaan berat dan tinggi badan secara rutin. Namun, data yang dikumpulkan seringkali tidak dianalisis secara sistematis, sehingga menghambat dihasilkannya informasi yang dapat ditindak lanjuti untuk generasi informasi yang *actionable* untuk pengambilan keputusan intervensi gizi berbasis data pemeriksaan setiap bulan (Roza et al., 2025)

Penilaian status gizi balita umumnya mengadopsi standar *Z-Score* WHO, yang mengklasifikasikan anak-anak ke dalam kategori seperti malnutrisi, kekurangan gizi, gizi baik, kelebihan gizi, dan obesitas. Data antropometri adalah data utama yang digunakan untuk mengukur komposisi tubuh guna menentukan status gizi, dengan indikator utama termasuk Berat Badan menurut Usia (BB/U), Tinggi Badan menurut Usia (TB/U), dan Berat Badan menurut Tinggi Badan (BB/TB). Pendekatan ini menyediakan basis data kuantitatif penting untuk penerapan algoritma pengelompokan seperti *K-Means* dan klasifikasi seperti *K-Nearest Neighbor* dalam menganalisis status gizi balita (Octaviyani et al., 2022).

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining yang memadukan teknik *unsupervised learning* dan *supervised learning*, yaitu algoritma K-Means Clustering dan K-Nearest Neighbor (K-NN). Pendekatan ini dipilih karena mampu mengolah data antropometri balita dalam jumlah besar serta menganalisis pola tersembunyi dan klasifikasi status gizi secara lebih akurat. Rancangan penelitian dimulai dari pengumpulan data antropometri balita Desa Suka Damai tahun 2024, dilanjutkan dengan tahap *preprocessing* berupa normalisasi data menggunakan metode *Min-Max Normalization*. Selanjutnya dilakukan proses clustering menggunakan K-Means dan klasifikasi menggunakan K-NN, kemudian hasil model dievaluasi menggunakan *Silhouette Coefficient* dan *Confusion Matrix* untuk memperoleh gambaran kondisi status gizi balita.

Populasi dalam penelitian ini adalah seluruh balita yang berada di Desa Suka Damai. Sampel penelitian merupakan balita yang terdaftar dalam kegiatan penimbangan dan pengukuran di Posyandu Desa Suka Damai selama tahun 2024 serta memiliki data antropometri lengkap berupa usia, berat badan, tinggi badan, dan lingkar kepala. Teknik pengambilan sampel menggunakan metode sampling jenuh, yaitu seluruh data yang memenuhi kriteria digunakan sebagai sampel penelitian. Teknik ini dipilih agar seluruh data dapat dimanfaatkan secara optimal dalam proses clustering dan klasifikasi sehingga menghasilkan analisis yang lebih akurat.

Teknik pengumpulan data dalam penelitian ini menggunakan data sekunder yang diperoleh dari buku register posyandu, rekapan penimbangan bulanan, serta catatan perkembangan balita yang dikelola kader posyandu dan bidan desa. Variabel penelitian meliputi usia (bulan), berat badan (kg), tinggi badan (cm), dan lingkar kepala (cm), dengan seluruh variabel bertipe data numerik. Data yang diperoleh kemudian diperiksa kelengkapannya dan dipersiapkan untuk proses analisis lebih lanjut.

Teknik analisis data dilakukan melalui beberapa tahapan, yaitu pengumpulan data, *preprocessing*, penentuan nilai k, proses clustering menggunakan algoritma K-Means, pembagian data training dan testing dengan rasio 80:20, serta klasifikasi menggunakan algoritma K-NN. Pada K-Means digunakan nilai k = 5 berdasarkan standar WHO yang terdiri atas kategori gizi buruk, gizi kurang, gizi normal, gizi lebih, dan obesitas. Selanjutnya dilakukan perhitungan jarak Euclidean pada data testing untuk menentukan tetangga terdekat dan menetapkan hasil klasifikasi berdasarkan mayoritas kelas. Hasil akhir penelitian kemudian dievaluasi dan diinterpretasikan untuk menarik kesimpulan mengenai status gizi balita di Desa Suka Damai.

HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini merupakan data tumbuh kembang balita yang bersumber dari arsip Posyandu di Desa Suka Damai,

Kecamatan Muara Badak, pada tahun 2024 sebanyak 149 balita. Data tersebut diperoleh dalam bentuk dokumen pencatatan rutin yang dilakukan oleh petugas Posyandu sebagai bagian dari kegiatan pemantauan kesehatan balita. Selanjutnya, seluruh data yang diperoleh akan diolah sebagai dasar untuk analisis lebih lanjut sesuai dengan metode yang diterapkan dalam penelitian ini.

Tabel 1. Data Tumbuh Kembang Balita

No	Balita	x_1	x_2	x_3	x_4
1.	M. Rafa Azka	57	14,2	101	50,9
2.	M. Ahwal Rajab	57	12,4	98,3	47,6
3.	Alifa	55	14,3	98,1	48,4
⋮	⋮	⋮	⋮	⋮	⋮
147.	Ayla	60	13,15	98,3	48
148.	Zafira	26	8,7	73,3	44
149.	Candra	60	16,4	106,1	50

Data tumbuh kembang balita Desa Suka Damai tahun 2024 yang di gunakan terdiri atas empat variabel, yaitu x_1 yang merupakan Umur balita dalam satuan bulan, x_2 yang merupakan berat badan balita dalam satuan kilogram (kg), x_3 yang merupakan tinggi badan balita dalam satuan sentimeter (cm), x_4 yang merupakan lingkar kepala balita dalam satuan sentimeter (cm). Data tersebut selanjutnya diproses melalui tahap normalisasi sebelum dilakukan pengelompokan menggunakan algoritma K-Means dan klasifikasi menggunakan algoritma K-Nearest Neighbor.

NORMALISASI DATA

Sebelum dilakukan proses clustering, seluruh data dinormalisasi menggunakan metode *Min-Max Normalization* agar setiap variabel memiliki rentang nilai yang sama, yaitu 0 hingga 1. Proses normalisasi dilakukan untuk menghindari dominasi variabel tertentu dalam perhitungan jarak pada algoritma K-Means dan K-NN. Hasil normalisasi menunjukkan bahwa data telah berada pada skala yang seragam sehingga dapat digunakan secara optimal dalam proses analisis.

Tabel 2. Hasil Normalisasi Data

No	Balita	x_1	x_2	x_3	x_4
1.	M. Rafa Azka	0,949	0,392	0,853	0,675
2.	M. Ahwal Rajab	0,949	0,323	0,810	0,551
3.	Alifa	0,915	0,396	0,807	0,581
⋮	⋮	⋮	⋮	⋮	⋮
147.	Ayla	1,000	0,352	0,810	0,556
148.	Zafira	0,424	0,181	0,415	0,415
149.	Candra	1,000	0,477	0,934	0,642

Normalisasi data sangat penting dalam penelitian berbasis jarak (*distance-based method*) karena perbedaan skala antar variabel dapat memengaruhi hasil pengelompokan dan klasifikasi. Dengan skala data yang seragam, proses pengukuran kemiripan antar data menjadi lebih objektif.

HASIL *CLUSTERING* MENGGUNAKAN *K-MEANS*

Proses *clustering* dilakukan menggunakan algoritma *K-Means* dengan jumlah *cluster* sebanyak lima kelompok berdasarkan standar WHO, yaitu gizi buruk, gizi kurang, gizi normal, gizi lebih, dan obesitas. Setelah proses iterasi mencapai kondisi konvergen, diperoleh centroid akhir yang stabil sehingga proses clustering dihentikan.

Tabel 3. *Centroid* Akhir Hasil *Clustering*

***Centriod* Baru**	x_1	x_2	x_3	x_4
C1	0,056	0,061	0,174	0,193
C2	0,225	0,188	0,431	0,442
C3	0,529	0,311	0,645	0,535
C4	0,873	0,386	0,804	0,574
C5	0,952	0,622	0,932	0,704

Berdasarkan hasil clustering, diperoleh 23 balita pada kategori gizi buruk, 29 balita kategori gizi kurang, 49 balita kategori gizi normal, 35 balita kategori gizi lebih, dan 13 balita kategori obesitas.

Tabel 4. Hasil Pengelompokan Status Gizi Balita

Cluster	**Status Gizi**	**Jumlah**
C1	Gizi Buruk	23
C2	Gizi Kurang	29
C3	Gizi Normal	49
C4	Gizi Lebih	35
C5	Obesitas	13

Hasil tersebut menunjukkan bahwa sebagian besar balita berada pada kategori gizi normal. Namun, masih ditemukan jumlah balita yang cukup tinggi pada kategori gizi buruk dan gizi kurang sehingga kondisi ini perlu menjadi perhatian dalam upaya peningkatan status gizi balita di Desa Suka Damai. Algoritma *K-Means* mampu mengelompokkan data berdasarkan kemiripan

karakteristik antropometri sehingga distribusi status gizi balita dapat terlihat dengan lebih jelas.

Kemudian kualitas hasil *clustering* dievaluasi menggunakan metode *Silhouette Coefficient*. Berdasarkan hasil evaluasi diperoleh rata-rata nilai silhouette sebesar 0,445. Nilai tersebut berada pada rentang 0,26–0,50 yang menunjukkan bahwa struktur cluster masih tergolong lemah. Hal ini mengindikasikan bahwa pemisahan antar cluster belum optimal dan masih terdapat kemungkinan tumpang tindih antar kelompok. Selain itu, terdapat beberapa data dengan nilai silhouette negatif yang menunjukkan bahwa beberapa objek memiliki kedekatan dengan cluster lain.

HASIL KLASIFIKASI MENGGUNAKAN *K-NEAREST NEIGHBOR*

Setelah proses *clustering* selesai, data diklasifikasikan menggunakan algoritma *K-Nearest Neighbor* (K-NN). Data dibagi menjadi data *training* sebesar 80% dan data *testing* sebesar 20%. Proses klasifikasi dilakukan berdasarkan kedekatan jarak *Euclidean* dengan nilai k = 5.

Hasil pengujian menggunakan *confusion matrix* menunjukkan bahwa seluruh data *testing* berhasil diklasifikasikan dengan benar pada masing-masing kelas status gizi.

Tabel 5. *Confusion Matrix* Klasifikasi Status Gizi

Aktual	Prediksi				
	Gizi Buruk	**Gizi Kurang**	**Gizi Normal**	**Gizi Lebih**	**Obesitas**
Gizi Buruk	5	0	0	0	0
Gizi Kurang	0	6	0	0	0
Gizi Normal	0	0	10	0	0
Gizi Lebih	0	0	0	7	0
Obesits	0	0	0	0	2

Confusion matrix menunjukkan bahwa model mampu mengklasifikasikan seluruh data uji secara tepat tanpa kesalahan klasifikasi. Seluruh kelas memiliki nilai *True Positive* yang tinggi dan tidak ditemukan nilai *False Positive* maupun *False Negatif*.

Berdasarkan hasil *confusion matrix*, diperoleh nilai akurasi sebesar 100%, sedangkan nilai precision, recall, dan F1-score pada seluruh kelas masing-masing sebesar 1.

Tabel 6. Hasil Evaluasi Algoritma K-NN

Kategori	**Precission**	**Recall**	**F1-Score**
Gizi Buruk	1	1	1
Gizi Kurang	1	1	1
Gizi Normal	1	1	1
Gizi Lebih	1	1	1
Obesits	1	1	1

Hasil evaluasi menunjukkan bahwa algoritma K-NN memiliki performa yang sangat baik dalam mengklasifikasikan status gizi balita berdasarkan data antropometri yang digunakan. Nilai *precision, recall*, dan *F1-score* yang sempurna menunjukkan bahwa model mampu mengenali seluruh kelas secara akurat dan konsisten.

Temuan penelitian ini menunjukkan bahwa kombinasi algoritma *K-Means* dan K-NN dapat digunakan sebagai pendekatan yang efektif dalam analisis status gizi balita. *K-Means* mampu memberikan gambaran distribusi kelompok status gizi, sedangkan K-NN mampu melakukan klasifikasi data baru dengan tingkat akurasi yang sangat tinggi.

PENUTUP

SIMPULAN

Berdasarkan hasil penelitian Algoritma *K-Means* berhasil mengelompokkan data status gizi balita di Desa Suka Damai tahun 2024 ke dalam lima *cluster*, yaitu gizi buruk sebanyak 23 balita, gizi kurang 29 balita, gizi normal 49 balita, gizi lebih 35 balita, dan obesitas 13 balita. Hasil ini menunjukkan adanya variasi kondisi status gizi balita yang cukup beragam di Desa Suka Damai. Namun, berdasarkan nilai rata-rata *silhouette* sebesar 0,445 yang berada pada rentang 0,26–0,50, kualitas *clustering* termasuk dalam kategori lemah. Hal ini mengindikasikan bahwa pemisahan antar *cluster* belum optimal dan masih terdapat kemungkinan adanya tumpang tindih antar kelompok data.

Di sisi lain algoritma *K-Nearest Neighbor* berhasil melakukan klasifikasi status gizi balita dengan pembagian data sebesar 80% data *training* (119 data) dan 20% data *testing* (30 data). Hasil pengujian menunjukkan bahwa seluruh data uji berhasil diklasifikasikan sesuai dengan kelas sebenarnya, dengan nilai akurasi sebesar 100%, *precision, recall*, dan *F1-Score* semua bernilai sempurna yaitu 1. Hal ini menunjukkan bahwa model klasifikasi memiliki performa yang sangat baik dalam mengenali pola data berdasarkan data training yang digunakan.

Kombinasi algoritma *K-Means* dan *K-Nearest Neighbor* secara keseluruhan cukup efektif dalam membantu analisis status gizi balita di Desa Suka Damai. Meskipun hasil *clustering* masih berada pada kategori lemah, metode ini tetap mampu memberikan gambaran awal terkait pengelompokan status gizi balita. Di sisi lain, proses klasifikasi menunjukkan hasil yang sangat optimal. Dengan demikian, pendekatan ini dapat dimanfaatkan sebagai alat bantu dalam pemantauan dan pengambilan keputusan terkait penanganan status gizi balita, meskipun masih diperlukan pengembangan lebih lanjut untuk meningkatkan kualitas pengelompokan data.

SARAN

Berdasarkan penelitian yang telah dilakukan, penelitian selanjutnya disarankan untuk mencoba metode clustering lain seperti *Hierarchical Clustering* guna membandingkan metode yang lebih sesuai dengan karakteristik data. Selain itu, penambahan variabel seperti asupan gizi, kondisi ekonomi keluarga, dan riwayat kesehatan balita juga perlu dilakukan agar hasil pengelompokan dan klasifikasi status gizi menjadi lebih akurat dan representatif.

DAFTAR PUSTAKA

Amna, Wahyuddin, S., Gede Iwan Sudipa, I., Andi Putra, T. E., Jurnaidi Wahidin, A., Alfa Syukrilla, W., Khrisna Wardhani, A., Heryana, N., Indriyani, T., Willyanto Santoso Tutuk Indriyani, L., & Willyanto Santoso, L. (2023). *DATA MINING*. www.globaleksekutifteknologi.co.id

Ananda, M. R., Nurul, S. M., Fadhila, E., & Rahma, A. (2023). *Data Mining Dalam Perusahaan PT Indofood Lubuk Pakam*. *02*(1), 97–102. https://doi.org/10.47233/jemb.v2i1.1009

Siregar, A. H., Tumanggor, A. A., & Rahmadani, A. (2023). Penerapan K-Nearest Neighbors (KNN) dalam Memprediksi dan Menghitung Akurasi Data Penyakit Stroke. *Jurnal Penelitian Rumpun Ilmu Teknik*, *2*(4), 146–154. https://doi.org/10.55606/juprit.v2i4.3040

Iqbal, M., Nurul Huda, M., & Syaripuddin. (2023). *Implementasi Algoritma K-Means Clustering dengan Jarak Euclidean dalam Mengelompokkan Daerah Penyebaran COVID-19 di Kabupaten Bogor*. *2*(1), 47–56. http://jurnal.fmipa.unmul.ac.id/index.php/basis

Loka, S. K. P., & Marsal, A. (2023). *MALCOM: Indonesian Journal of Machine Learning and Computer Science Comparison Algorithm of K-Nearest Neighbor and Naïve Bayes Classifier for Classifying Nutritional Status in Toddlers Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier Untuk Klasifikasi Status Gizi Pada Balita*. *3*, 8–14.

Octaviyani, N. R., Mayasari, R., & Susilawati. (2022). Implementasi Algoritma K-Means Clustering Status Gizi Balita. *Jurnal Ilmiah Wahana Pendidikan*, *2022*(13), 370–381. https://doi.org/10.5281/zenodo.6962588

Permana, A. A., Wahyuddin, S., Willyanto, S. L., Wahyu, G., Wibowo, N., Khrisna, A., Rahmaddeni, W., Wahidin, A. J., Eka, G.,Elisawati, Y., Rizqi, R., & Abdurrasyid, W. (2023). *MACHINE LEARNING*. www.globaleksekutifteknologi.co.id

Roza, Y. B., Defit, S., & Arlis, S. (2025). Analisis Cluster Algoritma K-Means Untuk Pengelompokan Kondisi Gizi Balita Pada Posyandu. *Bulletin of Computer Science Research*, *5*(5), 1182–1187. https://doi.org/10.47065/bulletincsr.v5i5.752

Napitu, S., & Hutabarat, H. D. M. (2024). Napitu & Hutabarat, 2024. *JURNAL ILMU KOMPUTER DAN TEKNOLOGI (IKOMTI)*, (Vol. 5 No. 3 (2024): Jurnal Ilmu Komputer dan Teknologi). https://doi.org/https://doi.org/10.35960/ikomti.v5i3.1648

Susanto, S., & Suryadi, D. (2010). *PENGANTAR DATA MINING*.

Yudhana, A., Hartanta, A. J. S., & Sunardi. (2020). ALGORITMA K-NN DENGAN EUCLIDEAN DISTANCE UNTUK PREDIKSI HASIL PENGGERGAJIAN KAYU SENGON. *TRANSMISI*, *22*(4). https://doi.org/10.14710/transmisi.22.4.107-141