

# Identifikasi Status Gizi Balita Menggunakan Metode *K-Means Clustering* dan *Naïve Bayes*

Sela Ramadani <sup>1\*</sup>, Musli Yanto <sup>1</sup>, Gunadi Widi Nurcahyo <sup>1</sup>

<sup>1</sup> Fakultas Ilmu Komputer, Magister Teknik Informatika; Universitas Putra Indonesia YPTK Padang; Padang ; e-mail: [selar0706@gmail.com](mailto:selar0706@gmail.com)  
[musli\\_yanto@upiyptk.ac.id](mailto:musli_yanto@upiyptk.ac.id) , [gwidinurcahyo@gmail.com](mailto:gwidinurcahyo@gmail.com)

\* Korespondensi: e-mail: [selar0706@gmail.com](mailto:selar0706@gmail.com)

Diterima: 24 Januari 2026; Review: 05 Februari 2026; Disetujui: 20 Februari 2026

Cara sitasi: Ramadani S, Yanto S, Nurcahyo GW. 2026. Identifikasi Status Gizi Balita Menggunakan Metode *K-Means Clustering* dan *Naïve Bayes*. Vol 11(1): 1-10.

**Abstrak:** Penelitian ini dilatarbelakangi oleh pentingnya kesehatan balita sebagai indikator keberhasilan pembangunan yang berkaitan erat dengan kondisi gizi serta berdampak langsung pada pertumbuhan dan perkembangan anak. Di wilayah kerja Puskesmas Cubadak, pemantauan kondisi gizi balita dilakukan berdasarkan data antropometri, sehingga diperlukan metode analisis yang mampu mengolah dan mengelompokkan data secara tepat dan objektif. Penelitian ini bertujuan menerapkan algoritma *K-Means Clustering* dan *Naïve Bayes* dalam mengelompokkan serta mengklasifikasikan kondisi gizi balita sebagai pendukung pengambilan keputusan. Metode penelitian menggunakan pendekatan machine learning dengan data antropometri balita yang meliputi usia, jenis kelamin, berat badan, dan tinggi badan. Algoritma *K-Means Clustering* digunakan untuk mengelompokkan data berdasarkan tingkat kemiripan karakteristik, sedangkan algoritma *Naïve Bayes* digunakan untuk melakukan klasifikasi kondisi gizi balita. Hasil penelitian menunjukkan bahwa kombinasi kedua metode menghasilkan akurasi sebesar 97%, sehingga model mampu mengelompokkan dan mengklasifikasikan kondisi gizi balita secara akurat dan konsisten sebagai sistem pendukung pengambilan keputusan di Puskesmas Cubadak.

**Kata kunci:** Gizi Balita, *Machine Learning*, *K-Means Clustering*, *Naïve Bayes*, Puskesmas Cubadak.

**Abstract:** This research is motivated by the importance of toddler health as an indicator of development success that is closely related to nutritional conditions and has a direct impact on children's growth and development. In the Cubadak Community Health Center working area, monitoring of toddler nutritional conditions is carried out based on anthropometric data, so an analytical method is needed that is able to process and group data accurately and objectively. This study aims to apply the *K-Means Clustering* and *Naïve Bayes* algorithms in grouping and classifying toddler nutritional conditions as a support for decision making. The research method uses a machine learning approach with toddler anthropometric data including age, gender, weight, and height. The *K-Means Clustering* algorithm is used to group data based on the level of similarity of characteristics, while the *Naïve Bayes* algorithm is used to classify toddler nutritional conditions. The results show that the combination of the two methods produces an accuracy of 97%, so that the model is able to group and classify toddler nutritional conditions accurately and consistently as a decision support system at the Cubadak Community Health Center.

**Keywords:** Toddler Nutrition, *Machine Learning*, *K-Means Clustering*, *Naïve Bayes*, Cubadak Health Center

## 1. Pendahuluan

Masalah gizi pada anak merupakan salah satu isu kesehatan global yang memerlukan perhatian serius. Kekurangan gizi pada usia balita dapat menyebabkan pertumbuhan yang tidak optimal, daya tahan tubuh yang rendah, bahkan meningkatkan risiko kematian. Di Indonesia, permasalahan gizi masih menjadi tantangan besar, terutama kasus stunting, gizi buruk, maupun obesitas pada anak. Oleh karena itu, penentuan status gizi balita menjadi sangat penting dalam rangka mendukung program kesehatan masyarakat terutama di Kawasan wilayah puskesmas cubadak. Status gizi balita di Puskesmas Cubadak umumnya dilakukan melalui pengukuran antropometri, seperti berat badan, tinggi badan, umur, serta perhitungan indeks massa tubuh (IMT) dan *z-score*. Namun, cara manual ini sering membutuhkan waktu lama, bergantung pada tenaga ahli, dan berisiko terjadi kesalahan analisis. Dengan perkembangan teknologi informasi, khususnya di bidang *machine learning* ML, tersedia alternatif solusi berupa pemanfaatan *algoritma klasifikasi* otomatis untuk membantu mempercepat, mempermudah, dan meningkatkan akurasi penentuan status gizi balita di Puskesmas Cubadak.

Solusi dalam mengatasi keterbatasan tersebut, maka penerapan Kecerdasan buatan (*Artificial Intelligence/AI*) merupakan bidang yang semakin penting dalam perkembangan teknologi saat ini. AI adalah salah satu cabang ilmu komputer yang bertujuan untuk mengembangkan sistem atau mesin yang dapat menirukan kemampuan kognitif manusia, seperti kemampuan berpikir, belajar, serta membuat keputusan secara otomatis. Pemanfaatan teknologi dalam bidang kesehatan, khususnya teknologi informasi dan data mining, telah berkembang pesat dan memberikan kontribusi yang signifikan dalam pengolahan serta analisis data kesehatan. Salah satu cabang AI tersebut adalah *machine learning* (ML), yaitu metode komputasi dalam kecerdasan buatan (*Artificial Intelligence*) yang memungkinkan komputer belajar dari data, menemukan pola, dan membuat prediksi atau keputusan tanpa perlu diprogram secara eksplisit [1].

Salah satu metode yang banyak digunakan dalam *machine learning* adalah *clustering* yaitu teknik dalam *unsupervised learning* yang bertujuan untuk mengelompokkan data ke dalam beberapa kelompok (*cluster*) berdasarkan tingkat kemiripan atau kedekatan karakteristik [2]. Data dalam satu *cluster* memiliki kesamaan yang tinggi (*high intra-cluster similarity*), sedangkan antar *cluster* berbeda satu sama lain (*low inter-cluster similarity*) [3]. Metode ini sangat berguna untuk menemukan pola, struktur tersembunyi, atau hubungan alami dalam data yang belum teridentifikasi. Proses *clustering* dilakukan dengan menganalisis atribut-atribut tertentu (misalnya umur, berat badan, dan tinggi badan) lalu menghitung tingkat kesamaan sehingga objek yang mirip akan dikelompokkan ke dalam *cluster* yang sama [4]. *Algoritma clustering* yang populer adalah *K-Means*. Metode ini bekerja dengan cara membagi data ke dalam sejumlah klaster yang ditentukan sebelumnya, lalu menghitung pusat klaster (*centroid*) untuk meminimalkan jarak antara data dan pusat klaster [5]. Penelitian yang sama juga menjelaskan bahwa Hasil menunjukkan bahwa dari 20 balita, 6 masuk kategori obesitas, 13 gizi lebih, dan 1 balita berisiko gizi lebih. Informasi ini dapat menjadi dasar evaluasi untuk orang tua dan Posyandu dalam meningkatkan asupan gizi yang tepat bagi anak [6].

Metode lain yang banyak digunakan dalam konsep *machine learning* adalah *klasifikasi* [7]. *klasifikasi* digunakan untuk mengelompokkan atau mengkategorikan data ke dalam kelas tertentu berdasarkan atribut atau fitur yang dimiliki. Metode ini bertujuan untuk memprediksi kategori data baru yang belum diketahui kelasnya, dengan memanfaatkan pola atau informasi yang diperoleh dari data latih. Penelitian yang sama juga menjelaskan bahwa penerapan konsep ML dengan *klasifikasi* juga memberikan hasil penelitian menunjukkan bahwa metode berbasis objek lebih efektif untuk citra beresolusi tinggi, sedangkan metode berbasis pemandangan lebih akurat untuk dataset penggunaan lahan dengan kompleksitas tinggi. Salah satu *algoritma klasifikasi* yang sederhana namun efektif adalah *Naïve Bayes*. *Algoritma* ini bekerja berdasarkan *teorema Bayes* dengan asumsi independensi antar atribut. *Teorema Bayes* sendiri merupakan aturan *probabilitas* bersyarat yang digunakan untuk memprediksi kemungkinan suatu kejadian berdasarkan informasi sebelumnya (*prior probability*) [8]. Penelitian yang sama juga menjelaskan bahwa Evaluasi model menggunakan *metrik presisi*, *recall*, dan akurasi menunjukkan hasil *presisi* 85%, *recall* 91%, dan akurasi 80%, sehingga metode ini terbukti efektif dalam mendukung identifikasi stunting sejak dini [9].

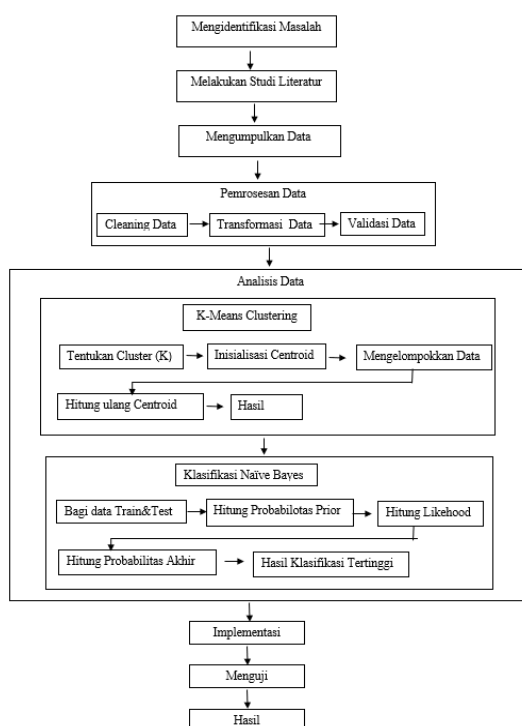
Nilai kebaruan (*Novelty*) dari penelitian ini terletak pada penggunaan kombinasi metode *K-Means Clustering* dan *Naïve Bayes* dalam penentuan status gizi balita. Pada penelitian ini, *K-Means* dimanfaatkan untuk mengelompokkan data berdasarkan kesamaan atribut antropometri,

sedangkan *Naïve Bayes* digunakan untuk membangun model *klasifikasi* yang mampu memprediksi kategori status gizi balita secara otomatis. Selain itu, penelitian ini menggunakan data balita yang diperoleh secara langsung dari wilayah kerja Puskesmas Cubadak, sehingga diharapkan dapat membantu tenaga kesehatan dalam menentukan status gizi dengan lebih cepat, tepat, dan efisien dibandingkan proses manual.

Berdasarkan uraian latar belakang tersebut, Maka Metode ini dipilih karena *K-Means* mampu mengelompokkan data anak berdasarkan kemiripan atribut. Selanjutnya, *Naïve Bayes* digunakan untuk menghasilkan model *klasifikasi* dengan tingkat akurasi yang lebih tinggi. Hasil penelitian ini diharapkan membantu tenaga kesehatan menentukan kondisi gizi anak secara cepat, akurat, dan efisien. Tujuan Penelitian ini digunakan untuk Menerapkan konsep *machine learning* dengan metode *K-Means Clustering* dan *Naïve Bayes* dalam menentukan kondisi gizi balita di puskesmas cubadak.

## 2. Metode Penelitian

Metodologi penelitian ini menjelaskan tahapan dan prosedur yang digunakan dalam penelitian untuk menentukan kondisi gizi balita menggunakan metode *K-Means Clustering* dan *Naive Bayes* pada studi kasus di Puskesmas Cubadak. Metode ini bertujuan menghasilkan model klasifikasi yang akurat dan dapat membantu tenaga kesehatan dalam mengelompokkan status gizi balita secara otomatis dan efektif. Kerangka kerja penelitian terdiri dari beberapa tahap utama mulai dari analisis kebutuhan, pengumpulan data, praproses data, penerapan metode *K-Means* untuk pengelompokan awal, pembangunan model klasifikasi menggunakan *Naive Bayes*, hingga evaluasi hasil klasifikasi. Adapun tahapan proses kerja penelitian ini dapat dilihat pada gambar 1 berikut:



Sumber: Hasil Penelitian(2026)

Gambar 1 Kerangka Kerja Penelitian

Berdasarkan kerangka kerja penelitian yang ditampilkan, tahapan penelitian dimulai dari proses identifikasi masalah untuk merumuskan permasalahan utama yang akan dikaji, yaitu perlunya pengelompokan data gizi balita menggunakan metode *K-Means Clustering* serta klasifikasi kondisi gizi dengan *algoritma Naïve Bayes*. Selanjutnya dilakukan studi literatur dengan menelaah teori terkait gizi balita, konsep *K-Means* dan *Naïve Bayes*, serta penelitian terdahulu yang relevan sebagai dasar pengembangan penelitian. Tahap berikutnya adalah pengumpulan data yang bersumber dari data primer berupa data gizi balita di Puskesmas Cubadak serta data sekunder dari laporan dan publikasi terkait. Data yang diperoleh kemudian

diproses melalui tahap preprocessing yang mencakup pembersihan, transformasi, dan validasi data untuk memastikan kualitas dataset sehingga siap digunakan dalam analisis.

Pada tahap analisis, metode *K-Means Clustering* diterapkan untuk mengelompokkan data balita berdasarkan kemiripan atribut melalui penentuan jumlah cluster, penetapan centroid awal, proses pengelompokan, pembaruan centroid, serta iterasi hingga mencapai kondisi konvergen. Hasil pengelompokan selanjutnya digunakan sebagai input pada proses klasifikasi menggunakan algoritma *Naïve Bayes*, yang meliputi pembagian data menjadi data latih dan data uji, perhitungan *probabilitas prior*, *likelihood*, dan *posterior*, serta penentuan kelas status gizi berdasarkan nilai probabilitas tertinggi. Implementasi algoritma dilakukan dengan bantuan perangkat lunak RapidMiner Studio 2026 pada spesifikasi perangkat keras dan perangkat lunak yang telah ditetapkan. Model yang dihasilkan dievaluasi menggunakan metrik kinerja seperti akurasi, *precision*, *recall*, *F1-score*, serta *confusion matrix*. Hasil akhir penelitian disajikan dalam bentuk tabel dan grafik sebagai dasar rekomendasi dalam penentuan kondisi gizi balita pada wilayah studi.

### 2.1 Kecerdasan Buatan (*Artificial Intelligence*)

Kecerdasan buatan (*Artificial Intelligence/AI*) merupakan cabang ilmu komputer yang bertujuan untuk mengembangkan sistem yang mampu meniru kecerdasan manusia, seperti belajar, berpikir, dan mengambil keputusan. AI telah diterapkan pada berbagai bidang, termasuk kesehatan, pendidikan, industri, dan sistem rekomendasi. Selain itu, AI mampu mengolah data dalam jumlah besar dengan cepat dan akurat sehingga membantu pengambilan keputusan yang lebih efektif. Penggunaan AI juga dapat meningkatkan efisiensi kerja manusia dengan mengotomatisasi proses yang kompleks. Dengan perkembangan teknologi yang pesat, AI menjadi salah satu bidang penelitian yang terus berkembang dan banyak dimanfaatkan dalam berbagai aplikasi modern [18].

### 2.2 Machine Learning

*Machine learning* merupakan subbidang AI yang memungkinkan komputer belajar dari data tanpa diprogram secara eksplisit. *Machine learning* dibagi menjadi *supervised learning*, *unsupervised learning*, *semi-supervised learning*, dan *reinforcement learning*. Dalam penelitian ini digunakan metode *unsupervised learning (clustering)* dan *supervised learning (klasifikasi)*. Pendekatan ini memungkinkan sistem untuk mengenali pola data secara otomatis berdasarkan karakteristik yang dimiliki oleh setiap objek. Proses pembelajaran model dilakukan melalui data latih sehingga sistem mampu melakukan prediksi terhadap data baru. Dengan penerapan *machine learning*, analisis data dapat dilakukan secara lebih cepat, objektif, dan akurat dibandingkan dengan metode manual [19].

### 2.3 Data Mining

*Data mining* adalah proses untuk menemukan pola atau pengetahuan baru dari data dalam jumlah besar. Tahapan data mining meliputi pengumpulan data, pembersihan data, transformasi data, pemodelan, evaluasi, dan interpretasi hasil. Proses ini bertujuan untuk menghasilkan informasi yang bermanfaat sebagai dasar pengambilan keputusan. Teknik data mining banyak digunakan dalam berbagai bidang seperti kesehatan, pendidikan, dan bisnis untuk menganalisis data secara mendalam. Dengan data mining, pola tersembunyi dalam data dapat diidentifikasi sehingga dapat memberikan insight yang akurat dan bernilai bagi pengguna [20].

### 2.4 Pengumpulan dan Analisa Data

Penelitian ini diawali dengan pengumpulan data kondisi gizi balita dari Puskesmas Cubadak periode tahun 2024 dengan jumlah 179 data balita. Data mencakup atribut jenis kelamin, usia, berat badan, tinggi badan, dan status gizi. Analisis data merupakan proses menganalisis dan mengolah data agar siap digunakan dalam penelitian. Tahap ini dilakukan setelah data dikumpulkan dan dibersihkan dari kesalahan atau data yang tidak relevan. Data yang telah siap kemudian diolah menggunakan metode yang telah ditentukan sesuai dengan tujuan penelitian. Melalui analisis data, peneliti dapat menemukan pola, hubungan, atau pengelompokan tertentu dari data yang ada.

### 2.5 K-Means Clustering

Data yang telah dikumpulkan selanjutnya dianalisis menggunakan metode *K-Means Clustering* untuk mengelompokkan kondisi gizi balita ke dalam tiga kategori, yaitu gizi buruk, obesitas, dan gizi baik. *K-Means* merupakan algoritma *clustering* bertipe partisi yang menentukan jumlah *cluster* dan nilai pusat *cluster* (*centroid*) sebagai titik awal pengelompokan. *Algoritma* ini bekerja secara iteratif hingga diperoleh hasil pengelompokan yang stabil dan optimal. Adapun tahapan-tahapan metode *k-means* sebagai berikut:

- a. Menentukan jumlah kluster yang digunakan.
- b. *inisialisasi centroid*.
- c. pengelompokan data [12].  

$$de = \sqrt{(x_i + s_i)^2 + (y_i - t_i)^2} \dots\dots\dots(1)$$

De merupakan *Euclidean Distance*, di mana i menyatakan banyak objek, (x, y) adalah koordinat objek, dan (s, t) merupakan koordinat *centroid*.
- d. Menentukan *centroid* baru  
 Penentuan *centroid* baru dilakukan dengan menghitung rata-rata nilai dari seluruh data yang termasuk dalam satu cluster yang sama. Proses ini berlangsung secara iteratif hingga posisi setiap cluster tidak lagi mengalami perubahan.  

$$C1 = (R1 + R2 + R3 \dots + Rn) / (\sum R) \dots\dots\dots(2)$$

C1 merupakan *centroid* baru yang dihitung dari R1 sebagai nilai data ke-1 hingga Rn sebagai nilai data ke-n, dengan  $\sum R$  sebagai jumlah seluruh data.
- e. Iterasi: Langkah pengelompokan dan perhitungan *centroid* diulang terus menerus hingga tidak ada perubahan.

**2.6 Naïve Bayes**

Metode *Naïve Bayes* bekerja dengan menghitung *probabilitas* suatu data termasuk ke dalam kelas tertentu berdasarkan nilai atributnya [13]. Algoritma ini sangat populer digunakan pada berbagai bidang seperti analisis teks (spam filtering, sentiment analysis), klasifikasi dokumen, prediksi kesehatan, dan banyak kasus klasifikasi lainnya, termasuk dalam pengelompokan kondisi gizi balita [14]. Adapun tahapan-tahapan metode *naïve bayes* sebagai berikut:

- a. Bagi data *training* dan *testing*
- b. Menghitung *probabilitas prior* dengan persamaan berikut:  

$$Pc_i = \frac{S_i}{S} \dots\dots\dots(3)$$

Pci merupakan *probabilitas* suatu data berada pada kelas ke-i, yang dihitung berdasarkan perbandingan antara jumlah data pada kelas tersebut dan jumlah seluruh data. Si menyatakan jumlah data yang termasuk ke dalam kelas ke-i, sedangkan S adalah jumlah keseluruhan data yang digunakan dalam proses analisis atau klasifikasi. Dengan demikian, nilai Pci dapat digunakan untuk menggambarkan peluang kemunculan data pada setiap kelas dalam dataset
- c. Menghitung *probabilitas likelihood* dengan persamaan *Gaussian* berikut:  

$$\mu = \frac{\sum x_i}{n} \dots\dots\dots(4)$$

$\mu$  merupakan nilai rata-rata (*mean*) dari seluruh data yang dihitung berdasarkan nilai setiap data yang ada. xi adalah nilai data ke-i, sedangkan n menyatakan jumlah total data yang digunakan dalam perhitungan rata-rata tersebut.

$$\sigma_c = \sqrt{\frac{\sum (x_i - \mu)^2}{n - 1}} \dots\dots\dots(5)$$

$\sigma_c$  merupakan simpangan baku pada kelas c yang menunjukkan tingkat penyebaran data terhadap nilai rata-rata. xi adalah nilai data ke-i,  $\mu$  merupakan nilai rata-rata data, dan n menyatakan jumlah seluruh data yang digunakan dalam perhitungan.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}((x-\mu)/\sigma)^2} \dots\dots\dots(6)$$

f(x) merupakan nilai *probabilitas* untuk atribut *numerik* berdasarkan nilai data yang diuji. x adalah nilai data yang diuji,  $\sigma$  adalah simpangan baku data, e merupakan bilangan eksponensial, dan  $\pi$  adalah konstanta pi dengan nilai 3,14159.

- d. Menghitung probabilitas akhir dengan persamaan berikut:

$$P(A|B) = P(B|A) * P(A|B) \dots\dots\dots(7)$$

$P(A|B)$  adalah *probabilitas* A jika diketahui B,  $P(B|A)$  adalah *probabilitas* B jika diketahui A, dan  $P(A)$  merupakan *probabilitas awal (prior)* dari kelas A.

## 2.7 Status Gizi Balita

Status gizi balita ditentukan berdasarkan indikator antropometri seperti berat badan menurut umur, tinggi badan menurut umur, dan berat badan menurut tinggi badan. Status gizi dikategorikan menjadi gizi buruk, gizi kurang, gizi baik, dan obesitas. Indikator antropometri tersebut digunakan karena mencerminkan kondisi pertumbuhan dan perkembangan fisik balita. Penilaian status gizi dilakukan dengan membandingkan hasil pengukuran dengan standar pertumbuhan yang ditetapkan oleh *World Health Organization (WHO)*. Informasi status gizi balita sangat penting sebagai dasar perencanaan intervensi kesehatan dan pemantauan pertumbuhan anak secara berkala.

## 3. Hasil dan Pembahasan

Bagian ini menyajikan hasil penelitian dan pembahasan mengenai pengelompokan serta klasifikasi kondisi gizi balita menggunakan metode *K-Means Clustering* dan *Naïve Bayes*. Hasil penelitian diperoleh melalui tahapan pengolahan data yang meliputi preprocessing, pengelompokan data, dan pengujian model klasifikasi untuk mengetahui tingkat akurasi sistem dalam menentukan status gizi balita. Penelitian ini diawali dengan pengumpulan data kondisi gizi balita dari Puskesmas Cubadak periode tahun 2024 dengan jumlah 179 data balita. Data mencakup atribut jenis kelamin, usia, berat badan, tinggi badan, dan status gizi. Selanjutnya dilakukan *preprocessing* yang meliputi pembersihan, transformasi, dan validasi data. Atribut yang kurang relevan seperti nama balita, alamat, tanggal lahir, cara ukur, dan LILA dihapus sehingga atribut utama yang digunakan adalah jenis kelamin, usia, berat badan, dan tinggi badan. *Transformasi* dilakukan dengan mengubah atribut kategorikal menjadi numerik serta mengonversi usia ke dalam satuan bulan, kemudian data divalidasi untuk memastikan kelengkapan dan konsistensi sebelum digunakan dalam pemodelan *K-Means Clustering* dan *Naïve Bayes*.

### 3.1 Pengumpulan dan Analisa Data

Data merupakan sekumpulan fakta, angka, atau informasi yang dikumpulkan dari suatu objek atau peristiwa untuk dianalisis dan diolah menjadi informasi yang bermakna. Pada penelitian ini, data utama yang digunakan adalah data puskesmas cubadak pada tahun 2024. Adapun data balita puskesmas cubadak dapat dilihat pada tabel 1.

Tabel 1. Data Balita

Data Ke	Jenis Kelamin	Umur (Bulan)	Berat	Tinggi
Data ke 1	0	58	15	99,8
Data ke 2	1	56	15	100
Data ke 3	0	59	11,5	95
Data ke 4	1	60	15,4	102,2
Data ke 5	0	56	13,3	94,2
Data ke 6	1	58	14,1	102
Data ke 7	0	60	16	103,6
Data ke 8	0	57	14,5	98,8
.....	.....	.....	.....	.....
Data ke 172	1	1	4,7	54,5
Data ke 173	0	1	3,7	51,3
Data ke 174	1	55	15,3	99
Data ke 175	0	0	2,9	49
Data ke 176	0	0	4,3	51
Data ke 177	0	44	13,7	94,8

Data Ke	Jenis Kelamin	Umur (Bulan)	Berat	Tinggi
Data ke 178	1	0	2,7	49
Data ke 179	1	0	3,7	51

Sumber: Data Balita Penelitian(2026)

Pada Tabel 1. Menampilkan data penelitian yang digunakan dalam penelitian ini Data yang ditampilkan meliputi jenis kelamin, umur dalam bulan, berat badan, dan tinggi badan untuk setiap sampel balita.masing-masing variable ini yang akan digunakan sebagai atribut untuk analisis lebih lanjut yaitu dengan proses pengelompokan *K- Means Clustering* dan *Naïve Bayes*

### 3.2 K-Means Clustering

*K-Means Clustering* adalah algoritma pengelompokan data yang membagi data ke dalam beberapa *cluster* berdasarkan kesamaan karakteristik menggunakan pusat cluster (*centroid*) [16]. *Algoritma* ini bekerja secara iteratif dengan memperbarui posisi *centroid* hingga diperoleh hasil pengelompokan yang optimal dan stabil [17]. Proses ini dilakukan dengan menentukan jumlah cluster terlebih dahulu, kemudian algoritma akan mencari pusat cluster (*centroid*) secara iteratif. Setiap data akan dihitung jaraknya terhadap *centroid* dan ditempatkan pada *cluster* terdekat. Setelah itu, posisi *centroid* akan diperbarui hingga tidak ada lagi perubahan signifikan pada hasil pengelompokan. Data yang digunakan untuk proses clustering terdiri dari jenis kelamin, umur, tinggi badan, dan berat badan. Tahapan *K-Means* meliputi penentuan jumlah *cluster*, penentuan nilai *centroid* awal, perhitungan jarak ke *centroid*, dan pengelompokan data berdasarkan jarak terdekat. Jika terjadi perubahan hasil pengelompokan, maka nilai *centroid* diperbarui hingga tidak ada perubahan lagi. Hasil perhitungan *K-Means Clustering* dapat dilihat pada tabel 2.

Tabel 2 Hasil *K-Means Clustering*

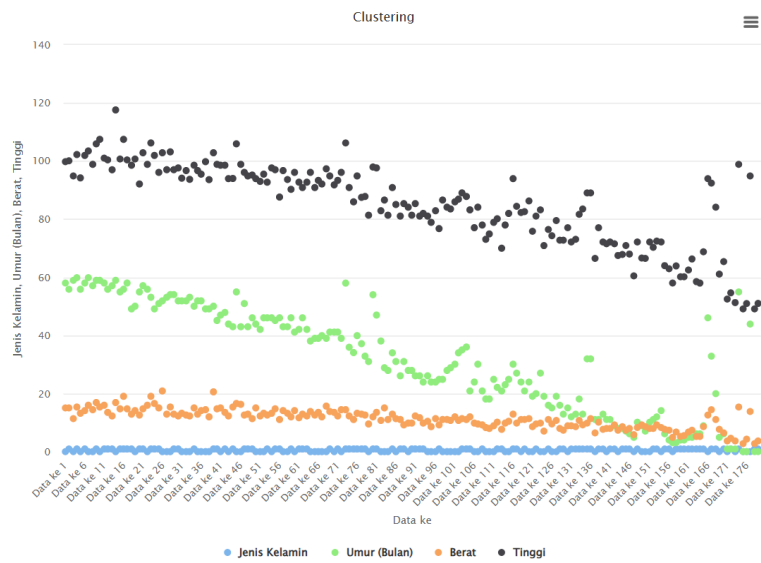
Data Ke	C0	C1	C2	Minimum	Cluster
Data ke 1	35,45389	8,841853	60,98014	8,841853	1
Data ke 2	33,80297	6,998816	59,47065	6,998816	1
Data ke 3	34,20223	10,23622	58,9086	10,23622	1
.....	.....	.....	.....	.....	.....
Data ke 77	11,18988	16,14289	36,71268	11,18988	0
Data ke 78	7,804109	19,23089	33,84063	7,804109	0
Data ke 79	4,866439	24,98759	27,95499	4,866439	0
.....	.....	.....	.....	.....	.....
Data ke 128	15,42118	42,10841	10,76869	10,76869	2
Data ke 129	17,85974	44,71475	8,630836	8,630836	2
Data ke 130	13,60297	40,52621	13,47482	13,47482	2

Sumber: Hasil Penelitian(2026)

Pada tabel 2 diperoleh hasil pembagian data ke dalam tiga *cluster*, yaitu C0 sebanyak 53 data, C1 sebanyak 79 data, dan C2 sebanyak 47 data. Tidak adanya perubahan jumlah anggota pada setiap *cluster* menunjukkan bahwa proses klasterisasi telah mencapai kondisi stabil. Dengan demikian, iterasi dihentikan karena posisi *centroid* sudah konvergen dan hasil pengelompokan dianggap optimal. *Algoritma K-Means Clustering* berhasil mengelompokkan data menjadi tiga *cluster*, yaitu:

1. *Cluster* 0 (C0) berisi 53 Data dengan kondisi gizi Obesitas
2. *Cluster* 1 (C1) berisi 79 Data dengan kondisi gizi Baik
3. *Cluster* 2 (C2) berisi 47 Data dengan kondisi gizi Buruk

Hasil klasterisasi ini memberikan gambaran nyata tentang distribusi kondisi gizi balita. Data di Puskesmas Cubadak, yang dapat dimanfaatkan oleh tenaga kesehatan dan pemangku kebijakan untuk menyusun program intervensi gizi yang lebih tepat sasaran dan berbasis data. Adapun hasil pengelompokan gizi balita divisualisasikan menggunakan *grafik* dalam *RapidMiner Visualisasi* ini mempermudah dalam melihat distribusi data balita ke dalam masing masing klaster yang terbentuk. Adapun visualisasi *k-means cluster* dalam pengelompokan gizi balita disajikan pada Gambar 1.



Sumber: Hasil Penelitian(2026)

Gambar 1 Visualisasi Hasil *K-Means Clustering*

Gambar 1 menunjukkan visualisasi hasil pengelompokan data kondisi gizi balita menggunakan metode *K-Means Clustering* pada aplikasi *RapidMiner*. Setiap warna pada grafik merepresentasikan cluster yang berbeda berdasarkan atribut jenis kelamin, umur, berat badan, dan tinggi badan balita. Visualisasi ini memperlihatkan bahwa data balita dengan karakteristik yang serupa berhasil dikelompokkan ke dalam *cluster* yang sama.

### 3.3 Naïve Bayes

*Naïve bayes* adalah *algoritma klasifikasi* berbasis *probabilitas* yang menggunakan *Teorema Bayes* dengan asumsi bahwa setiap fitur (atribut) bersifat independen satu sama lain [15]. *Naïve bayes* ini digunakan setelah dapat hasil dari *K-Means* kemudian di *naïve bayes* kan. Untuk dataset yang digunakan adalah dataset gizi balita di mana terdiri dari 4 atribut yaitu jensis kelamin, umur, tinggi, dan berat dan satu kolom target yaitu kondisi gizi balita. Hasil perhitungan *Naïve Bayes* dapat dilihat pada tabel 3 berikut:

Tabel 3 Hasil Klasifikasi *Naïve Bayes*

Data ke	Jenis Kelamin	Umur (Bulan)	Berat	Tinggi	Hasil
Data ke 144	Laki laki	8	8,6	67,9	gizi buruk
Data ke 145	Laki laki	7	7,4	71	gizi buruk
Data ke 146	Laki laki	6	7,9	68	gizi buruk
.....	.....	.....	.....	.....	.....
Data ke 166	Perempuan	46	12,5	93,8	gizi Baik
Data ke 167	Laki laki	33	14,3	92,5	Gizi Baik
Data ke 174	Laki laki	55	15,3	99	gizi baik
.....	.....	.....	.....	.....	.....
Data ke 168	Laki laki	20	10,9	84	obesitas

Sumber: Hasil Penelitian(2026)

Berdasarkan hasil klasifikasi pada Tabel 3, terlihat bahwa sebagian besar data *testing* dikategorikan dalam kelompok gizi buruk. Hal ini menunjukkan bahwa model *Naïve Bayes* mampu mengenali pola data dengan cukup konsisten terhadap karakteristik gizi balita. Beberapa data juga teridentifikasi sebagai gizi baik, menandakan adanya variasi kondisi balita dalam dataset. Secara keseluruhan, hasil prediksi ini memberikan gambaran umum tentang distribusi kondisi gizi balita pada data *testing*.

### 3.4 Evaluasi

*Evaluasi* dilakukan untuk menilai kinerja metode *K-Means Clustering* dan *Naïve Bayes* dalam mengelompokkan dan mengklasifikasikan kondisi gizi balita. Hasil *evaluasi* diperoleh dengan membandingkan hasil *prediksi model* dengan data aktual menggunakan parameter pengukuran

kinerja berupa akurasi. Tahap evaluasi ini bertujuan untuk mengetahui tingkat ketepatan *model* dalam memberikan hasil klasifikasi kondisi gizi balita. Berikut ditampilkan tabel 4 *confusion matrix*.

Tabel 4. *Confusion Matrix*

Prediksi / Aktual	Gizi Buruk	Gizi Baik	Obesitas	Precision
Pred. Gizi Buruk	31	0	0	100%
Pred. Gizi Baik	0	3	1	75%
Pred. Obesitas	0	0	1	100%
Recall	100%	100%	50%	

Sumber: Hasil Penelitian(2026)

Berdasarkan tabel 4 *confusion matrix* menunjukkan hasil klasifikasi kondisi gizi balita menggunakan *algoritma Naïve Bayes*. Berdasarkan hasil pengujian, model yang dibangun memperoleh nilai akurasi sebesar 97%, yang menandakan bahwa model memiliki performa klasifikasi yang sangat baik. Pada kategori gizi buruk, seluruh data sebanyak 31 data berhasil diklasifikasikan secara tepat tanpa adanya kesalahan, sehingga nilai *precision* dan *recall* mencapai 100%. Pada kategori gizi baik, sebanyak 3 data berhasil diklasifikasikan dengan benar, namun terdapat 1 data yang keliru diklasifikasikan sebagai obesitas, sehingga nilai *precision* pada kelas gizi baik sebesar 75% dan nilai *recall* sebesar 100%. Pada kelas obesitas, terdapat 1 data yang berhasil diklasifikasikan dengan benar, namun *recall* hanya sebesar 50% karena adanya keterbatasan jumlah data obesitas yang digunakan. Secara keseluruhan, hasil *confusion matrix* menunjukkan bahwa kombinasi metode *K-Means Clustering* dan *Naïve Bayes* mampu mengklasifikasikan kondisi gizi balita secara akurat dan konsisten, sehingga dapat digunakan sebagai sistem pendukung pengambilan keputusan dalam pemantauan status gizi balita.

#### 4. Kesimpulan

Berdasarkan hasil pengujian yang telah dilakukan terhadap kondisi gizi balita menggunakan metode *K-Means Clustering* dan *Naïve Bayes* di puskesmas cubadak dalam mengelompokkan kondisi gizi balita, maka dapat ditarik kesimpulan Penelitian ini menunjukkan bahwa penerapan metode *K-Means Clustering* dan *Naïve Bayes* dapat digunakan secara efektif dalam menentukan kondisi gizi balita. Selanjutnya, metode *Naïve Bayes* digunakan untuk melakukan klasifikasi kondisi gizi balita berdasarkan hasil pengelompokan tersebut. Hasil evaluasi kinerja menunjukkan tingkat akurasi sebesar 97%, yang menandakan bahwa model memiliki kemampuan klasifikasi yang sangat baik. Dengan demikian, kombinasi kedua metode tersebut dapat mendukung proses penentuan kondisi gizi balita secara lebih cepat, objektif, dan akurat. Penelitian selanjutnya disarankan untuk menggunakan bahasa *pemrograman Python* dalam proses pengolahan dan analisis data.

#### 5. Ucapan Terima Kasih

Peneliti menyampaikan terima kasih kepada dosen pembimbing dan pihak-pihak yang telah memberikan bimbingan serta masukan selama proses penelitian. Peneliti mengucapkan terima kasih juga kepada Puskesmas Cubadak yang telah memberikan data dan dukungan dalam pelaksanaan penelitian ini. Semoga hasil penelitian ini dapat memberikan kontribusi bagi pengembangan ilmu pengetahuan dan pengambilan keputusan di bidang kesehatan.

#### Referensi

- [1] Q. An, S. Rahman, J. Zhou, and J. J. Kang, "A Comprehensive Review on Machine Learning in Healthcare Industry: Classification, Restrictions, Opportunities and Challenges," May 01, 2023, *MDPI*. doi: 10.3390/s23094178.
- [2] M. Yanto, F. Hadi, and S. Arlis, "Determination of children's nutritional status with machine learning classification analysis approach," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 33, no. 1, pp. 303–313, Jan. 2024, doi: 10.11591/ijeecs.v33.i1.pp303-313.
- [3] L. Barua, S. Ahamed, M. Atikur Rahman, M. Rasel Mia, and M. Karam Newaz, "A Comprehensive Study of Different Clustering Algorithms Based on Big Data Analysis," *International Journal of Advances in Engineering and Management (IJAEM)*, vol. 5, pp. 799–806, 2023, doi: 10.35629/5252-0502799806.

- [4] T. Dinh *et al.*, “Data clustering: a fundamental method in data science and management,” *Data Science and Management*, Aug. 2025, doi: 10.1016/j.dsm.2025.08.001.
- [5] D. Janner Lubis and G. Karunia Gusti, “Teknois : Jurnal Ilmiah Teknologi Informasi dan Sains [58] Penerapan Algoritma Naïve Bayes Untuk Penentuan Balita Penerima Makanan Tambahan (PMT) Berdasarkan Status Gizi Di Pos Pelayanan Terpadu (POSYANDU),” vol. 13, no. 1, pp. 58–66, 2023, doi: 10.36350/jbs.v13i1.
- [6] D. Desyanti, W. Desriyati, and M. Mesran, “Implementasi Metode K-Means Clustering dalam Mengukur Tingkat Gizi Balita Berdasarkan Z-Score,” *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 3, pp. 1808–1817, Dec. 2024, doi: 10.47065/bits.v6i3.6177.
- [7] A. Abdillah, A. Thobirin, and D. Eka Wijayanti, “Klasifikasi penentuan status gizi balita dengan metode naive bayes,” *Jurnal Ilmiah Matematika*, vol. 11, no. 1, pp. 62–78, 2024, doi: 10.26555/konvergensi.30139.
- [8] O. Pahlevi, Y. Handrianto, and C. Author, “Optimasi Algoritma Naïve Bayes Berbasis Particle Swarm Optimization Untuk Klasifikasi Status Stunting,” 2024. [Online]. Available: <https://www.kaggle.com/datasets/muhtarom/stunting-dataset>.
- [9] P. Ciomas Bogor, M. Lutfi, N. Suryana, I. Saepudin, A. Husain, and S. Handayani, “Klasifikasi Status Gizi Balita Menggunakan Naïve Bayes Classification Di Kelurahan KLASIFIKASI STATUS GIZI BALITA MENGGUNAKAN NAÏVE BAYES CLASSIFICATION DI KELURAHAN PADASUKA CIOMAS BOGOR,” 2025.
- [10] R. R. Aria, “Implementasi Algoritma K-Means untuk Pengelompokan Data Imunisasi Balita dengan Metode CRISP-DM,” *remik*, vol. 9, no. 1, pp. 189–197, Jan. 2025, doi: 10.33395/remik.v9i1.14391.
- [11] S. Cahyo Auliya Rahayu, R. Dwiyanaputra, and A. Yudo Husodo, “Klustering Topik Pada Kolom Komentar Instagram Tentang Kabinet Merah Putih Menggunakan Metode K-Means (Clustering Topics in Instagram About Cabinet Merah Putih Using the K-Means Method).” [Online]. Available: <http://jtika.if.unram.ac.id/index.php/JTIKA/>
- [12] P. Ciomas Bogor, M. Lutfi, N. Suryana, I. Saepudin, A. Husain, and S. Minang, “Klasifikasi Status Gizi Balita Menggunakan Naïve Bayes Classification Di Kelurahan KLASIFIKASI STATUS GIZI BALITA MENGGUNAKAN NAÏVE BAYES CLASSIFICATION DI KELURAHAN PADASUKA CIOMAS BOGOR,” 2025.
- [13] T. Hardiani and R. N. Putri, “Implementasi Metode Naïve Bayes Classifier Untuk Klasifikasi Stunting Pada Balita,” *Digital Transformation Technology*, vol. 4, no. 1, pp. 621–627, Aug. 2024, doi: 10.47709/digitech.v4i1.4481.
- [14] E. R. D. Eko, D. Maharani, and A. K. Syahputra, “Pemanfaatan Metode Naive Bayes Untuk Klasifikasi Status Gizi Balita Pada Kelurahan Karang Anyer,” *Decode: Jurnal Pendidikan Teknologi Informasi*, vol. 4, no. 2, pp. 392–405, May 2024, doi: 10.51454/decode.v4i2.279.
- [15] U. Arfan and N. Paraga, “Perbandingan Algoritma K-Means, Naïve Bayes dan Decision Tree Dalam Memprediksi Penjualan Bahan Bakar Minyak,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 4, pp. 1379–1389, Jul. 2024, doi: 10.57152/malcom.v4i4.1566.
- [16] I. S. Tinendung and I. Zufria, “Pengelompokan Status Stunting Pada Anak Menggunakan Metode K-Means Clustering,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, no. 4, p. 2014, Oct. 2023, doi: 10.30865/mib.v7i4.6908.
- [17] G. Vardakas, I. Papakostas, and A. Likas, “Deep Clustering Using the Soft Silhouette Score: Towards Compact and Well-Separated Clusters,” Feb. 2024, [Online]. Available: <http://arxiv.org/abs/2402.00608>
- [18] K. Ofosu-Ampong, “Artificial intelligence research: A review on dominant themes, methods, frameworks and future research directions,” Jun. 01, 2024, *Elsevier B.V.* doi: 10.1016/j.teler.2024.100127.
- [19] S. Niazi, “Title: Big Data Analytics with Machine Learning: Challenges, Innovations, and Applications 38 BIG DATA ANALYTICS WITH MACHINE LEARNING: CHALLENGES, INNOVATIONS, AND APPLICATIONS.” [Online]. Available: <https://jecir.com/>
- [20] Z. Wu, “Data mining in AI: Evolution, applications, and future directions,” *Applied and Computational Engineering*, vol. 104, no. 1, pp. 1–6, Nov. 2024, doi: 10.54254/2755-2721/104/20240904.