

**PEMETAAN POLA GIZI pada KELOMPOK POPULASI:
ANALISIS CLUSTERING MENGGUNAKAN MODEL K-
MEANS BERDASARKAN TINGGI BADAN dan BERAT
BADAN**

**PROYEK UAS PEMBELAJARAN MESIN
KELAS C**



OLEH (14)

AMALIA DAMAYANTI HUSAINI

202131002

**FAKULTAS TELEMATIKA ENERGI
INSTITUT TEKNOLOGI PERUSAHAAN LISTRIK NEGARA
JAKARTA**

2024

Abstrak

Penelitian ini menggunakan metode analisis clustering, yaitu K-Means. Metode ini bertujuan untuk mengidentifikasi pola gizi dalam kelompok populasi berdasarkan tinggi badan dan berat badan menggunakan model K-Means. Kombinasi variabel tinggi badan dan berat badan dipilih sebagai parameter utama untuk analisis cluster guna memahami keragaman pola makan dan kecukupan gizi dalam masyarakat. Dataset yang digunakan mencakup informasi tentang tinggi badan, berat badan, dan data gizi lainnya dari sejumlah responden.

Metode K-Means digunakan untuk mengelompokkan individu ke dalam cluster yang memiliki karakteristik gizi serupa. Hasil analisis cluster akan memberikan pemahaman yang lebih mendalam mengenai preferensi makanan, kecukupan nutrisi, dan pola gizi pada setiap kelompok. Dengan pemetaan pola gizi ini, dapat dikembangkan strategi intervensi yang lebih spesifik dan personalisasi untuk meningkatkan kesehatan gizi Masyarakat.

Temuan penelitian ini diharapkan dapat memberikan wawasan baru dalam pengelolaan gizi masyarakat, membantu perancangan program kesehatan yang lebih efektif, serta memberikan dasar bagi kebijakan gizi yang lebih terarah. Hasil analisis cluster ini dapat menjadi panduan bagi pihak kesehatan dan pemerintah untuk menyusun program edukasi gizi yang lebih tepat sasaran, dengan mempertimbangkan variasi kebutuhan gizi antar kelompok populasi. Dengan demikian, penelitian ini berkontribusi pada upaya meningkatkan kesehatan masyarakat melalui pendekatan yang lebih terfokus dan adaptif terhadap kebutuhan gizi individu.

Kata Kunci — K-Means, Gizi, Berat Badan, Tinggi Badan

Abstract

This research uses a clustering analysis method, namely K-Means. This method aims to identify nutritional patterns in population groups based on height and weight using the K-Means model. The combination of height and weight variables was chosen as the main parameter for cluster analysis to understand the diversity of diet and nutritional adequacy in the community. The dataset used included information on height, weight and other nutritional data from a number of respondents.

The K-Means method is used to group individuals into clusters that have similar nutritional characteristics. The results of the cluster analysis will provide a deeper understanding of the food preferences, nutritional adequacy, and nutritional patterns of each group. With this nutritional pattern mapping, more specific and personalized intervention strategies can be developed to improve the nutritional health of the community.

The findings of this study are expected to provide new insights into community nutrition management, help design more effective health programs, and provide a basis for more targeted nutrition policies. The results of the cluster analysis can serve as a guide for health authorities and the government to develop more targeted nutrition education programs, taking into account variations in nutritional needs between population groups. Thus, this study contributes to efforts to improve public health through a more targeted approach.

Keywords – K-Means, Nutrition, Height, Weight

DAFTAR ISI

Abstrak.....	ii
DAFTAR ISI.....	iv
BAB I.....	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	1
1.3 Tujuan	2
1.4 Manfaat	2
BAB II	3
KAJIAN PUSTAKA.....	3
2.1 Penelitian yang Relevan	3
2.2 Pembelajaran Mesin	8
2.3 Clustering.....	8
2.4 K – Means	9
BAB III	10
HASIL DAN PEMBAHASAN	10
3.1 K – MEANS.....	10
BAB IV	28
PENUTUP	28
4.1 Kesimpulan	28
4.2 Saran	28
DAFTAR PUSTAKA	29
LAMPIRAN – LAMPIRAN	30

BAB I

PENDAHULUAN

1.1 Latar Belakang

Gizi yang sehat merupakan aspek penting untuk menjamin kesehatan masyarakat. Namun, setiap individu memiliki kebutuhan nutrisi yang berbeda-beda, dipengaruhi oleh banyak faktor berbeda, termasuk tinggi badan dan berat badan. Saat ini, pendekatan umum dalam merancang program gizi masih bersifat umum dan belum mempertimbangkan keberagaman karakteristik antar individu. Oleh karena itu, diperlukan pendekatan yang lebih spesifik untuk memahami dan memenuhi kebutuhan gizi masyarakat secara lebih akurat.

Model K-Means, sebagai metode analisis cluster, menawarkan pendekatan potensial untuk mengelompokkan individu berdasarkan karakteristik gizi, terutama tinggi dan berat badan. Dengan memetakan pola gizi suatu kelompok penduduk menggunakan model ini, kita dapat menentukan pola makan yang khas dan tingkat kecukupan gizi pada setiap kelompok. Hal ini memungkinkan pengembangan strategi intervensi yang lebih bertarget dan personal, berdasarkan kebutuhan nutrisi spesifik setiap kelompok.

Penelitian ini bertujuan untuk mengisi kesenjangan pengetahuan dalam kesehatan gizi dengan menggabungkan analisis cluster dan data tinggi dan berat badan. Hasil penelitian ini diharapkan dapat memberikan informasi baru bagi pengambil kebijakan dan praktisi kesehatan untuk mengembangkan program gizi yang lebih efektif dan adaptif berdasarkan perubahan kebutuhan gizi masyarakat.

Dengan demikian, penelitian ini tidak hanya memberikan kontribusi terhadap pengembangan ilmu pengetahuan tetapi juga memberikan dampak positif terhadap upaya pencegahan penyakit dan peningkatan kesehatan masyarakat.

1.2 Rumusan Masalah

1. bagaimana pola gizi dalam kelompok populasi dapat dianalisis berdasarkan tinggi badan dan berat badan menggunakan model K-Means ?
2. Apakah terdapat perbedaan signifikan dalam pemetaan pola gizi antar kelompok populasi berdasarkan tinggi badan dan berat badan menggunakan model K-Means ?

1.3 Tujuan

Tujuan dari penelitian ini adalah untuk menganalisis pola gizi dalam kelompok populasi berdasarkan tinggi badan dan berat badan menggunakan model K-Means. Untuk memahami secara mendalam karakteristik gizi yang muncul dalam kelompok-kelompok yang terbentuk melalui analisis cluster, sehingga dapat memberikan wawasan yang lebih akurat terkait kebutuhan gizi khas pada setiap kelompok. Penelitian ini akan menghasilkan pemetaan pola gizi pada kelompok populasi, memberikan gambaran menyeluruh mengenai preferensi makanan, kecukupan nutrisi, dan variabilitas gizi antar kelompok. Tujuan akhir adalah menyediakan dasar yang kuat untuk pengembangan strategi intervensi gizi yang lebih terfokus dan personalisasi, dengan harapan dapat meningkatkan kesehatan gizi masyarakat secara holistik.

1.4 Manfaat

- **Manfaat Bagi Praktis**

Penelitian ini memberikan wawasan yang berharga bagi para pemangku kebijakan Kesehatan dan praktisi gizi. Pemetaan pola gizi pada kelompok populasi dapat digunakan sebagai dasar untuk menyusun program gizi yang lebih efektif dan tepat sasaran, dengan mempertimbangkan kebutuhan gizi yang berbeda antar kelompok. Dengan demikian, hasil penelitian ini dapat menjadi panduan praktis bagi upaya pencegahan penyakit, peningkatan kesehatan, dan perbaikan gizi dalam skala populasi.

- **Manfaat Bagi Akademik**

Penelitian ini memberikan kontribusi yang signifikan untuk pengembangan pengetahuan di bidang Kesehatan gizi dan analisis cluster. Penelitian ini dapat membuka pintu untuk penelitian lanjutan, memberikan pemahaman lebih mendalam tentang faktor yang mempengaruhi pola gizi, serta merangsang pengembangan metodologi analisis cluster yang lebih canggih dan relevan.

BAB II

KAJIAN PUSTAKA

2.1 Penelitian yang Relevan

Untuk memperkuat hasil penelitian, pada Bab ini berisikan tentang beberapa penelitian terdahulu yang akan dibahas sebagai pembandingan serta pedoman dalam memahami dan merancang sebuah metode yang digunakan. Sebagai pembandingan penelitian maka akan dirangkum penelitian terdahulu pada Tabel 2.1 sebagai berikut :

Tabel 2.1 Perbandingan Penelitian Dengan Penelitian yang Relevan

No.	1.
Judul	Clustering of Child Nutrition Status using Hierarchical Agglomerative Clustering Algorithm in Bekasi City (Ozzi ardhianto et al., 2023)
Penulis	1. Ozzi Ardhianto 2. Muhammad Salam Asyidqi 3. Ajif Yunizar Pratama Yusuf, S.Si, M.Eng 4. Dr. Tb. Ai Munandar, S.Kom., MT
Tahun	2023
Hasil	<p>Penelitian ini bertujuan untuk menganalisis pengelompokan nutrisi bayi berdasarkan berat badan, tinggi, dan usia untuk mengidentifikasi kelompok bayi yang membutuhkan perhatian khusus mengenai asupan gizi mereka. Metode yang digunakan dalam penelitian ini adalah pengelompokan hierarkis, yang membantu dalam pengelompokan data berdasarkan kesamaan tinggi dan berat badan pada usia tertentu.</p> <p>Hasil analisis pengelompokan memberikan wawasan berharga bagi para ahli gizi, dokter anak, dan petugas kesehatan masyarakat dalam mengembangkan program intervensi yang tepat untuk meningkatkan pola makan bayi dan memenuhi kebutuhan gizi mereka.</p> <p>Analisis pengelompokan juga membantu dalam mengidentifikasi kelompok bayi yang membutuhkan perhatian khusus mengenai kebutuhan gizi mereka, sehingga meminimalkan risiko kekurangan gizi dan pertumbuhan yang tidak sehat pada bayi.</p>

	<p>Dari analisis dan pengelompokan, terlihat lebih banyak bayi dengan status “Gizi Baik” dibandingkan dengan status gizi lainnya di salah satu kecamatan di Kota Bekasi. Analisis juga menunjukkan variasi pola makan bayi berdasarkan berat, tinggi, dan usia mereka.</p> <p>Penelitian menyarankan penggunaan lebih banyak data baru untuk studi masa depan dan memberikan temuan spesifik terkait gender dan posyandu (pos layanan kesehatan terpadu) dalam hal status gizi yang berbeda.</p>
Keterkaitan Penelitian	<p>Keterkaitan dengan penelitian pada jurnal yaitu menganalisis pengelompokan nutrisi bayi berdasarkan berat badan, tinggi, dan usia untuk mengidentifikasi kelompok bayi yang membutuhkan perhatian khusus mengenai asupan gizi mereka.</p>

No.	2.
Judul	Implementation of Machine Learning Using the K-Nearest Neighbor Classification Model in Diagnosing Malnutrition in Children (Mutammimul Ula et al., 2022)
Penulis	<ol style="list-style-type: none"> 1. Mutammimul Ula 2. Ananda Faridhatul Ulva 3. Ilham Saputra 4. Mauliza 5. Ivan Maulana
Tahun	2022
Hasil	<p>Penelitian ini bertujuan untuk mendeteksi malnutrisi pada balita dan mengklasifikasikan data menggunakan metode pengelompokan k-mean. Sistem yang dibangun dapat digunakan sebagai acuan untuk memantau pertumbuhan dan perkembangan anak.</p> <p>Status gizi anak diklasifikasikan berdasarkan kriteria Berat Badan untuk Usia (W/U), Tinggi untuk Usia (TB/U), dan Berat untuk Tinggi (W/TB). Nilai data berat badan, tinggi badan, dan jenis kelamin digunakan untuk menentukan status gizi.</p> <p>Studi ini menganalisis data awal 40 anak di setiap daerah/desa, menilai status gizi mereka. Klasifikasi termasuk Malnutrisi kurang dari 3 SD (70%), Malnutrisi - 3 SD hingga $+2 < -2$ SD (80%), Good Nutrition -2 SD to +2 SD, and Over Nutrition $> SD$.</p> <p>Hasil penelitian menunjukkan bahwa teknologi model K-</p>

	NN, yang diimplementasikan dengan model klasifikasi tetangga K-terdekat, berhasil mendeteksi kekurangan gizi pada anak-anak. Deteksi didasarkan pada kriteria status gizi dan analisis data menggunakan metode KNN
Keterkaitan Penelitian	Keterkaitan pada penelitian ini yaitu mengklasifikasikan data menggunakan metode pengelompokan k-mean. Sistem yang dibangun dapat digunakan sebagai acuan untuk memantau pertumbuhan dan perkembangan anak.

No.	3.
Judul	IMPLEMENTATION OF CLUSTERING USING K-MEANS METHOD TO DETERMINE NUTRITIONAL STATUS (Nagari & Inayati, 2020)
Penulis	1. Stefanny Surya Nagari 2. Lilik Inayati
Tahun	2020
Hasil	<p>Penelitian ini bertujuan untuk mengklasifikasikan status gizi bayi di bawah 60 bulan menggunakan metode pengelompokan K-Means. Ini menggunakan data sekunder dari Ponkesdes Mayangrejo, Bojonegoro, tanpa interaksi langsung dengan subjek. Studi ini menyimpulkan bahwa status gizi bayi dapat dikelompokkan menjadi empat kelompok: 23 balita kekurangan gizi, 17 balita kekurangan gizi, 7 balita bergizi, dan 10 balita bergizi berlebihan.</p> <p>Studi ini merekomendasikan penelitian lebih lanjut menggunakan algoritma lain untuk meningkatkan akurasi dalam mengelompokkan status gizi anak-anak. Metode K-Means ditemukan untuk mengelompokkan status gizi anak-anak berdasarkan parameter berat badan menurut usia menjadi empat kelompok: status gizi buruk (23 balita), status gizi buruk (17 balita), status gizi yang baik (7 balita), dan status gizi berlebihan (10 balita).</p> <p>Analisis cluster adalah teknik multivariat yang menggunakan Analisis cluster bertujuan untuk mengelompokkan objek data dengan karakteristik serupa menjadi satu cluster. Ini membantu dalam mengidentifikasi objek dengan karakteristik serupa dan mengelompokkannya bersama-</p>

	<p>sama.</p> <p>Penerapan teknologi informasi di sektor kesehatan, seperti pengelompokan data, dapat memfasilitasi pemrosesan data dan alokasi kepada pihak berwenang untuk tindakan tindak lanjut</p>
Keterkaitan Penelitian	Keterkaitan pada penelitian ini yaitu mengklasifikasikan status gizi bayi di bawah 60 bulan menggunakan metode pengelompokan K-Means.

No.	3.
Judul	Geographic Information System Mapping and Management of Child with the Highest Nutritional Potential in Prabumulih City Using K-Means Clustering Method (Novianti et al., 2022)
Penulis	<ol style="list-style-type: none"> 1. Leni Novianti 2. Robinson Robinson 3. Ienda Meiriska 4. Resti Atika Sari
Tahun	2021
Hasil	<p>Penelitian ini menggunakan metode pengelompokan K-Means untuk mengidentifikasi daerah-daerah dengan tingkat potensi gizi yang berbeda di Kota Prabumulih. Tiga kelompok diidentifikasi: C1 untuk daerah dengan potensi gizi tinggi, C2 untuk daerah dengan potensi gizi sedang, dan C3 untuk daerah dengan potensi kekurangan gizi rendah.</p> <p>Penelitian ini bertujuan untuk memberikan informasi untuk pengambilan keputusan dan memaksimalkan pemantauan kesehatan pada balita untuk mencegah potensi masalah kesehatan. Metode pengembangan sistem yang digunakan adalah eXtreme Programming (XP).</p> <p>Studi ini mengumpulkan data pada balita dengan potensi buruk dari 2018-2020 dan menghasilkan identifikasi cluster yang mewakili berbagai tingkat potensi nutrisi. Penerapan sistem pemetaan menggunakan metode</p>

	<p>pengelompokan K-mean berguna untuk mengklasifikasikan anak kurang gizi di bawah lima tahun, yang dapat membantu dalam mengatasi masalah kekurangan gizi di Kota Prabumulih.</p> <p>Studi ini menyoroti pentingnya nutrisi untuk pertumbuhan dan perkembangan balita dan kebutuhan untuk memetakan area dengan anak-anak kekurangan gizi untuk mengatasi masalah ini secara efektif. Studi ini menggunakan metode pengumpulan data primer dan sekunder, termasuk observasi, wawancara, dan jurnal penelitian, untuk mengumpulkan informasi yang relevan.</p> <p>Studi ini bertujuan untuk mengembangkan sistem informasi geografis yang berfokus pada penanganan kebutuhan nutrisi balita dengan potensi kekurangan gizi, menyediakan alat yang berharga untuk pemantauan dan pengambilan Keputusan. Studi ini menekankan pentingnya mengatasi kekurangan gizi sebagai perhatian bagi pemerintah daerah dan beban bagi negara bagian dan keluarga.</p>
Keterkaitan Penelitian	Keterkaitan pada penelitian ini yaitu mengidentifikasi daerah-daerah dengan tingkat potensi gizi yang berbeda.

No.	5.
Judul	Application of Clustering-Based Data Mining for the Assessment of Nutritional Status in Toddlers at Community Health Centers (Fianty et al., 2023)
Penulis	<ol style="list-style-type: none"> 1. Melissa Indah Fianty 2. Monika Evelin Johan 3. Azka Aulia 4. Mella Margareta Veronica
Tahun	2023

Hasil	<p>Penggunaan teknik penambangan data dengan algoritma pengelompokan k-mean memungkinkan pemahaman yang lebih dalam tentang status gizi balita.</p> <p>Hasilnya menyoroti pentingnya intervensi yang hati-hati dan tepat waktu untuk mengatasi masalah kekurangan gizi pada balita.</p> <p>Pemantauan kesehatan secara teratur oleh pusat-pusat kesehatan masyarakat dan keterlibatan orang tua yang aktif sangat penting untuk mendukung pertumbuhan dan perkembangan anak-anak ini secara optimal.</p> <p>Penelitian ini memberikan wawasan berharga untuk mengatasi masalah kekurangan gizi dan mendukung pembentukan generasi masa depan yang lebih sehat dan lebih menjanjikan di Indonesia.</p>
Keterkaitan Penelitian	Keterkaitan pada penelitian ini yaitu Penggunaan teknik penambangan data dengan algoritma pengelompokan k-mean memungkinkan pemahaman yang lebih dalam tentang status gizi balita.

2.2 Pembelajaran Mesin

Pembelajaran mesin adalah “Bidang studi yang memberikan komputer kemampuan untuk belajar tanpa diprogram secara eksplisit(Arthur Samuel, 1959). Pembelajaran mesin adalah studi tentang algoritme komputer yang memungkinkan program komputer ditingkatkan secara otomatis melalui pengalaman(Tom M. Mitchell (1997)). Machine learning dapat didefinisikan sebagai metode komputasi berdasarkan pengalaman untuk meningkatkan performa atau membuat prediksi yang akurat(Mohri et.al, 2012). Pembelajaran mesin adalah disiplin ilmu yang memfokuskan pada pengembangan teknik-teknik yang memungkinkan komputer untuk mengatasi tugas-tugas yang kompleks dengan bantuan data.(Christopher M. Bishop.,2006)

2.3 Clustering

Clustering adalah proses pengelompokan benda serupa ke dalam kelompok yang berbeda, atau lebih tepatnya partisi dari sebuah data set kedalam subset, sehingga data dalam setiap subset memiliki arti yang bermanfaat. Dimana sebuah cluster terdiri dari kumpulan benda-benda yang mirip antara satu dengan yang lainnya dan berbeda dengan benda yang terdapat pada cluster lainnya. Algoritma clustering terdiri dari dua bagian yaitu secara hirarkis dan secara partitional. Algoritma hirarkis menemukan cluster secara berurutan dimana

cluster ditetapkan sebelumnya, sedangkan algoritma partitional menentukan semua kelompok pada waktu tertentu (Madhulatha, 2012). Clustering juga bisa dikatakan suatu proses dimana mengelompokkan dan membagi pola data menjadi beberapa jumlah data set sehingga akan membentuk pola yang serupa dan dikelompokkan pada cluster yang sama dan memisahkan diri dengan membentuk pola yang berbeda ke cluster yang berbeda (HUNG et al., 2005).

Clustering dapat memainkan peran penting dalam kehidupan sehari-hari, karena tidak bisa lepas dengan sejumlah data yang menghasilkan informasi untuk memenuhi kebutuhan hidup. Salah satu sarana yang paling penting dalam hubungan dengan data adalah untuk mengklasifikasikan atau mengelompokkan data tersebut ke dalam seperangkat kategori atau cluster. Clustering dapat ditemukan di beberapa aplikasi yang ada di berbagai bidang. Sebagai contoh pengelompokan data yang digunakan untuk menganalisa data statistik seperti pengelompokan untuk pembelajaran mesin, data mining, pengenalan pola, analisis citra dan bioinformatika (Bataineh et al., 2011).

2.4 K – Means

Algoritma K-means merupakan salah satu algoritma dengan partitional, karena K-Means didasarkan pada penentuan jumlah awal kelompok dengan mendefinisikan nilai centroid awalnya (Madhulatha, 2012). Algoritma K - means menggunakan proses secara berulang-ulang 14 untuk mendapatkan basis data cluster. Dibutuhkan jumlah cluster awal yang diinginkan sebagai masukan dan menghasilkan titik centroid akhir sebagai output. Metode K-means akan memilih pola k sebagai titik awal centroid secara acak atau random. Jumlah iterasi untuk mencapai cluster centroid akan dipengaruhi oleh calon cluster centroid awal secara random. Sehingga didapat cara dalam pengembangan algoritma dengan menentukan centroid cluster yang dilihat dari kepadatan data awal yang tinggi agar mendapatkan kinerja yang lebih tinggi (HUNG et al., 2005, Saranya & Punithavalli, 2011, Eltibi & Ashour, 2011)

Dalam penyelesaiannya, algoritma K-Means akan menghasilkan titik centroid yang dijadikan tujuan dari algoritma K-Means. Setelah iterasi KMeans berhenti, setiap objek dalam dataset menjadi anggota dari suatu cluster. Nilai cluster ditentukan dengan mencari seluruh objek untuk menemukan cluster dengan jarak terdekat ke objek. Algoritma K -means akan mengelompokkan item data dalam suatu dataset ke suatu cluster berdasarkan jarak terdekat (Bangoria et al., 2013).

BAB III

HASIL DAN PEMBAHASAN

3.1 K – MEANS

3.1.1 Proses Tahapan Pengolahan Data dan Clustering hingga Evaluasi

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

Pembahasan :

- import numpy as np : untuk membuat struktur data array yang efisien.
- import pandas as pd : untuk mengelola dan menganalisis data tabular.
- import matplotlib.pyplot as plt : untuk membuat visualisasi seperti grafik dan plot.
- import seaborn as sns : untuk membuat visualisasi statistik yang menarik.
- %matplotlib inline : untuk Jupyter Notebook yang memastikan bahwa plot yang dihasilkan oleh Matplotlib akan ditampilkan langsung di dalam notebook, tanpa perlu menggunakan fungsi plt.show().

```
In [3]: import warnings
warnings.filterwarnings('ignore')
```

Pembahasan :

- import warnings : pengaturan yang terkait peringatan atau pemberitahuan yang dapat muncul selama eksekusi program.
- warnings.filterwarnings('ignore') : untuk menghilangkan tampilan yang mungkin tidak diinginkan atau tidak relevan, terutama pada saat pengembangan atau analisis data.

```
In [4]: df = pd.read_csv('gizi.csv')
df
```

Pembahasan :

- df = pd.read_csv('gizi.csv') : untuk membaca data dari file CSV ke dalam suatu DataFrame.
- df = pemanggilan untuk menampilkan data yang telah dibaca sebelumnya.

Berikut tampilan data nya :

Out[4]:

	NO	NAMA BALITA	JENIS KELAMIN	UMUR BULAN/TAHUN	Berat	Tinggi	BMI
0	1	Adhe Fitri	P	24	5.8	0.650	13.73
1	2	Andi Hariati	P	24	5.5	0.590	15.80
2	3	Anwar Amir	L	28	6.7	0.715	13.11
3	4	Asmar	L	30	8.1	0.725	15.41
4	5	Eka Andriyani	P	28	6.9	0.730	12.95
...
494	496	Moza Rafasya	L	23	11.0	78.300	43.11
495	497	Amora Nauli	P	13	8.8	71.400	28.30
496	498	Arsenio	L	11	8.1	74.500	23.60
497	499	Shaqueena Nur	P	37	10.7	65.400	47.30
498	500	Alenka Ikrima	L	25	9.9	66.800	18.70

499 rows × 7 columns

```
In [13]: df = df.drop(['NO'], axis = 1)
```

```
In [14]: df
```

Pembahasan :

- `df = df.drop(['NO'], axis = 1)` : sebuah function untuk menghapus kolom, yaitu kolom 'NO'.
- `df` : pemanggilan untuk menampilkan data.

Berikut tampilan data nya :

Out[14]:

	NAMA BALITA	JENIS KELAMIN	UMUR BULAN/TAHUN	Berat	Tinggi	BMI
0	Adhe Fitri	P	24	5.8	0.650	13.73
1	Andi Hariati	P	24	5.5	0.590	15.80
2	Anwar Amir	L	28	6.7	0.715	13.11
3	Asmar	L	30	8.1	0.725	15.41
4	Eka Andriyani	P	28	6.9	0.730	12.95
...
494	Moza Rafasya	L	23	11.0	78.300	43.11
495	Amora Nauli	P	13	8.8	71.400	28.30
496	Arsenio	L	11	8.1	74.500	23.60
497	Shaqueena Nur	P	37	10.7	65.400	47.30
498	Alenka Ikrima	L	25	9.9	66.800	18.70

499 rows × 6 columns


```
In [15]: df.head(10)
```

```
Out[15]:
```

	NAMA BALITA	JENIS KELAMIN	UMUR BULAN/TAHUN	Berat	Tinggi	BMI
0	Adhe Fitri	P	24	5.8	0.650	13.73
1	Andi Hariati	P	24	5.5	0.590	15.80
2	Anwar Amir	L	28	6.7	0.715	13.11
3	Asmar	L	30	8.1	0.725	15.41
4	Eka Andriyani	P	28	6.9	0.730	12.95
5	Evi Multazam	P	25	5.8	0.770	9.78
6	Firdaus Zubair	L	23	4.7	0.525	17.05
7	Ilham Wahyudi	L	28	6.8	0.540	23.32
8	Irmayaeni	P	21	3.5	0.550	11.57
9	Lieri Aprilyanti	P	27	5.8	0.530	20.65

Pembahasan :

- df.head(10) : menampilkan 10 data teratas

```
In [16]: df.tail(10)
```

```
Out[16]:
```

	NAMA BALITA	JENIS KELAMIN	UMUR BULAN/TAHUN	Berat	Tinggi	BMI
489	Isna Naila Putri	P	33	10.4	60.5	33.60
490	Rafasya Khiar	L	22	8.5	71.0	18.20
491	Adzka Naufal Abiyu	L	32	8.3	68.5	19.25
492	Muhammad	L	32	7.8	67.8	30.30
493	Arroyan Zavier	L	23	9.8	67.4	42.20
494	Moza Rafasya	L	23	11.0	78.3	43.11
495	Amora Nauli	P	13	8.8	71.4	28.30
496	Arsenio	L	11	8.1	74.5	23.60
497	Shaqueena Nur	P	37	10.7	65.4	47.30
498	Alenka Ikrima	L	25	9.9	66.8	18.70

Pembahasan :

- df.tail(10) : menampilkan 10 data terbawah

```
In [17]: df.shape
```

```
Out[17]: (499, 6)
```

```
In [18]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 499 entries, 0 to 498
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  -
0   NAMA BALITA           499 non-null    object
1   JENIS KELAMIN         499 non-null    object
2   UMUR BULAN/TAHUN      499 non-null    int64
3   Berat                 499 non-null    float64
4   Tinggi                499 non-null    float64
5   BMI                   499 non-null    float64
dtypes: float64(3), int64(1), object(2)
memory usage: 23.5+ KB
```

Pembahasan :

- df.shape : untuk menampilkan dimensi atau bentuk seperti kolom dan baris.
- df.info : untuk melihat jumlah kolom, non-null count, dan type data.

```
In [19]: df.isnull().sum()
```

```
Out[19]: NAMA BALITA           0
          JENIS KELAMIN       0
          UMUR BULAN/TAHUN    0
          Berat               0
          Tinggi              0
          BMI                 0
          dtype: int64
```

Pembahasan :

- df.isnull().sum() : untuk mengidentifikasi nilai – nilai yang hilang (Null) dalam data, kemudian menghitung jumlah nilai null untuk setiap kolom.

```
In [20]: df.describe()
```

```
Out[20]:
```

	UMUR BULAN/TAHUN	Berat	Tinggi	BMI
count	499.000000	499.000000	499.000000	499.000000
mean	22.242485	9.092385	42.978918	30.229038
std	11.033689	2.010890	38.814302	207.292107
min	4.000000	1.000000	0.490000	1.730000
25%	13.000000	7.900000	0.760000	13.845000
50%	21.000000	9.000000	66.800000	17.750000
75%	32.000000	10.000000	77.900000	26.320000
max	54.000000	15.700000	104.500000	4645.000000

Pembahasan :

- df.describe() : untuk menghasilkan ringkasan statistic untuk setiap kolom numerik pada data. Ringkasan ini berupa count, mean, std, min, kuartil pertama, kuartil kedua, kuartil ketiga dan max.

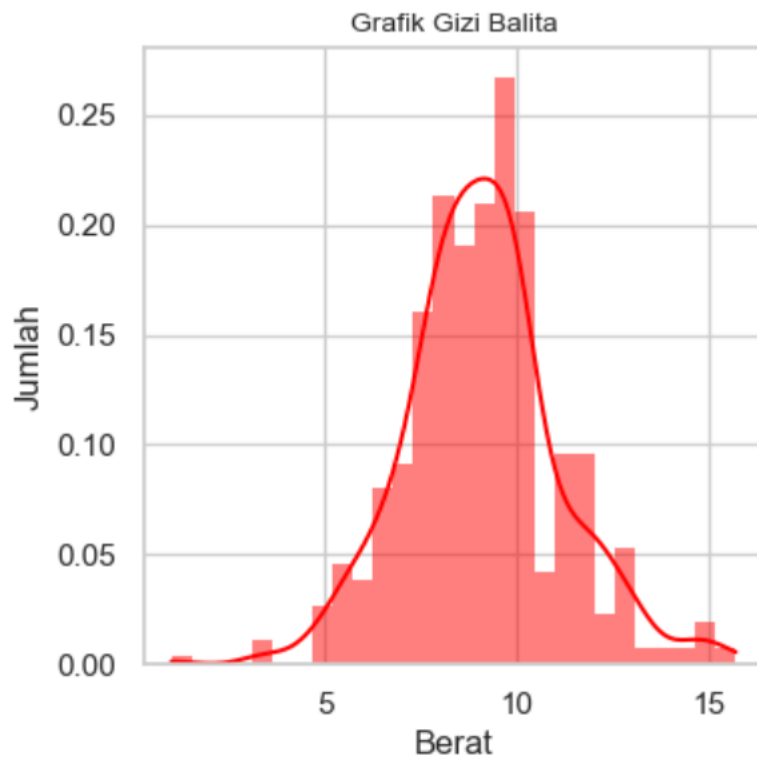
```
In [26]: plt.figure(figsize=(4,4))
sns.set(style= 'whitegrid')
sns.histplot(df['Berat'], color='red', kde=True,
              stat='density', linewidth=0)
plt.title('Grafik Gizi Balita',fontsize=10)
plt.ylabel('Jumlah')
plt.xlabel('Berat')
```

Pembahasan :

- untuk membuat histogram (grafik frekuensi) dari kolom 'Berat' pada data dengan library Matplotlib dan Seaborn.

Berikut hasil yang terlihat :

```
Out[26]: Text(0.5, 0, 'Berat')
```



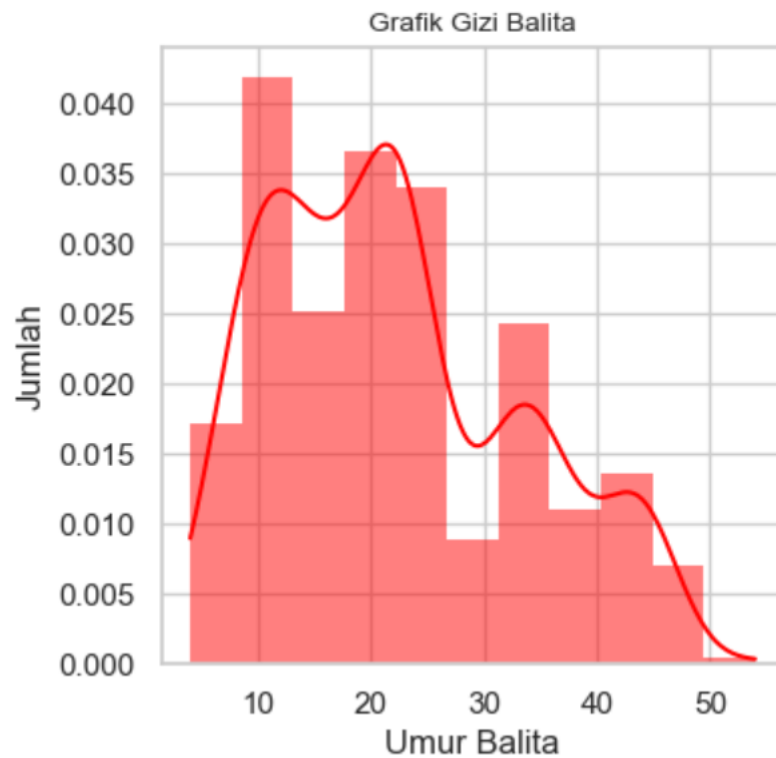
```
In [27]: plt.figure(figsize=(4,4))
sns.set(style= 'whitegrid')
sns.histplot(df['UMUR BULAN/TAHUN'], color='red',
             kde=True, stat='density', linewidth=0)
plt.title('Grafik Gizi Balita',fontsize=10)
plt.ylabel('Jumlah')
plt.xlabel('Umur Balita')
```

Pembahasan :

- untuk membuat histogram (grafik frekuensi) dari kolom 'Umur Bulan / Tahun' pada data dengan library Matplotlib dan Seaborn.

Berikut hasil yang terlihat :

```
Out[27]: Text(0.5, 0, 'Umur Balita')
```

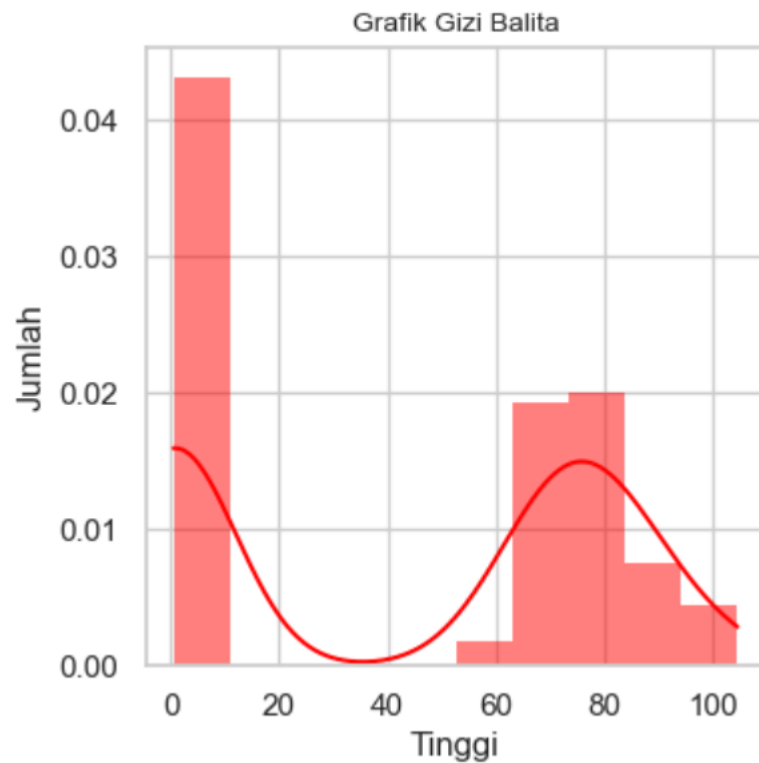


```
In [28]: plt.figure(figsize=(4,4))
sns.set(style= 'whitegrid')
sns.histplot(df['Tinggi'], color='red',
             kde=True, stat='density', linewidth=0)
plt.title('Grafik Gizi Balita',fontsize=10)
plt.ylabel('Jumlah')
plt.xlabel('Tinggi')
```

- untuk membuat histogram (grafik frekuensi) dari kolom 'Tinggi' pada data dengan library Matplotlib dan Seaborn.

Berikut hasil yang terlihat :

```
Out[28]: Text(0.5, 0, 'Tinggi')
```



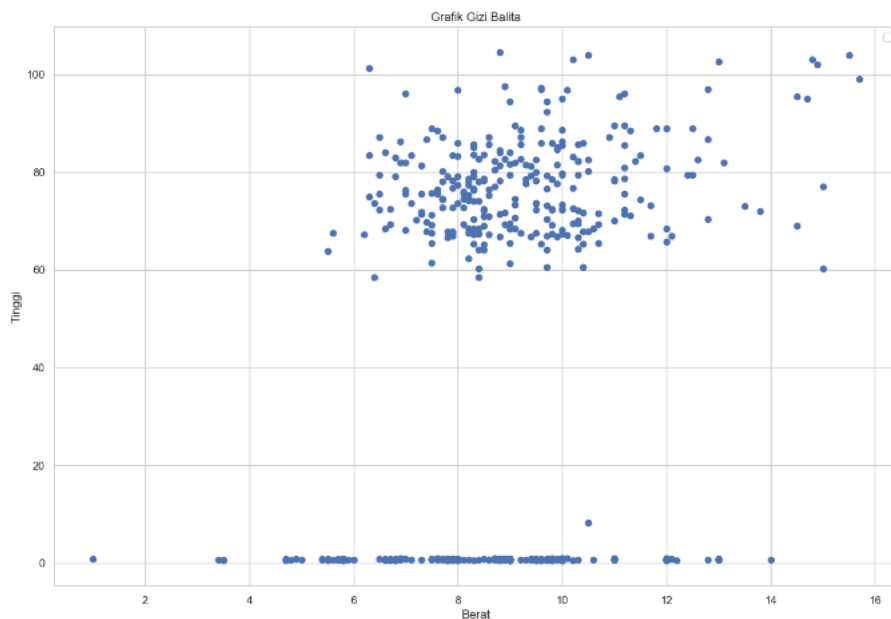
```
In [29]: plt.figure(figsize=(15, 10))
plt.scatter(df['Berat'], df['Tinggi'])
plt.xlabel("Berat")
plt.ylabel("Tinggi")
plt.title("Grafik Gizi Balita")
plt.legend()
plt.show()
```

No artists with labels found to put in legend. Note that artists whose label start with an underscore are ignored when legend() is called with no argument.

Pembahasan :

- untuk membuat scatter plot (grafik sebaran) dari dua kolom dalam DataFrame df, yaitu 'Berat' dan 'Tinggi', menggunakan library Matplotlib.

Berikut tampilan datanya :



```
In [30]: print(df.columns)
```

```
Index(['NAMA BALITA', 'JENIS KELAMIN', 'UMUR BULAN/TAHUN',  
      'Berat', 'Tinggi',  
      'BMI'],  
      dtype='object')
```

```
In [31]: df.columns = df.columns.str.strip()
```

```
In [32]: x = df[['Berat', 'Tinggi']].values
```

```
In [33]: from sklearn.preprocessing import MinMaxScaler  
scaler = MinMaxScaler()  
x = scaler.fit_transform(x)
```

Pembahasan :

- `print(df.columns)` : untuk mencetak nama kolom dari data
- `df.columns = df.columns.str.strip()` : untuk menghapus spasi yang mungkin ada di awal atau akhir nama kolom
- `x = df[['Berat', 'Tinggi']].values` : untuk membuat variable 'x' yang berisi array Numpy yang berisi data dari kolom 'Berat' dan 'Tinggi'
- `from sklearn.preprocessing import MinMaxScaler` : untuk melakukan penskalaan fitur, sehingga nilai-nilai dalam fitur akan diubah menjadi rentang tertentu.
- `scaler = MinMaxScaler()` : untuk mentransformasi data dengan menggunakan metode penskalaan min-max.

- `x = scaler.fit_transform(x)` : untuk menghitung nilai minimum dan maksimum dari setiap fitur variable 'x'

In [34]: x

```
Out[34]: array([[3.26530612e-01, 1.53831362e-03],
                [3.06122449e-01, 9.61446015e-04],
                [3.87755102e-01, 2.16325353e-03],
                [4.82993197e-01, 2.25939813e-03],
                [4.01360544e-01, 2.30747044e-03],
                [3.26530612e-01, 2.69204884e-03],
                [2.51700680e-01, 3.36506105e-04],
                [3.94557823e-01, 4.80723007e-04],
                [1.70068027e-01, 5.76867609e-04],
                [3.26530612e-01, 3.84578406e-04],
                [4.14965986e-01, 1.24987982e-03],
                [5.71428571e-01, 1.15373522e-03],
                [4.89795918e-01, 1.53831362e-03],
                [2.72108844e-01, 1.82674743e-03],
                [3.40136054e-01, 1.73060283e-03],
                [6.25850340e-01, 9.61446015e-05],
                [8.02721088e-01, 2.21132583e-03],
                [6.19047619e-01, 4.42265167e-03],
                [5.71428571e-01, 2.49975964e-03],
```

Pembahasan :

- x : untuk mencetak nilai x dengan array

In [35]: `from sklearn.cluster import KMeans`

```
kmeans = KMeans(n_clusters=4, random_state=0)
kmeans.fit(x)
```

Out[35]: `KMeans(n_clusters=4, random_state=0)`

In [36]: `y_cluster = kmeans.fit_predict(x)`

In [37]: `y_cluster`

Pembahasan :

- In [35] : digunakan untuk melakukan pengelompokkan data dalam variable 'x' ke dalam cluster-cluster tertentu.
- In [36] : untuk mengembalikan label cluster untuk setiap observasi dalam data x.
- In [37] : mencetak atau mengembalikan label cluster yang telah ditetapkan untuk setiap data point dalam variabel x.

Berikut tampilan data dari y_cluster :

```
Out[37]: array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 2, 2,
2, 2, 2, 2, 2,
0, 0, 0, 2, 2, 2, 2, 2, 2, 2, 2, 2, 0, 2, 2, 2, 2,
2, 2, 0, 0, 0,
0, 0, 0, 2, 0, 0, 0, 0, 0, 2, 2, 2, 2, 2, 2, 2, 2,
0, 0, 0, 2, 2,
2, 0, 2, 0, 2, 0, 0, 0, 2, 2, 2, 2, 2, 2, 2, 2,
2, 2, 2, 2, 2,
2, 0, 0, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 0, 0, 2, 2,
2, 2, 2, 2, 2,
0, 0, 2, 2, 2, 0, 2, 0, 0, 0, 2, 0, 0, 2, 2, 2, 2,
2, 0, 0, 0, 0,
2, 2, 2, 0, 2, 0, 0, 2, 2, 2, 2, 0, 0, 0, 0, 2, 2,
2, 2, 2, 0, 0,
2, 2, 2, 2, 2, 2, 2, 2, 2, 0, 2, 2, 2, 2, 0, 2, 2,
2, 2, 2, 2, 0,
0, 2, 0, 0, 0, 2, 2, 2, 0, 2, 2, 2, 0, 0, 0, 0, 0,
0, 2, 0, 2, 2,
2, 2, 1, 1, 1, 1, 3, 3, 3, 1, 1, 1, 3, 3, 3, 1, 3,
3, 3, 3, 1, 2,
3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 1, 1, 1, 3, 1, 1, 1,
1, 3, 1, 3, 1,
1, 1, 1, 3, 3, 1, 1, 3, 3, 1, 1, 1, 1, 3, 1, 1, 3,
1, 3, 3, 3, 3,
```

```
In [38]: df['Cluster'] = y_cluster
```

Pembahasan :

- Dengan menggabungkan kedua operasi tersebut, kita dapat melihat hasil clustering, yaitu label cluster untuk setiap data point dalam DataFrame df dan koordinat pusat dari setiap cluster.

```
In [39]: df.head(10)
```

```
Out[39]:
```

	NAMA BALITA	JENIS KELAMIN	UMUR BULAN/TAHUN	Berat	Tinggi	BMI	Cluster
0	Adhe Fitri	P	24	5.8	0.650	13.73	0
1	Andi Hariati	P	24	5.5	0.590	15.80	0
2	Anwar Amir	L	28	6.7	0.715	13.11	0
3	Asmar	L	30	8.1	0.725	15.41	0
4	Eka Andriyani	P	28	6.9	0.730	12.95	0
5	Evi Multazam	P	25	5.8	0.770	9.78	0
6	Firdaus Zubair	L	23	4.7	0.525	17.05	0
7	Ilham Wahyudi	L	28	6.8	0.540	23.32	0
8	Irmayaeni	P	21	3.5	0.550	11.57	0
9	Lieri Aprilyanti	P	27	5.8	0.530	20.65	0

Pembahasan :

- Menampilkan 10 data yang sudah melakukan clustering

```
In [40]: kmeans.cluster_centers_
```

```
Out[40]: array([[0.38455864, 0.00211518],  
                [0.50557543, 0.71141752],  
                [0.61957833, 0.00288161],  
                [0.70790291, 0.80553864]])
```

Pembahasan :

- Untuk mengakses atribut yang ada

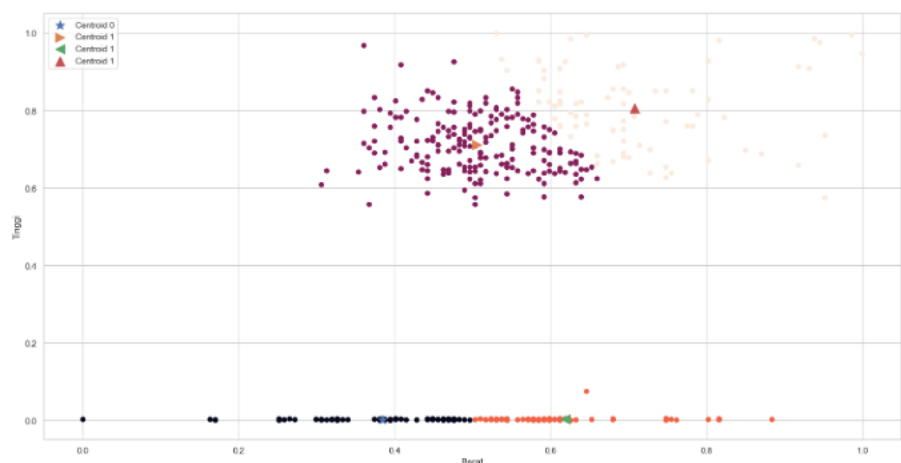
```
In [41]: plt.figure(figsize=(20, 10))
plt.scatter(x[:,0],x[:,1],c=kmeans.labels_)
plt.scatter(kmeans.cluster_centers_[0,0],
            kmeans.cluster_centers_[0,1],
            cmap='rainbow',marker='*',
            s=150,label='Centroid 0')
plt.scatter(kmeans.cluster_centers_[1,0],
            kmeans.cluster_centers_[1,1],
            cmap='rainbow',marker='>',
            s=150,label='Centroid 1')
plt.scatter(kmeans.cluster_centers_[2,0],
            kmeans.cluster_centers_[2,1],
            cmap='rainbow',marker='<',
            s=150,label='Centroid 1')
plt.scatter(kmeans.cluster_centers_[3,0],
            kmeans.cluster_centers_[3,1],
            cmap='rainbow',marker='^',
            s=150,label='Centroid 1')
plt.xlabel("Berat")
plt.ylabel("Tinggi")
plt.legend()
```

Pembahasan :

- scatter plot yang menunjukkan data dalam cluster-cluster yang berbeda dengan centroid masing-masing cluster. Plot ini membantu dalam visualisasi pemahaman tentang bagaimana data telah terkelompok berdasarkan hasil clustering yang telah dilakukan.

Berikut tampilan data :

Out[41]: <matplotlib.legend.Legend at 0x1ecfa164fa0>



```
In [46]: from sklearn.metrics import silhouette_score

silhouette_avg = silhouette_score(x, kmeans.labels_)
print("Silhouette Coefficient:", silhouette_avg)

Silhouette Coefficient: 0.4789508559842194
```

```
In [47]: from sklearn.cluster import KMeans
cs = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters = i)
    kmeans.fit(x)
    cs.append(kmeans.inertia_)
cs
```

```
Out[47]: [78.67163435484544,
          11.57046856493527,
          8.684694770783896,
          5.83975044130192,
          4.825355663808294,
          3.8256678264660366,
          2.977786642919814,
          2.594392155844976,
          2.2128109548366486,
          1.9063242898276562]
```

Pembahasan :

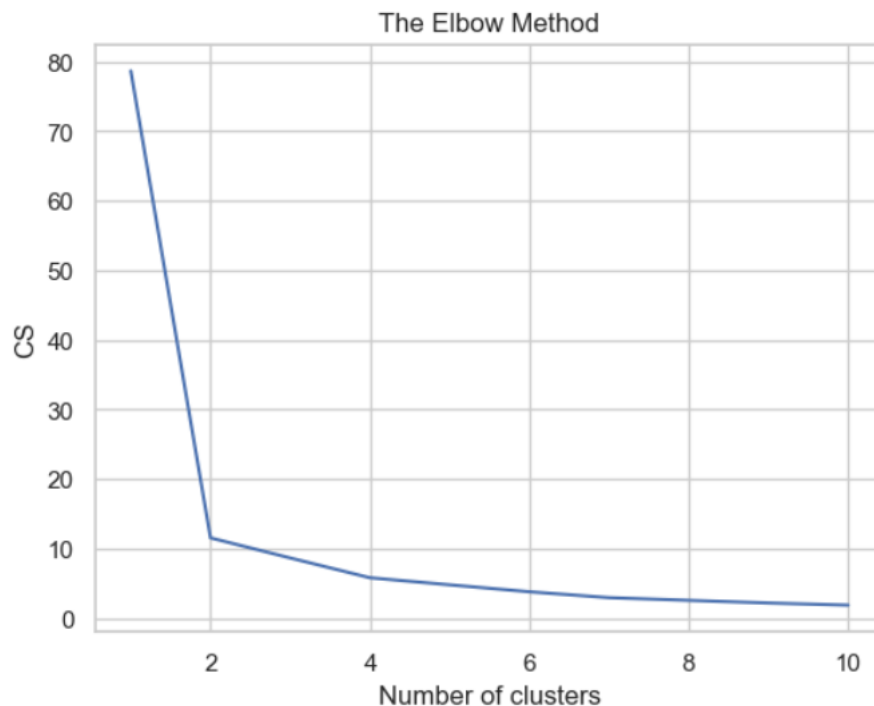
- In [46] : untuk menghitung dan mencetak nilai Silhouette Coefficient untuk hasil clustering yang telah dilakukan menggunakan algoritma K-Means.
- In [47] : untuk menghitung nilai inersia (inertia) untuk berbagai jumlah cluster (K) menggunakan algoritma K-Means.

```
In [48]: plt.plot(range(1, 11), cs)
plt.title('The Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('CS')
plt.show()
```

Pembahasan :

- untuk membuat grafik yang disebut "Elbow Method" atau metode siku. Grafik ini membantu dalam menentukan jumlah optimal dari cluster dalam analisis clustering.

Berikut tampilan grafik elbow :



```
In [50]: df['Cluster'] = df['Cluster'].replace([0,1,2,3],
                                                ['buruk', 'kurang',
                                                'normal', 'lebih'])
```

Pembahasan :

- pada baris ini menggunakan metode `replace` pada kolom 'Cluster' dalam DataFrame `df` untuk mengganti nilai-nilai label cluster yang semula berupa angka (0, 1, 2, 3) dengan label yang lebih deskriptif ('buruk', 'kurang', 'normal', 'lebih'). Penggantian ini membantu dalam memahami makna masing-masing cluster.

```
In [51]: print(df.head(10))
```

	NAMA BALITA	JENIS KELAMIN	UMUR BULAN/TAHUN	Berat	Tinggi	BMI \
0	Adhe Fitri	P	24	5.8	0.650	13.73
1	Andi Hariati	P	24	5.5	0.590	15.80
2	Anwar Amir	L	28	6.7	0.715	13.11
3	Asmar	L	30	8.1	0.725	15.41
4	Eka Andriyani	P	28	6.9	0.730	12.95
5	Evi Multazam	P	25	5.8	0.770	9.78
6	Firdaus Zubair	L	23	4.7	0.525	17.05
7	Ilham Wahyudi	L	28	6.8	0.540	23.32
8	Irmayaeni	P	21	3.5	0.550	11.57
9	Lieri Aprilyanti	P	27	5.8	0.530	20.65

	Cluster
0	buruk
1	buruk
2	buruk
3	buruk
4	buruk
5	buruk
6	buruk
7	buruk
8	buruk
9	buruk

Pembahasan :

- mencetak 10 data teratas dan cluster pada data 10 teratas dinyatakan buruk.

```
In [52]: print(df.tail(10))
```

	NAMA BALITA	JENIS KELAMIN	UMUR BULAN/TAHUN	Berat	Tinggi	BMI \
489	Isna Naila Putri	P	33	10.4	60.5	33.60
490	Rafasya Khiair	L	22	8.5	71.0	18.20
491	Adzka Naufal Abiyu	L	32	8.3	68.5	19.25
492	Muhammad	L	32	7.8	67.8	30.30
493	Arroyan Zavier	L	23	9.8	67.4	42.20
494	Moza Rafasya	L	23	11.0	78.3	43.11
495	Amora Nauli	P	13	8.8	71.4	28.30
496	Arsenio	L	11	8.1	74.5	23.60
497	Shaqueena Nur	P	37	10.7	65.4	47.30
498	Alenka Ikrima	L	25	9.9	66.8	18.70

	Cluster
489	kurang
490	kurang
491	kurang
492	kurang
493	kurang
494	lebih
495	kurang
496	kurang
497	kurang
498	kurang

Pembahasan :

- mencetak 10 data terbawah dan cluster pada data 10 terbawah dinyatakan kurang dan lebih.

```
In [53]: print(df.iloc[60:75])
```

	NAMA	BALITA	JENIS	KELAMIN	UMUR	BULAN/TAHUN	Berat	Tinggi	BMI	Cluster
60	dandy			L		32	10.0	0.95	11.08	normal
61	fitria			P		36	5.4	0.75	9.60	buruk
62	faki			L		45	5.5	0.77	9.28	buruk
63	ekel			L		38	7.8	0.73	14.64	buruk
64	dans			L		43	8.9	0.72	17.17	normal
65	maryam			P		43	9.8	0.67	21.83	normal
66	agus			L		44	8.8	0.75	15.64	normal
67	putri			P		23	6.9	0.94	7.81	buruk
68	pia			P		45	9.0	0.88	11.62	normal
69	nawie			L		43	8.0	0.76	13.85	buruk
70	anji			L		23	8.7	0.91	10.51	normal
71	afgan			L		19	7.8	0.69	16.38	buruk
72	rosa			P		29	6.6	0.54	22.63	buruk
73	dani			L		40	7.5	0.68	16.22	buruk
74	basmalla			P		34	8.9	0.77	15.01	normal

```
In [54]: from sklearn.metrics import silhouette_score

silhouette_avg = silhouette_score(x, kmeans.labels_)
print("Silhouette Coefficient:", silhouette_avg)
```

Silhouette Coefficient: 0.4690221099567972

Pembahasan :

- In [53] : untuk mencetak baris data dalam rentang indeks dari 60 hingga 74 (indeks 75 tidak termasuk) dari DataFrame df.
- In [54] : untuk mencetak baris data dalam rentang indeks dari 60 hingga 74 (indeks 75 tidak termasuk) dari DataFrame df.

BAB IV

PENUTUP

4.1 Kesimpulan

Penelitian ini bertujuan untuk melakukan meta-analisis data gizi berdasarkan tinggi dan berat badan dengan menggunakan model K-Means. Pendekatan ini akan menciptakan peta pola gizi kelompok populasi tertentu yang lebih akurat. Dengan menggunakan algoritma K-Means, penelitian ini mencoba mengelompokkan individu ke dalam kelompok dengan karakteristik gizi serupa. Hasil pengelompokan ini dapat memberikan wawasan lebih lanjut mengenai perbedaan pola gizi dalam suatu populasi, membantu mengidentifikasi kelompok yang mungkin memerlukan perhatian atau intervensi gizi khusus. Temuan ini menyoroti pentingnya analisis pengelompokan dalam mendukung pemahaman pola gizi dan memberikan dasar untuk pengembangan strategi gizi yang lebih tepat sasaran.

4.2 Saran

Untuk penyempurnaan pembuatan laporan ini, kami mengharapkan adanya saran dari semua pihak baik dosen, seluruh mahasiswa, serta dokter spesialis gizi, yang membaca laporan hasil penelitian ini.

DAFTAR PUSTAKA

- Fianty, M. I., Johan, M. E., Aulia, A., & Veronica, M. (2023). *Application of Clustering-Based Data Mining for the Assessment of Nutritional Status in Toddlers at Community Health Centers*. 5(4), 1350–1362. <https://doi.org/10.51519/journalisi.v5i4.586>
- Mutammimul Ula, Ananda Faridhatul Ulva, Ilham Saputra, Mauliza Mauliza, & Ivan Maulana. (2022). Implementation of Machine Learning Using the K-Nearest Neighbor Classification Model in Diagnosing Malnutrition in Children. *Multica Science and Technology (Mst) Journal*, 2(1), 94–99. <https://doi.org/10.47002/mst.v2i1.326>
- Nagari, S. S., & Inayati, L. (2020). Implementation of Clustering Using K-Means Method To Determine Nutritional Status. *Jurnal Biometrika Dan Kependudukan*, 9(1), 62. <https://doi.org/10.20473/jbk.v9i1.2020.62-68>
- Novianti, L., Robinson, R., Meiriska, I., & Sari, R. A. (2022). Geographic Information System Mapping and Management of Child with the Highest Nutritional Potential in Prabumulih City Using K-Means Clustering Method. *Proceedings of the 5th FIRST T1 T2 2021 International Conference (FIRST-T1-T2 2021)*, 9, 369–374. <https://doi.org/10.2991/ahe.k.220205.065>
- Ozzi ardhianto, Muhammad Salam Asyidqi, Ajif Yunizar Pratama Yusuf, S.Si, M.Eng, & Dr. Tb. Ai Munandar, S.Kom., MT. (2023). Clustering of Child Nutrition Status using Hierarchical Agglomerative Clustering Algorithm in Bekasi City. *International Journal of Information Technology and Computer Science Applications*, 1(3), 122–128. <https://doi.org/10.58776/ijitcsa.v1i3.42>

LAMPIRAN – LAMPIRAN

- **Lampiran 1**

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

import warnings
warnings.filterwarnings('ignore')

df = pd.read_csv('gizi.csv')
df

df = df.drop(['NO'], axis = 1)
df

df.head(10)
df.tail(10)
df.shape
df.info()
df.isnull().sum()
df.describe()

plt.figure(figsize=(4,4))
sns.set(style= 'whitegrid')
sns.histplot(df['Berat'], color='red', kde=True,
              stat='density', linewidth=0)
plt.title('Grafik Gizi Balita',fontsize=10)
plt.ylabel('Jumlah')
plt.xlabel('Berat')

plt.figure(figsize=(4,4))
sns.set(style= 'whitegrid')
sns.histplot(df['UMUR BULAN/TAHUN'], color='red',
              kde=True, stat='density', linewidth=0)
plt.title('Grafik Gizi Balita',fontsize=10)
plt.ylabel('Jumlah')
plt.xlabel('Umur Balita')

plt.figure(figsize=(4,4))
sns.set(style= 'whitegrid')
sns.histplot(df['Tinggi'], color='red',
              kde=True, stat='density', linewidth=0)
plt.title('Grafik Gizi Balita',fontsize=10)
plt.ylabel('Jumlah')
```

```

plt.xlabel('Tinggi')

plt.figure(figsize=(15, 10))
plt.scatter(df['Berat'], df['Tinggi'])
plt.xlabel("Berat")
plt.ylabel("Tinggi")
plt.title("Grafik Gizi Balita")
plt.legend()
plt.show()

print(df.columns)

df.columns = df.columns.str.strip()

x = df[['Berat', 'Tinggi']].values

from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
x = scaler.fit_transform(x)
x

from sklearn.cluster import KMeans
kmeans = KMeans(n_clusters=4, random_state=0)
kmeans.fit(x)

y_cluster = kmeans.fit_predict(x)
y_cluster
df['Cluster'] = y_cluster
df.head(10)
kmeans.cluster_centers_

plt.figure(figsize=(20, 10))
plt.scatter(x[:,0],x[:,1],c=kmeans.labels_)
plt.scatter(kmeans.cluster_centers_[0,0],
            kmeans.cluster_centers_[0,1],
            cmap='rainbow',marker='*',
            s=150,label='Centroid 0')
plt.scatter(kmeans.cluster_centers_[1,0],
            kmeans.cluster_centers_[1,1],
            cmap='rainbow',marker='>',
            s=150,label='Centroid 1')
plt.scatter(kmeans.cluster_centers_[2,0],
            kmeans.cluster_centers_[2,1],
            cmap='rainbow',marker='<',
            s=150,label='Centroid 1')
plt.scatter(kmeans.cluster_centers_[3,0],

```

```

        kmeans.cluster_centers_[3,1],
        cmap='rainbow',marker='^',
        s=150,label='Centroid 1')
plt.xlabel("Berat")
plt.ylabel("Tinggi")
plt.legend()

from sklearn.metrics import silhouette_score
silhouette_avg = silhouette_score(x, kmeans.labels_)
print("Silhouette Coefficient:", silhouette_avg)

from sklearn.cluster import KMeans
cs = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters = i)
    kmeans.fit(x)
    cs.append(kmeans.inertia_)
cs

plt.plot(range(1, 11), cs)
plt.title('The Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('CS')
plt.show()

kmeans.inertia_

df['Cluster'] = df['Cluster'].replace([0,1,2,3],
                                      ['buruk','kurang',
                                      'normal','lebih'])

```

- **Lampiran 2**

Link github : <https://github.com/amaliadamayant/K-Means-UAS-PM-C-202131002-AmaliaDH>

- **Lampiran 3**

Link dataset dari Kaggle :

<https://www.kaggle.com/code/ichaanawai/cluster>