

Implementasi Metode *K-Means Clustering* untuk Menentukan Kondisi Gizi Balita (Studi Kasus : Puskesmas Mamsena)

Yulita Eko¹, Yasinta Oktaviana Legu Rema², Hevi Herlina Ullu³, Budiman Baso⁴

^{1,2,3,4}Fakultas Pertanian Sains dan Kesehatan, Program Studi Teknologi Informasi, Universitas Negeri Timor, Timor Tengah Utara, Indonesia

Email: ¹yulitaeko2002@gmail.com, ²rema.ivana@gmail.com, ³heviherlina@unimor.ac.id,

⁴budimanbaso@gmail.com

Abstrak—Kondisi gizi balita sangat memengaruhi pertumbuhan dan kesehatan pada balita. Penentuan kondisi gizi balita perlu dilakukan untuk mengetahui presentase kondisi gizi dari balita tersebut. Penelitian ini dilakukan di Puskesmas Mamsena dengan tujuan untuk mengelompokkan kondisi gizi balita dengan sampel penelitian 348 data balita dengan menggunakan data antropometri balita. Antropometri ialah data yang merujuk pada pengukuran masa tubuh atau fisik dari balita. Jenis kelamin, usia, berat badan, tinggi badan, dan lingkaran lengan merupakan data antropometri yang digunakan dalam penelitian ini pada 348 balita. Penelitian ini mengimplementasikan salah satu metode *clustering* khususnya strategi pengelompokan *K-Means*, untuk mengelompokkan informasi yang sama menjadi sekumpulan informasi ke dalam kelompok-kelompok tertentu. Metode ini diimplementasikan dengan menggunakan bahasa pemrograman python untuk menganalisis data. Jumlah kelompok atau K terbaik diperoleh dengan menggunakan *elbow method* atau teknik siku dengan jumlah K = 5 yaitu gizi buruk, gizi kurang, gizi cukup, gizi baik dan kegemukan. Hasil pengelompokan yang diperoleh untuk setiap *cluster* yaitu *cluster* nol memiliki presentase = 28, 16 % atau sama dengan 98 data balita, *cluster* satu memiliki presentase = 16, 95 % atau sama dengan 59 data balita, *cluster* dua memiliki presentase = 23, 85 % atau sama dengan 83 data balita, *cluster* tiga memiliki presentase = 16, 09% atau sama dengan 56 data balita dan *cluster* empat memiliki presentase = 14, 94% atau sama dengan 52 data balita. Evaluasi hasil *clustering* dilakukan menggunakan metode DBI dengan nilai = 0,2520626580022645. Dari perolehan nilai DBI maka dapat disimpulkan bahwa kualitas hasil *cluster* dengan jumlah K=5 memiliki kualitas yang bagus karena hasil akurasi yang diperoleh mendekati 0.

Kata Kunci: *K-means clustering*, *Data mining*, Kondisi Gizi balita

Abstract— *Toddler nutritional conditions greatly affect growth and health in toddlers. Determination of toddler nutritional conditions needs to be done to determine the nutritional conditions of the toddlers. This study was conducted at the Mamsena Health Center with the aim of grouping toddler nutritional conditions with a research sample of 348 toddler data using toddler anthropometric data. Anthropometry is data that refers to the measurement of body mass or physical of toddlers. Gender, age, weight, height, and arm circumference are anthropometric data used in this study on 348 toddlers. This study implements one of the clustering methods, especially the K-Means grouping strategy, to group the same information into a group of information into certain groups. This method is implemented using the python programming language to analyze data. The best number of groups or K is obtained by using the elbow method or elbow technique with the number of K = 5, namely poor, malnutrition, adequate nutrition, good nutrition and obesity. The grouping results obtained for each cluster are cluster zero has a percentage = 28.16% or equal to 98 toddler data, cluster one has a percentage = 16.95% or equal to 59 toddler data, cluster two has a percentage = 23.85% or equal to 83 toddler data, cluster three has a percentage = 16.09% or equal to 56 toddler data and cluster four has a percentage = 14.94% or equal to 52 toddler data. Evaluation of clustering results is carried out using the DBI method with a value = 0.2520626580022645. From the DBI value obtained, it can be concluded that the quality of the cluster results with the number K = 5 has good quality because the accuracy results obtained are close to 0.*

Keywords: : *K-means clustering*, *Data mining*, *Nutritional Conditions of Toddlers*

1. PENDAHULUAN

Salah satu faktor utama yang menentukan kesehatan pada balita yaitu peningkatan status gizi bayi Menurut [1], kondisi kesehatan seseorang sebagai akibat dari mengonsumsi makanan dan memanfaatkan zat gizi disebut sebagai "status gizi". Pola makan yang seimbang diperlukan untuk membantu balita tumbuh dan berkembang dengan baik. Dalam situasi ini, status gizi dapat dibagi menjadi beberapa kelas, seperti, kurang gizi, gizi cukup, gizi baik dan kegemukan yang diperkirakan menggunakan data *antropometri* atau perkiraan aktual balita yang mencakup hal-hal seperti tinggi, berat, lingkaran lengan atas (LILA) dan jenis kelamin. Kecerdasan, daya tahan tubuh, kematian bayi dan kematian ibu dapat dipengaruhi oleh status gizi bayi. Selain itu, status gizi balita dipengaruhi oleh status keuangan, perilaku pola makan, dan pendapatan keluarga. Pemeriksaan dan penilaian status gizi sangat penting untuk mengenali masalah dan mendorong solusi yang efektif. Usia, tinggi, berat, dan

lingkar lengan atas adalah beberapa karakteristik yang digunakan untuk menentukan status gizi balita. Estimasi ini membutuhkan banyak aset, terutama jika dilakukan pada data anak kecil yang tak terhitung jumlahnya. Dalam upaya atau proyek untuk lebih meningkatkan perawatan bayi dengan mengembangkan metodologi yang lebih efektif dan efisien.

Teknologi informasi merupakan suatu keharusan, khususnya untuk mengorganisasikan dan menyimpan data. Salah satu dari sekian banyak teknik pengelolaan dan analisis data di bidang kesehatan adalah *data mining*. Pengorganisasian data gizi balita menjadi tantangan tersendiri bagi Puskesmas Mamsena. Saat ini pengelompokan data masih menggunakan *Ms. Excel*. Jumlah data yang sangat banyak dapat menyebabkan aplikasi *Excel* berjalan lambat. Di Puskesmas Mamsena belum ada satu teknik khusus yang dapat digunakan untuk mengelompokkan masalah gizi anak dalam jumlah besar secara mekanis dan metodis. Oleh karena itu, salah satu strategi yang dapat digunakan untuk mengatasi masalah tersebut adalah penerapan teknik *data mining* dengan metode *K-Means Clustering*. Untuk membantu pengambilan keputusan di masa mendatang, *data mining* sebagaimana yang didefinisikan oleh [2] adalah proses mengidentifikasi pola dan informasi terkait dalam basis data yang sangat besar. Teknik ini dapat digunakan untuk mengorganisasikan data dan mengidentifikasi pola-pola penting yang mungkin tidak langsung terlihat. Metode-metode yang dapat digunakan dalam teknik *data mining* antara lain klasifikasi, asosiasi, *clustering*, *sequencing*, deskriptif, regresi, dan peramalan/prediksi.

Algoritma *K-Means* merupakan salah satu algoritma *clustering* yang sangat populer dalam mengelompokkan data sesuai dengan karakteristik kenyamanan. Pengelompokkan data yang diperoleh disebut sebagai *cluster*/klaster [14]. Algoritma *K-means Clustering* merupakan suatu metode penganalisaan data atau metode *data mining* yang mengelompokkan data sebagai partisi. *K-Means Clustering* memiliki objek yaitu meminimalisasi *object function* yang telah di atur pada proses klasterisasi dengan cara minimalisasi variasi antar *cluster* 1 dengan maksimalisasi variasi dengan data *cluster* lainnya [13].

Tujuan dari penelitian ini adalah untuk mengkategorikan data anak menurut status gizi masing-masing [3]. Pengelompokan hierarkis dan non-hierarkis merupakan dua metode dasar yang digunakan dalam teknik pengelompokan. Data dapat dikelompokkan ke dalam kelompok yang mirip atau berbeda satu sama lain menggunakan algoritma pengelompokan. Pengelompokan *K-means* merupakan salah satu teknik non-hierarkis yang digunakan dalam penelitian ini. Pendekatan ini mengumpulkan data ke dalam kelompok-kelompok dengan kualitas yang sama dalam kelompok yang sama. Metode siku digunakan untuk menghasilkan angka acak (*K*) untuk jumlah klaster dalam algoritma *K-Means*. Titik data terdekat dengan centroid atau pusat klaster ditemukan menggunakan rumus Jarak Euclidean [4]. *K-Means* bertujuan untuk memaksimalkan kesamaan data dalam klaster dan mengurangi kesamaan data antar klaster. Pendekatan Pengelompokan *K-Means* memiliki beberapa manfaat, seperti kemudahan penggunaan dan kecepatan dalam memproses data dalam jumlah besar dengan cepat, bahkan untuk individu yang tidak memiliki pengalaman pemrograman. Puskesmas Mamsena dapat mengumpulkan data tentang masalah kesehatan balita secara real time dengan menggunakan metode ini. Pendekatan *K-Means* akan digunakan dalam penelitian ini untuk mengkategorikan data mengenai status gizi balita di Puskesmas Mamsena. Mengingat berbagai masalah yang disebutkan, penelitian ini diharapkan dapat mengatasi tantangan pengelolaan informasi yang berkaitan dengan kesejahteraan dan menawarkan kemajuan signifikan bagi mereka yang terlibat dalam administrasi kesejahteraan, khususnya dalam penyediaan dan pengumpulan data kesehatan balita. Hasilnya, penelitian ini tidak hanya akan membantu dalam pemusatan perhatian pada pengelompokan data tetapi juga akan berkontribusi atau meningkatkan inisiatif yang ditujukan untuk meningkatkan program gizi balita.

2. METODE PENELITIAN

2.1 Tahapan Penelitian

Tahapan eksplorasi merupakan rangkaian yang mencakup cara pelaksanaan penelitian dari awal hingga akhir. Berikut ini adalah alur dari penelitian ini :

Berikut uraian secara rinci dari tahapan penelitian:

a. Observasi Tempat Penelitian

Pada tahap awal, peneliti melakukan observasi atau pengamatan di tempat penelitian secara langsung dan cermat dengan tujuan untuk dapat mengetahui kondisi yang terjadi atau membuktikan kebenaran terkait permasalahan yang terjadi sesuai dengan topik pada penelitian ini.

b. Analisis Kebutuhan

Setelah melakukan observasi atau pengamatan di tempat penelitian, tahap selanjutnya yaitu analisis kebutuhan. Analisis kebutuhan dilakukan dengan tujuan untuk mengetahui kebutuhan terkait dengan penelitian ini. Analisis kebutuhan ini yang kemudian akan menjadi acuan dalam menyelesaikan masalah yang terdapat dalam penelitian ini. Terdapat dua tahapan analisis kebutuhan yaitu:

1. Identifikasi Masalah dan Tujuan

Tahap pertama analisis kebutuhan yaitu mengidentifikasi masalah yang ada di Puskesmas Mamsena. Salah satu masalah yang ada di puskesmas yaitu belum tersedia metode atau teknik untuk

mengelompokkan kondisi gizi balita secara efektif dan efisien. Dari masalah yang diidentifikasi, peneliti memutuskan tujuan dari penelitian ini untuk membuat pengelompokkan kondisi gizi balita dengan teknik *data mining* menggunakan metode *K-means Clustering*.

2. Rumusan Masalah

Pada tahapan perumusan masalah ini, sesuai dengan masalah yang sedang dialami oleh pihak puskesmas sendiri maka rumusan masalah yang akan di pecahkan dalam penelitian ini yaitu bagaimana implementasi metode *K-Means Clustering* dalam pengelompokkan kondisi gizi balita di Puskesmas Mamsena? bagaimana kontribusi atau nilai tambah dari implementasi metode ini khususnya dalam hal evaluasi atau program perbaikan gizi di Puskesmas Mamsena?

Dari hasil analisis kebutuhan kemudian didapatkan 5 *cluster* yang akan digunakan untuk mengelompokkan status gizi balita dengan menggunakan metode penentuan *cluster* atau K terbaik yaitu metode *elbow* atau metode siku. *Cluster* yang digunakan berjumlah 5 *cluster* mulai dari *cluster* 0 sampai *cluster* 4 dengan kelompok *cluster* yaitu *cluster* 0 = gizi buruk, *cluster* 1 = gizi kurang, *cluster* 2 = gizi cukup, *cluster* 3 = gizi baik dan *cluster* 4 = obesitas atau kegemukan. Data gizi balita di *cluster* atau dikelompokkan dengan menggunakan 5 fitur atau atribut yaitu dari data *antropometri* balita yang tercatat di Puskesmas Mamsena yaitu : Usia (bulan), Berat Badan (kg), Tinggi Badan (cm), Lingkar Lengan Atas (LiLA) dan Jenis Kelamin.

c. Pengumpulan Data

Tahapan selanjutnya yaitu pengumpulan data. Terdapat dua teknik pengumpulan data yang digunakan dalam penelitian yaitu:

1. Studi Pustaka

Studi pustaka yang digunakan dalam penelitian ini diperoleh melalui jurnal, buku, artikel maupun internet yang digunakan untuk mencari informasi yang relevan dengan penelitian ini.

2 Wawancara Pada tahap ini, peneliti berbicara langsung kepada petugas gizi di Puskesmas Mamsena mengenai topik penelitian yang diusulkan. Selain data, wawancara ini juga ditujukan untuk mengumpulkan informasi yang dibutuhkan dalam kajian ini.

3 Pengolahan dan analisis data

Tahapan ini merupakan proses sistematis untuk menguraikan, menginterpretasikan dan mengolah data mentah menjadi suatu kesimpulan atau informasi yang bermakna dan berguna.

Penelitian ini menggunakan data balita 3 tahun terakhir mulai dari tahun 2021, 2022 dan 2023 dengan populasi data sebanyak 2.664 data balita. Dari populasi tersebut kemudian di hitung menggunakan formula *slovin* untuk memperoleh jumlah sampel yang akan digunakan. Jumlah data sampel yang diperoleh sebanyak 348 data balita. Penelitian ini menggunakan data sampel yang ada yaitu data balita yang berjumlah 348 dikarenakan keterbatasan sumber daya manusia dari peneliti.

Penggunaan tahapan *K-means* yang digunakan pada penelitian ini yaitu pada tahapan untuk menentukan kondisi gizi balita balita mulai dari penentuan jumlah *cluster* atau K, menentukan titik pusat *cluster* atau *centroid* awal secara *random* atau acak, menghitung jarak setiap data ke *centroid* menggunakan rumus *De Euclidean*, kemudian dari jarak tersebut data di kelompokkan berdasarkan jarak terdekat untuk dikelompokkan ke dalam *cluster* tersebut, perhitungan dilakukan hingga setiap *cluster* bernilai optimal atau anggota pada suatu *cluster* tidak lagi berpindah ke *cluster* lain.

Metode *Elbow* digunakan untuk mengetahui jumlah K atau *cluster* terbaik yang akan digunakan untuk memperoleh nilai *error* atau SSE yang semakin kecil atau cenderung menurun dan pada jumlah K tersebut membentuk siku atau *elbow* maka jumlah K terbaik yang digunakan. Jumlah K atau *cluster* terbaik atau *cluster* yang membentuk siku berada pada K=5, maka pada penelitian ini menggunakan 5 *cluster* untuk mengelompokkan kondisi gizi balita di Puskesmas Mamsena.

Penggunaan formula *slovin* digunakan pada tahapan pengumpulan data yang digunakan untuk memperoleh data sampel dari jumlah populasi data. Pada tahapan ini peneliti hanya menggunakan sampel dari data tersebut dengan jumlah data sebanyak 348 data balita, dikarenakan keterbatasan tenaga dan kemampuan dari peneliti.

d. Laporan Hasil dan Evaluasi

Tahap terakhir dari penelitian ini yaitu laporan hasil dan evaluasi. Setelah pengolahan data di lakukan kemudian hasil dari pengolahan tersebut dibuatkan laporan dan kemudian dilakukan evaluasi terhadap hasil dari penelitian ini.

2.2 Landasan Teori

2.2.1 Data mining

Ada sejumlah cara untuk menganalisis dan mengekstrak informasi secara otomatis dalam teknik penambangan data [5].

Dalam teknik *data mining*, terdapat beberapa metode untuk menganalisis dan mengekstraksi informasi secara otomatis [3]. Selain itu, pengelompokan data digunakan dalam *data mining* untuk mengidentifikasi pola dan langkah-langkah selanjutnya. Menurut [4] tujuan dari *data mining* adalah untuk memfasilitasi kegiatan evaluasi dengan cara yang memastikan bahwa kegiatan tersebut memenuhi harapan. *Data mining* dapat digunakan dalam berbagai bidang, termasuk asuransi, pendidikan, keuangan, pemasaran, kesehatan, dan sebagainya. Ada dua sifat dalam *data mining* yaitu :

- a. Prediktif : menghasilkan modal berdasarkan kumpulan informasi yang dapat digunakan untuk mengukur nilai berbagai informasi. Teknik yang menggabungkan penambangan informasi
- b. Deskriptif : menghasilkan data baru dengan melacak contoh atau koneksi dalam informasi. Dalam penambangan informasi ekspresif.

2.2.3 K-Means

K-Means adalah strategi pengelompokan data non-progresif yang mengelompokkan data ke dalam setidaknya satu kelompok. Ini adalah salah satu metode komputasi pengelompokan. Pengelompokan adalah teknik yang umum digunakan yang semakin populer seiring dengan peningkatan volume data. *K* dipahami sebagai konstanta untuk jumlah pengelompokan yang optimal. *K* adalah nilai rata-rata dari kumpulan data, yang disebut sebagai kelompok dalam konteks ini. Salah satu strategi kerangka kerja segmentasi dalam penambangan data adalah pengelompokan *K-Means*, yang merupakan strategi analisis data yang menunjukkan satu bagian data. *K-Means* Menurut [8], pengelompokan adalah teknik non-hierarkis yang membagi data menjadi dua kelompok:

- a) menghitung jumlah *k* dalam set data;
- b) mencari tahu nilai pusat, atau *centroid*. Pada hasilnya, nilai *centroid* ditentukan secara acak atau dengan memilih nilai yang sesuai dengan nilai terbesar untuk kelompok tinggi dan nilai terendah untuk kelompok rendah.
- c) Tentukan jarak antara setiap titik data dan *centroid*. Jarak *Euclidean* adalah rumus yang digunakan untuk menghitung jarak *centroid* :

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

Rumusnya adalah $D_e = \text{Euclidean}$. Jarak $i = \text{Titik Objek}$ (s, t) = Jumlah Objek (x, y) = Titik Pusat Kluster
 Untuk membuat *centroid* baru, kumpulkan item berdasarkan seberapa jauh mereka dari *centroid* terdekat. Setelah menghitung jarak dari titik konvergensi sebelumnya dan membagi hasilnya dengan total atribut di setiap kelompok, *centroid* baru diperoleh. Rumus berikut dapat digunakan untuk menemukan *centroid* baru:

$$C_k \left(\frac{1}{n_k} \right) \sum d_i \quad (2)$$

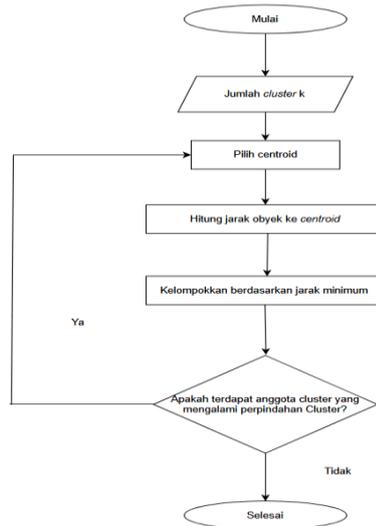
Keterangan :

n_k = jumlah dokumen dalam *cluster* k

d_i = dokumen dalam *cluster* k

- a) Lakukan iterasi seperti pada langkah ke dua hingga nilai setiap *cluster* optimal

Berikut merupakan *flowchart K-Means Clustering* yang umumnya digunakan:



Gambar 1. Flowchart K-Means Clustering

Teknik *K-Means Clustering* juga memiliki kelebihan dalam pengelompokan data karena dapat mengelompokkan sejumlah besar data dalam waktu singkat dan memiliki algoritma yang sederhana, mudah dipahami, dan diterapkan. Selain itu terdapat juga kekurangan yang dimiliki teknik *K-means* yaitu jika hasil penentuan *centroid* acak kurang tepat, hasil metode *K-Means Clustering* akan kurang optimal, sehingga cukup sulit untuk mencari data dengan jumlah dimensi yang besar.

2.2.4 Metode *Elbow*

Salah satu strategi untuk mencari jumlah kelompok terbaik adalah dengan melihat nilai setiap kelompok yang akan membentuk siku pada suatu titik dengan menggunakan metode siku atau sistem siku. Metode siku umumnya diberikan sebagai grafik frekuensi rendah atau nilai kesalahan [9]. Nilai K ideal dipilih dengan menggunakan metode ini. Untuk menggunakan metode siku, ada rumus yang dapat digunakan yang umumnya disebut nilai total WSS (*Internal Group Number of Squares*) atau biasa disebut SSE (*Total Square Error*) dengan rumus sebagai berikut :

$$SSE = \sum_k^k = 1 \sum_{x_i \in C_k} (x_i - \varphi_k)^2 \quad (4)$$

Keterangan :

C_k = Jumlah K yang terbentuk

K = banyaknya *cluster*

x_i = data x pada fitur ke-i

φ_k = rata-rata *cluster* pada nilai K

Metode *elbow* memiliki kelebihan karena dapat menentukan jumlah kluster terbaik dengan cepat dan efektif. Metode ini menggunakan SSE untuk menunjukkan jumlah kluster yang tepat, dan grafik SSE yang dihasilkan mudah dipahami. Metode *elbow* juga memiliki beberapa kekurangan, yaitu metode ini hanya dapat digunakan dengan algoritma *K-Means*, tidak cocok untuk semua jenis data, terutama data asimetris dengan banyak *noisy*, bergantung pada SSE yang dapat menyebabkan kesalahan saat menentukan jumlah kluster dan grafik yang dihasilkan terkadang ambigu atau tidak jelas.

2.2.5 *Davies Bouldin Index (DBI)*

Salah satu metode untuk mengevaluasi kualitas setiap kluster dalam pengelompokan data adalah *Indeks Davies-Bouldin*. Metode ini melihat seberapa baik setiap objek dalam suatu kluster. Kohesi dan pemisahan digabungkan dalam *Indeks Davies-Bouldin*. Pemisahan adalah jarak antara pusat kluster (*centroid*) dari berbagai kluster, sedangkan kohesi adalah derajat kedekatan setiap data dengan pusat kluster tempat data tersebut berada. Dengan menggunakan metode *K-Means*, metode ini digunakan untuk mengevaluasi kinerja dan mengukur akurasi [2]. Kluster semakin baik jika nilai DBI semakin kecil (non-negatif ≥ 0). DBI memiliki sejumlah manfaat, salah satunya dapat digunakan untuk mengevaluasi kluster dengan menghitung kuantitas dan atribut dataset yang ada untuk memberikan validasi yang baik. Kelemahan metode DBI adalah nilai akurasi DBI akan kurang baik atau

bahkan lebih buruk jika klaster yang ditentukan di awal tidak optimal. DBI juga dapat digunakan untuk menentukan nilai K terbaik untuk setiap K secara grafis, selain digunakan untuk mengevaluasi hasil klaster.

2.2.6 Formula Slovin

Persamaan Slovin adalah rumus yang digunakan dalam pengujian kuantitatif untuk menentukan jumlah informasi pengujian yang mencakup jumlah atau populasi informasi yang digunakan dengan ruang gerak tertentu. Rumus Slovin dijelaskan dalam paragraf berikut.

$$n = \frac{N}{1+N(e)^2} \quad (5)$$

Keterangan :

n = jumlah sampel yang dibutuhkan

N = jumlah populasi

e = margin kesalahan atau *error*

Setelah kuantitas pengujian diperoleh, untuk memutuskan setiap ukuran contoh ke dalam subgrup populasi berdasarkan luas setiap subgrup menggunakan resep distribusi relatif. Rumus untuk menentukan alokasi proporsional adalah sebagai berikut:

$$n_i = \frac{N_i}{N} \times n \quad (6)$$

Keterangan :

n_i = ukuran sampel untuk sub kelompok i

n = ukuran sampel total yang dihitung menggunakan rumus *slovin*

N_i = ukuran sub kelompok i

N = ukuran total populasi

2.2.7 Kondisi Gizi Balita

Pada tahun 2020, Kementerian Kesehatan Republik Indonesia menyatakan bahwa balita secara umum didefinisikan sebagai anak yang berusia di bawah lima tahun, atau bayi sebagai kelompok anak yang berusia 0-59 bulan atau 0-5 tahun. Bayi usia 0-28 hari, 0-11 bulan, dan 12-59 bulan semuanya dianggap sebagai usia balita. Kementerian Kesehatan Indonesia dan Organisasi Kesehatan Dunia mendefinisikan status atau kondisi gizi sebagai keadaan atau kondisi yang dihasilkan dari ketidakseimbangan antara asupan makanan dengan kebutuhan gizi yang dibutuhkan tubuh untuk pertumbuhan dan perkembangan pencernaan. Agar anak-anak tumbuh dan berkembang, gizi sangat penting. Masalah kesehatan dapat muncul jika kebutuhan gizi balita tidak terpenuhi. Anak-anak, misalnya, mudah lelah karena energi rendah, kelainan otak, dan penyebab lainnya. Status gizi bayi sangat penting karena pertumbuhan dan perkembangan balita memiliki dampak jangka panjang yang signifikan terhadap kesehatan mereka.

2.2.8 Puskesmas Mamsena

Puskesmas Mamsena merupakan pusat kesehatan masyarakat non rawat inap yang berada di kecamatan Insana Barat. Puskesmas Mamsena beralamat di Jl. Mamsena, Desa Atmen kecamatan Insana Barat. Salah satu program yang ada di puskesmas mamsena yaitu Upaya Kesehatan Masyarakat (UKM) program gizi. Program gizi adalah salah satu program esensial yang ada di puskesmas. Salah satu program gizi di puskesmas yaitu konseling gizi balita (*sumber : website manual mutu puskesmas mamsena : <https://www.scribd.com/document/434891692/MM>*).

Data gizi balita setiap tahun mengalami perubahan. Data yang banyak diperlukan tempat yang besar untuk menampung data tersebut, dan untuk lebih memudahkan dalam hal evaluasi dan penanganan untuk upaya peningkatan kesehatan balita, diperlukan pengelompokkan data untuk setiap balita yang ada di puskesmas. Di Puskesmas Mamsena belum tersedia metode atau cara yang efektif dan efisien dalam pengelompokkan data gizi balita tersebut, sehingga proses pengelompokkan data membutuhkan waktu dan tenaga yang cukup banyak.

2.2.9 Bahasa Pemrograman Python

Dalam bidang analisis data, bahasa pemrograman *Python* banyak digunakan. Hal ini dikarenakan kemudahan penggunaan *Python*. Selain itu, bahasa pemrograman *Python* bersifat sumber terbuka, artinya siapa pun dapat menggunakannya di sistem operasi apa pun berkat berbagai pustakanya, yang masing-masing memiliki tujuannya sendiri. *NumPy*, *Pandas*, *Matplotlib* dan *Scikit-learn* adalah contoh pustaka *Python* yang dapat digunakan untuk pemodelan statistik, visualisasi data, analisis data, dan pembelajaran mesin. Dalam berbagai bidang teknis,

termasuk ilmu data, penambangan data, pembelajaran mesin, kecerdasan buatan, dan pemrosesan gambar, Python digunakan secara luas. dan lain sebagainya. Selain itu, *Python* mudah diintegrasikan dengan teknologi lain yang terkait dengan analisis data, seperti basis data, alat big data, kerangka kerja web, dan sebagainya, untuk mengakses dan mengelola data dari berbagai sumber [6]. *Python* juga dapat digunakan untuk pengembangan aplikasi karena merupakan bahasa skrip yang mudah dipelajari, canggih, dan mudah beradaptasi. Karena sifatnya yang mudah ditafsirkan dan sintaksisnya yang sangat dinamis. Pola pemrograman seperti pemrograman berorientasi objek, imperatif, fungsional, dan prosedural semuanya didukung oleh *Python*.

3. HASIL DAN PEMBAHASAN

Penelitian ini ditujukan untuk mengelompokkan kondisi gizi balita di Puskesmas Mamsena menjadi lima kelompok, yaitu kelompok gizi buruk, gizi kurang, gizi cukup, gizi baik, dan kegemukan atau obesitas dengan menggunakan bahasa pemrograman *python*. Puskesmas diharapkan dapat memperoleh manfaat dari hasil penelitian ini, terutama dalam hal pengelompokan kondisi gizi balita secara efisien dan efektif.

3.1 Hasil Pengumpulan Data

Dari populasi data yang diperoleh dari tiga tahun terakhir mulai dari tahun 2021, 2022 dan 2023 sebanyak 2.664 data balita, data sampel yang digunakan dalam penelitian ini diperoleh melalui formula Slovin ($\alpha = 5\%$) dengan menggunakan persamaan (5) yaitu :

$$n = \frac{N}{1 + N(e)^2} = \frac{2.664}{1 + 2.664(0,05)^2} = 347,780 = 348$$

Setelah sampel dari keseluruhan data diketahui, kemudian ditentukan juga sampel pada setiap tahun data balita, dengan menggunakan rumus alokasi proposional menggunakan persamaan (2.6)

$$n_i = \frac{N_i}{N} \times n$$

Berikut merupakan perhitungan alokasi proposional untuk menentukan sampel data balita setiap tahun menggunakan formula persamaan (6) yaitu :

$$N = 2.664$$

$$n = 348$$

Tabel 1. Data Sampel Penelitian

No	Data Tahun ke-	N_i	n_i
1.	2021	885	116
2.	2022	905	118
3.	2023	874	114
	Total	2.664	348

Oleh karena itu, sebanyak 348 data digunakan sebagai sampel. Data antropometri (Usia, Berat Badan, Tinggi Badan, Lingkar Lengan Atas/LiLA, dan Jenis Kelamin) merupakan variabel atau atribut yang digunakan dalam penelitian ini.. Detail informasi balita yang digunakan dalam penelitian ini dapat dilihat pada Tabel 2 di bawah ini :

Tabel 2. Data Balita

No	Nama Balita	Jenis Kelamin	Usia (bulan)	Berat Badan (kg)	Tinggi Badan (cm)	LiLA (cm)
1	Avelino Reynaldi Tonbesi	L	25	10,1	80,5	14,5
2	Abraham Ka'auni	L	22	9,1	80,5	13
3	Detalia D. Amora Maumabe	P	28	7,7	72,2	14,5
4	Aldiano Andre Naikofi	L	18	8,7	81	12,8
5	Abraham F. Fallo	L	42	12,8	89,6	14,5
6	Sesilia A. Tas'au	P	20	9	79,5	12,7
7	Alexander J. Usboko	L	37	11,4	88,3	14,5
8	Helene Kelu	P	27	10,4	85,6	13,5
9	Jino K. Misa	L	46	13,6	95,7	14,5

10	Kaira Fautnine	P	48	11,2	92,4	13
11	Karmel A. Kurniati	P	27	12,3	87,6	16
12	Klarisa A. Kolo	P	23	7,6	71,5	13
13	Ludgardis G. Naisoko	P	18	7,8	67	14,5
14	Alfonsus R. Naisoko	L	19	8,4	69	15,5
15	Alfaro Sengkoen	L	37	11,2	87,2	13,2
...
348	Maria G Seran	P	29	7,7	73,5	12,8

Dari 348 data balita yang digunakan terdapat 52% balita yang berjenis kelamin laki-laki atau sama dengan 181 balita dan 48% balita yang berjenis kelamin perempuan atau sama dengan 167 balita. Selanjutnya untuk usia balita yang diteliti paling rendah yaitu balita yang berusia 8 bulan dan balita yang paling tua memiliki usia 54 bulan atau yang setara dengan 4 tahun 8 bulan. Rata-rata usia balita yang diteliti dari 348 data balita yaitu 28,23 bulan, atau dua tahun empat bulan. Balita yang diteliti memiliki berat rata-rata 18,1 kg, sedangkan balita yang diteliti memiliki berat rata-rata 2,9 kg. Dari total 348 data balita, berat rata-rata bayi yang diteliti adalah 10,70 kg. Selanjutnya tinggi badan balita yang diteliti paling rendah yaitu balita yang memiliki tinggi badan 37 cm dan balita yang memiliki tinggi badan maksimum yaitu balita dengan tinggi badan 106 cm. Rata-rata tinggi badan balita yang diteliti dari 348 data balita yaitu 82,77 cm. Kemudian dari jumlah 348 data balita yang diteliti balita yang memiliki lingkaran lengan atas (LiLA) yang paling rendah yaitu 9,7 cm dan balita yang memiliki LiLA maksimum yaitu 18 cm. Untuk mengonversi tipe data objek atribut gender dari tabel 2 ke numerik atau angka, diperlukan transformasi. Balita laki-laki diberi nilai satu, sedangkan balita perempuan diberi nilai dua. Lebih jauh, normalisasi diperlukan untuk mengendalikan keseragaman nilai antar fitur guna mencegah jarak yang signifikan atau tidak ada di antara data. Tabel 3 di bawah ini menampilkan data yang digunakan untuk pengelompokan berdasarkan hasil transformasi dan normalisasi yang dilakukan.

Tabel 3. Data Balita hasil Tranformasi dan Normalisasi

No	Nama Balita	Usia (Bulan)	Berat Badan (kg)	Tinggi Badan (cm)	LiLA	Jenis Kelamin
1	Avelino Reynaldi Tonbesi	0,354166667	0,473684211	0,630434783	0,578313253	0
2	Abraham Ka'auni Detalia D. Amora	0,291666667	0,407894737	0,630434783	0,397590361	0
3	Maumabe	0,416666667	0,315789474	0,510144928	0,578313253	1
4	Aldiano Andre Naikofi	0,208333333	0,381578947	0,637681159	0,373493976	0
5	Abraham F. Fallo	0,708333333	0,651315789	0,762318841	0,578313253	0
6	Sesilia A. Tas'au	0,25	0,401315789	0,615942029	0,361445783	1
7	Alexander J. Usboko	0,604166667	0,559210526	0,743478261	0,578313253	0
8	Helene Kelu	0,395833333	0,493421053	0,704347826	0,457831325	1
9	Jino K. Misa	0,791666667	0,703947368	0,850724638	0,578313253	0
10	Kaira Fautnine	0,833333333	0,546052632	0,802898551	0,397590361	1
11	Karmel A. Kurniati	0,395833333	0,618421053	0,733333333	0,759036145	1
12	Klarisa A. Kolo	0,3125	0,309210526	0,5	0,397590361	1
13	Ludgardis G. Naisoko	0,208333333	0,322368421	0,434782609	0,578313253	1
14	Alfonsus R. Naisoko	0,229166667	0,361842105	0,463768116	0,698795181	0
15	Alfaro Sengkoen	0,604166667	0,546052632	0,727536232	0,421686747	0
...
348	Maria G Seran	0,4375	0,315789474	0,528985507	0,373493976	1

3.2 Hasil Pengolahan Data

Penelitian ini bertujuan untuk mengelompokkan kondisi gizi balita ke dalam kelas-kelas tertentu dan juga untuk memudahkan pihak Puskesmas dalam upaya perbaikan gizi balita.

3.2.1 Import library python

Tahapan pertama yang dilakukan yaitu mengimport library yang dibutuhkan dalam proses *K-means clustering* untuk menentukan kondisi gizi balita. Proses *clustering* dilakukan menggunakan *software google colab*.

Library yang digunakan yaitu *numpy*, *matplotlib*, *pandas*, *seaborn* dan *sklearn*. Pada library *sklearn* terdapat beberapa modul yang digunakan yaitu *KMeans*, *metrics*, *spatial distance*, *datasets* dan *davies bouldin score*.

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
import sklearn
from sklearn.cluster import KMeans
from sklearn import metrics
from scipy.spatial.distance import cdist
from sklearn.datasets import make_blobs
from sklearn.cluster import KMeans
from sklearn.metrics import davies_bouldin_score
```

Gambar 2. Import library python

3.2.2 Menginput dan Membaca Dataset

Setelah tahapan seperti pada gambar 3 dilakukan, langkah selanjutnya yaitu menginput *dataset* dan membaca *dataset* yang digunakan yaitu data balita yang tercatat di Puskesmas Mamsena sebanyak 348 data

No	Nama Balita	Usia (bulan)	Berat Badan (kg)	Tinggi Badan (cm)	LiLA (cm)	Jenis Kelamin
0	Avelino Reynaldi Tonbesi	25	10.1	80.5	14.5	1
1	Abraham Ka'auni	22	9.1	80.5	13.0	1
2	Detalia D. Amora Maumabe	28	7.7	72.2	14.5	2
3	Aldiano Andre Naikofi	18	8.7	81.0	12.8	1
4	Abraham F. Fallo	42	12.8	89.6	14.5	1
...
343	Seravina Naiheli	19	9.1	79.5	13.3	2
344	Skolastika K. Lina	23	12.8	88.6	16.0	2
345	Stevania Tahoni	28	9.1	78.5	13.5	2
346	Theresia I. Metboki	12	15.2	99.8	15.5	2
347	Maria G Seran	29	7.7	73.5	12.8	2

348 rows x 7 columns

Gambar 3. Dataset

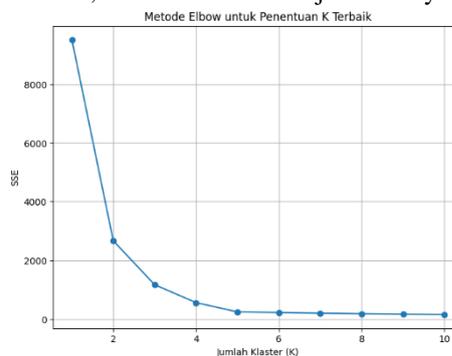
Terdapat 348 baris informasi dan 7 segmen yang terdiri dari bagian untuk nomor, nama bayi, orientasi, usia (bulan), berat (kg), level (cm) dan sirkuit lengan atas (cm) seperti yang ditampilkan pada Gambar 3.

3.2.4 Proses K-means Clustering

Tahap selanjutnya setelah *training dataset* yaitu proses *K-means clustering*. Terdapat beberapa tahapan yang akan di lakukan pada proses *K-means clustering* yaitu :

- a. Penentuan jumlah K terbaik

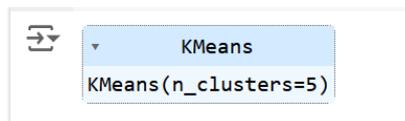
Pada tahap ini, penentuan jumlah kelompok (K) terbaik diselesaikan dengan menggunakan teknik siku (*elbow method*). Jumlah K terbaik dari metode siku dapat dilihat pada Gambar 4, di mana semakin rendah nilai inersia atau error yang dihasilkan, maka semakin baik jumlah K yang dipilih.



Gambar 4. K terbaik dengan metode Elbow

Dari output grafik metode *elbow* seperti gambar 7 dapat dilihat jumlah K mulai dari *range* 1 sampai 10 mengalami penurunan nilai *inertia*. *Inertia* merupakan sebuah metrik untuk mengukur seberapa baik data yang akan di *cluster* atau yang biasa disebut WCSS atau SSE, dimana penurunan nilai *inertia* mengalami penurunan yang dapat memberikan indikasi jumlah *cluster* yang efektif. Dari hasil penentuan jumlah K

terbaik menggunakan metode *elbow* jumlah K yang digunakan yaitu K=5 dimana jumlah K=5 memiliki nilai inertias atau *error* yang cenderung menurun atau semakin kecil. Setelah di tentukan K terbaik langkah selanjutnya yaitu menginput jumlah K yang akan digunakan pada proses *K-means* seperti pada gambar 5 berikut :

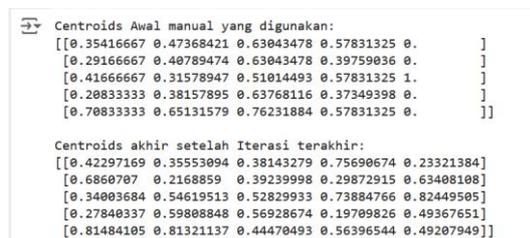


Gambar 5. K optimal = 5

Jadi, jumlah kelompok *cluster* (K) yang ditentukan untuk melakukan klasterisasi kondisi gizi balita adalah 5 *cluster*.

b. Menentukan *centroid* awal (titik pusat *cluster*)

Tahapan selanjutnya setelah menentukan jumlah K terbaik dalam proses *K-means clustering* yaitu menentukan *centroid* awal atau titik pusat *cluster*. Penentuan *centroid* awal dilakukan menggunakan secara acak dimana data balita ke-1 sampai 5 yang merupakan titik awal pusat *cluster* atau *centroid*



Gambar 6. Centroid awal dan akhir

Tampilan *output* seperti gambar di atas menampilkan titik *centroid* awal untuk iterasi ke-1 dan juga *centroid* untuk iterasi ke terakhir untuk masing-masing *cluster* mulai dari *cluster* 0 sampai *cluster* 4.

c. Menghitung jarak masing-masing data

Perhitungan manual di *excel* dilakukan untuk membandingkan hasil *cluster* dengan program *python*. Perhitungan dilakukan sebanyak 6 iterasi hingga masing-masing *cluster* bernilai optimal dan nilai SSE cenderung menurun. Berikut merupakan hasil perhitungan di *excel* pada iterasi ke 6

Tabel 4. Hasil perhitungan manual di *excel* pada iterasi ke-6

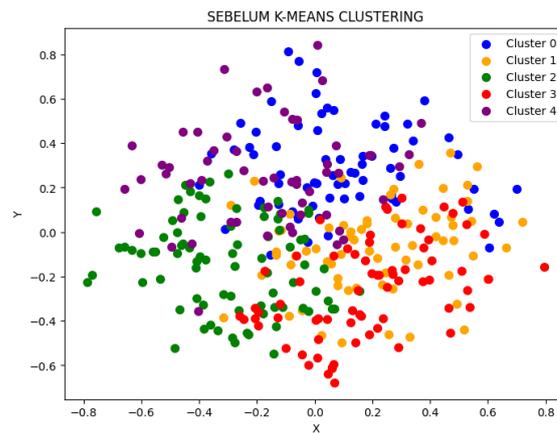
Nama Balita	Usia (Bulan)	Berat Badan (kg)	Tinggi Badan (cm)	LiLA	Jenis Kelamin	Kelompok Cluster
Avelino Reynaldi						
Tonbesi	0,354166667	0,473684211	0,630434783	0,578313253	0	0
Abraham Ka'auni	0,291666667	0,407894737	0,630434783	0,397590361	0	0
Detalia D. Amora						
Maumabe	0,416666667	0,315789474	0,510144928	0,578313253	1	1
Aldiano Andre Naikofi	0,208333333	0,381578947	0,637681159	0,373493976	0	0
Abraham F. Fallo	0,708333333	0,651315789	0,762318841	0,578313253	0	2
Sesilia A. Tas'au	0,25	0,401315789	0,615942029	0,361445783	1	1
Alexander J. Usboko	0,604166667	0,559210526	0,743478261	0,578313253	0	2
Helene Kelu	0,395833333	0,493421053	0,704347826	0,457831325	1	3
Jino K. Misa	0,791666667	0,703947368	0,850724638	0,578313253	0	2
Kaira Fautnine	0,833333333	0,546052632	0,802898551	0,397590361	1	4
Karmel A. Kurniati	0,395833333	0,618421053	0,733333333	0,759036145	1	3
Klarisa A. Kolo	0,3125	0,309210526	0,5	0,397590361	1	1
Ludgardis G. Naisoko	0,208333333	0,322368421	0,434782609	0,578313253	1	1
Alfonsus R. Naisoko	0,229166667	0,361842105	0,463768116	0,698795181	0	0
Alfaro Sengkoen	0,604166667	0,546052632	0,727536232	0,421686747	0	2

Nama Balita	Usia (Bulan)	Berat Badan (kg)	Tinggi Badan (cm)	LiLA	Jenis Kelamin	Kelompok Cluster
Maria G Seran	0,4375	0,315789474	0,528985507	0,373493976	1	1

d. Hasil Cluster

Tahapan selanjutnya setelah menghitung jarak data dengan titik *centroid* yaitu menampilkan hasil *cluster*. Hasil *cluster* ditampilkan melalui 2 tahapan yaitu menampilkan *dataset* sebelum *dicluster* dan menampilkan *dataset* setelah *dicluster*

1. Menampilkan *dataset* sebelum di *cluster*



Gambar 7. Tampilan data balita sebelum di *cluster*

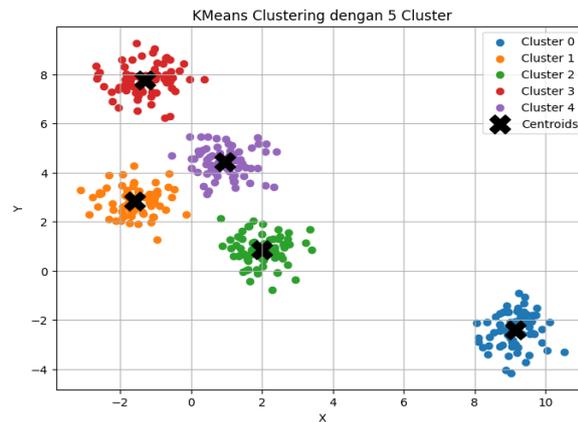
Dari tampilan seperti gambar 11 dapat dilihat data balita tidak beraturan dan belum berada dalam kelompok data yang memiliki kemiripan antar data, maka perlu dilakukan *clustering* agar dapat mengelompokkan data ke dalam kelas atau kelompok tertentu yang memiliki kemiripan antar data.

2. Menampilkan *dataset* setelah di *cluster*

```
array([0, 0, 1, 0, 2, 1, 2, 3, 2, 4, 3, 1, 1, 0, 2, 1, 4, 2, 4, 0, 0, 2,
4, 4, 2, 4, 1, 1, 1, 1, 4, 1, 2, 4, 2, 0, 0, 0, 0, 4, 0, 0, 4, 2,
4, 1, 4, 1, 3, 3, 0, 0, 2, 0, 2, 2, 2, 1, 3, 0, 1, 3, 0, 0, 2, 0,
0, 0, 2, 2, 2, 4, 0, 1, 1, 0, 2, 2, 3, 4, 4, 2, 4, 0, 2, 2, 3, 0,
0, 0, 2, 0, 2, 2, 1, 2, 2, 4, 1, 4, 3, 4, 2, 0, 0, 0, 2, 2, 0, 0,
4, 4, 4, 3, 0, 1, 2, 2, 4, 1, 2, 4, 3, 3, 4, 1, 2, 2, 2, 0, 2,
2, 1, 0, 2, 4, 2, 0, 0, 1, 4, 1, 1, 4, 1, 4, 1, 0, 2, 3, 2, 0, 0,
2, 2, 1, 0, 0, 2, 3, 4, 2, 0, 0, 3, 3, 3, 1, 4, 0, 4, 4, 1, 0, 0,
2, 0, 4, 2, 2, 0, 2, 2, 4, 0, 3, 0, 4, 4, 1, 0, 1, 4, 4, 4, 1, 1,
0, 2, 1, 1, 1, 1, 3, 1, 3, 1, 1, 1, 0, 0, 2, 0, 0, 0, 3, 1, 1, 3,
3, 1, 3, 3, 3, 0, 0, 2, 0, 2, 0, 0, 0, 2, 0, 0, 2, 2, 0, 0, 0, 2,
2, 0, 4, 4, 4, 4, 4, 3, 1, 4, 3, 3, 4, 1, 3, 1, 2, 2, 2, 2, 0, 2,
2, 0, 0, 0, 0, 0, 3, 4, 3, 3, 2, 2, 4, 3, 3, 3, 3, 4, 3, 3, 3, 1,
0, 0, 0, 2, 2, 2, 2, 0, 2, 0, 2, 2, 0, 0, 0, 0, 0, 1, 3, 3,
4, 3, 2, 2, 2, 0, 0, 2, 0, 0, 3, 3, 1, 1, 1, 1, 1, 3, 0, 0, 2,
2, 2, 3, 3, 0, 0, 3, 3, 3, 3, 3, 3, 1, 1, 3, 1, 3, 1], dtype=int32)
```

Gambar 8. Hasil *cluster* dalam bentuk *array*

Output dari hasil *cluster* seperti gambar 12 yaitu kelompok *cluster* dalam empat kelompok dan di tampilkan dalam bentuk *array* mulai dari 0 sampai dengan 4 dimana jumlah *cluster* balita yaitu 5 *cluster*. Dari *array* diatas sudah terlihat kelompok *cluster* pada setiap data pada 348 data balita yang *dicluster*.



Gambar 9. Tampilan data balita setelah di *cluster*

Gambar 13 menunjukkan hasil *cluster* pada masing-masing kelompok. Terdapat 5 cluster dimana *cluster* 0 ditandai dengan titik berwarna biru, *cluster* 1 ditandai dengan titik berwarna orange, *cluster* 2 ditandai dengan titik berwarna hijau, *cluster* 3 ditandai dengan titik berwarna merah dan *cluster* 4 berwarna ungu sedangkan masing-masing *centroid* pada ke-5 *cluster* ditandai dengan tanda silang berwarna hitam. Setelah hasil *cluster* diketahui langkah selanjutnya yaitu menghitung presentase dari masing-masing *cluster*. Berikut hasil dari presentasi untuk setiap *cluster* kondisi gizi balita

```

↔ Jumlah data dan persentase per cluster:
Cluster 0:
  Jumlah data: 98
  Persentase: 28.16%

Cluster 1:
  Jumlah data: 59
  Persentase: 16.95%

Cluster 2:
  Jumlah data: 83
  Persentase: 23.85%

Cluster 3:
  Jumlah data: 56
  Persentase: 16.09%

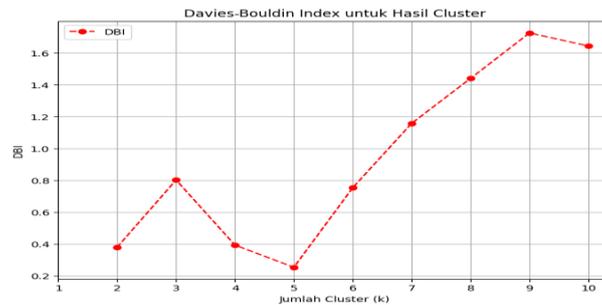
Cluster 4:
  Jumlah data: 52
  Persentase: 14.94%
    
```

Gambar 10. Hasil presentase setiap cluster

Dari hasil presentasi setiap *cluster* dapat disimpulkan bahwa *cluster* nol memiliki presentase = 28, 16 % atau sama dengan 98 data balita, *cluster* satu memiliki presentase = 16, 95 % atau sama dengan 59 data balita, *cluster* dua memiliki presentase = 23, 85 % atau sama dengan 83 data balita, *cluster* tiga memiliki presentase = 16, 09% atau sama dengan 56 data balita dan *cluster* empat memiliki presentase = 14, 94% atau sama dengan 52 data balita.

3.2.5 Hitung akurasi

Tahap terakhir dalam proses *clustering* yaitu evaluasi hasil *cluster* menggunakan DBI (*Davies Bouldin Index*). Evaluasi hasil *clustering* dapat ditampilkan dalam bentuk grafik juga. Berikut tampilan untuk akurasi *clustering* kondisi gizi balita mulai dari *range* K=2 sampai K=10



Gambar 11. Grafik evaluasi hasil cluster dengan DBI

Pada penelitian menggunakan jumlah K optimal = 4 cluster dimana dari hasil perhitungan akurasi seperti gambar 16 menunjukkan hasil K=5 yang memiliki akurasi paling rendah. Presentase hasil akurasi untuk K=5 dapat dilihat pada gambar 17 berikut :

Davies-Bouldin Index (DBI) untuk k=5: 0.2520626580022645

Gambar 12. Nilai DBI untuk K=4

Dari nilai evaluasi DBI pada gambar 17 dapat disimpulkan bahwa hasil *clustering* kondisi gizi balita dengan jumlah K=5 memiliki kualitas yang sangat baik karena bernilai hampir mendekati nol.

3.3 Pembahasan

Di Puskesmas Mamsena, metode *clustering K-means* digunakan untuk mengelompokkan status gizi balita secara akurat dan efisien untuk mengetahui status gizi balita. Hal utama yang dilakukan adalah pengumpulan data. Data yang digunakan adalah data antropometri balita, khususnya umur (bulan), berat badan (kg), tinggi badan (cm) dan LiLA (cm) serta orientasi. Setelah data terkumpul, tahap selanjutnya adalah menentukan jumlah kelompok (K) yang paling tepat untuk digunakan. Metode *Elbow* menghasilkan jumlah kelompok (K) yang optimal dalam penelitian ini, dengan jumlah kelompok optimal sebanyak lima kelompok. Dengan estimasi yang dilakukan dengan menggunakan alat WEKA, penelitian sebelumnya oleh [11] juga menghasilkan kelompok ideal dan memiliki hasil ideal sebanyak lima kelompok. Sementara itu, penelitian yang dilakukan oleh [12] menggunakan jumlah K yang digunakan, yakni 5 kelompok, dimana data yang digunakan berjumlah 150 data balita yang digunakan untuk siklus pengumpulan dan hasil akurasi yang dinilai dengan *Z-Score* menunjukkan tingkat ketepatan yang rendah dari perhitungan K-strategi. Rumus *De Euclidean Distance* digunakan untuk menghitung jarak setiap data ke titik pusat terdekat di *Excel* secara manual setelah titik pusat grup awal (centroid) dipilih secara acak. Jarak grup terdekat ditambahkan ke jarak total. Estimasi dilakukan hingga semua informasi di setiap kumpulan tidak berpindah ke kumpulan lain atau memiliki nilai optimal. Perhitungan dilakukan dalam studi ini hingga iterasi keenam, di mana setiap komponen di setiap grup memiliki nilai ideal.

Kontribusi dari implementasi metode *K-means clustering* dalam menentukan kondisi gizi balita dapat memberikan manfaat yang berarti bagi pihak Puskesmas. Dari hasil *clustering* yang diperoleh dari penelitian ini, pihak Puskesmas dapat mengidentifikasi kelompok balita yang memerlukan perhatian khusus seperti balita yang berada dalam kelompok gizi buruk dan gizi kurang, disamping itu penelitian yang dilakukan oleh [7] juga dapat memberikan kontribusi dalam upaya perbaikan kondisi gizi balita dengan evaluasi terhadap hasil *clustering* penentuan kondisi gizi balita memungkinkan perubahan kondisi gizi balita dari waktu ke waktu sehingga dapat memberikan umpan balik yang bermanfaat untuk upaya program yang akan meningkatkan gizi balita di masa mendatang. Selain itu data *cluster* yang akurat dapat membantu merancang strategi yang lebih tepat sasaran dan meminimalkan risiko kekurangan gizi di masa depan, sehingga meningkatkan efektivitas dan efisiensi program perbaikan gizi balita.

4. KESIMPULAN

Dalam menentukan kondisi gizi balita di Puskesmas Mamsena menggunakan metode *K-means Clustering* yang telah dilakukan, terdapat 3 kesimpulan dari hasil penelitian ini yaitu :

- Berdasarkan hasil penelitian untuk menentukan status kesehatan balita di Puskesmas Mamsena dengan teknik *data mining* menggunakan *K-means clustering*, terdapat 5 kelompok yang digunakan yaitu kurang sehat, gizi kurang, gizi cukup, gizi baik dan obesitas. Metode *elbow* dan DBI digunakan untuk menentukan nilai K

terbaik, dan ini menghasilkan jumlah kelompok. Dari jumlah tersebut sampel data yang digunakan sebanyak 348 data balita yang dikelompokkan mejadi 5 cluster dengan presentase masing-masing cluster yaitu : 5 cluster dengan presentase masing-masing cluster yaitu cluster nol memiliki presentase = 28, 16 % atau sama dengan 98 data balita, cluster satu memiliki presentase = 16, 95 % atau sama dengan 59 data balita, cluster dua memiliki presentase = 23, 85 % atau sama dengan 83 data balita, cluster tiga memiliki presentase = 16, 09% atau sama dengan 56 data balita dan cluster empat memiliki presentase = 14, 94% atau sama dengan 52 data balita. Dari hasil cluster yang diperoleh kemudian dievaluasi menggunakan metode DBI untuk mengetahui kualitas dari hasil cluster yang diperoleh. Nilai akurasi yang diperoleh dengan jumlah K=5 yaitu 0,2520626580022645. Dari hasil cluster yang diperoleh dapat disimpulkan bahwa jika data yang digunakan berjumlah banyak maka jumlah cluster atau kelompok juga berjumlah banyak agar akurasi atau kualitas dari hasil clustering semakin baik. Semakin kecil nilai akurasi atau mendekati nilai nol semakin baik kualitas dari hasil cluster tersebut.

- b. Dari hasil penelitian yang diperoleh hasil perhitungan manual dan python adalah sesuatu yang serupa, sehingga teknik K-means clustering merupakan metode yang efektif dan efisien untuk menganalisis data.
- c. Penelitian ini dapat memberikan nilai tambah atau kontribusi bagi pihak Puskesmas Mamsena dalam hal pemantauan ataupun evaluasi program perbaikan gizi balita di tingkat Puskesmas. Dengan adanya penelitian ini dapat memberikan efisiensi dan efektivitas dalam penentuan kondisi gizi balita.

5. SARAN

Ada beberapa saran dari penelitian ini, khususnya:

- a. Dapat mengkolaborasikan metode *K-means* dengan metode *Clustering* lainnya seperti *K-Medoids*, *Fuzzy C-Means* dan lain sebagainya untuk dapat memperoleh hasil cluster dengan kualitas yang lebih baik, efektif dan efisien.
- b. Dapat mengembangkan hasil dari penelitian ini ke dalam *website* agar lebih efektif dalam memvisualisasikan dan mengeksplorasi data.

UCAPAN TERIMA KASIH

Terimakasih yang berlimpah penulis ucapkan kepada semua pihak yang membantu dalam penyelesaian tugas akhir ini. Pihak Kampus, Pihak Puskesmas Mamsena yang sudah memeberikan izin dan mempermudah penulis dalam menyelesaikan tugas akhir.

REFERENCES

- [1] A. Ali, "Clustering Data Antropometri Balita Untuk Menentukan Status Gizi Balita Di Kelurahan Jumpat Rejo Sukodono Sidoarjo," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 3, pp. 395–407, 2020, doi: 10.35957/jatisi.v7i3.530.
- [2] C. Rahmat, H. Permatasari, E. Rasywir, and Y. Pratama, "Penerapan K-Means Untuk Clustering Kondisi Gizi Balita Pada Posyandu," vol. 7, no. 1, pp. 207–213, 2023, doi: 10.30865/mib.v7i1.5142.
- [3] D. Yudha, P. Widodo, and D. Febiharsa, "ANALISIS NILAI GIZI BALITA DI DESA MANGUNSARI KECAMATAN GUNUNGPATI KOTA SEMARANG DENGAN ALGORITMA K-MEANS CLUSTERING UNTUK PENCEGAHAN STUNTING Tumbuh kembang balita yang baik merupakan salah satu faktor pendukung kemajuan suatu negara . Akan tetapi persma," *Edustems*, vol. 1, no. 1, pp. 415–428, 2023.
- [4] E. R. Wijhati, R. Nuzuliana, and M. L. E. Pratiwi, "Analisis status gizi pada balita stunting," *J. Kebidanan*, vol. 10, no. 1, p. 1, 2021, doi: 10.26714/jk.10.1.2021.1-12.
- [5] I. dan I. A. S. Ahmad Zaki, "Penerapan K-Means Clustering dalam Pengelompokan Data (Studi Kasus Profil Mahasiswa Matematika FMIPA UNM)," 2022. [Online]. Available: <http://www.ojs.unm.ac.id/jmathcos>
- [6] K. Zannah, S. Sumarno, Z. M. Nasution, I. Parlina, and I. P. Sari, "MODEL CLUSTERING MENGGUNAKAN ALGORITMA K-MEAN DALAM MENENTUKAN KRITERIA KONDISI GIZI BALITA DAN ANAK," *JUTEKIN (Jurnal Tek. Inform.)*, vol. 10, no. 1, Jun. 2022, doi: 10.51530/jutekin.v10i1.568.
- [7] Lilik Inayati, Stefanny Surya Nagari, "IMPLEMENTASI CLUSTERING MENGGUNAKAN METODE K-MEANS PADA UNTUK MENENTUKAN STATUS GIZI BALITA," *J. Biometrika dan Kependud.*, vol. 9, no. 1, pp. 62–68, 2020, doi: 10.20473/jbk.v9i1.2020.62.
- [8] M. Miranda, N. Rahaningsih, and ..., "Analisis Clustering Data Anak Balita di Posyandu Kampung Sukarame Menggunakan Algoritma K-Means," *J. Inform. ...*, vol. 6, no. 1, pp. 136–141, 2024, [Online]. Available: <https://publikasiilmiah.unwahas.ac.id/JINRPL/article/view/10256>
- [9] N. A. Maori and E. Evanita, "Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 14, no. 2, pp. 277–288, 2023, doi: 10.24176/simet.v14i2.9630.

- [10] N. Komang Sri Julyantari *et al.*, “Implementasi K-Means Untuk Pengelompokan Status Gizi Balita (Studi Kasus Banjar Titih) Implementation of K-Means for Clustering the Nutritional Status of Toddlers (Banjar Titih Case Study),” *J. Janitra Inform. dan Sist. Inf.*, vol. 1, no. 2, pp. 92–101, 2021, doi: 10.25008/janitra.
- [11] R. Ranjawali, A. C. Talakua, and R. T. Abineno, “Clustering Stunting Pada Balita Dengan Metode K- Means Di Puskesmas Kanatang,” *SATI Sustain. Agric. Technol. Innov.*, pp. 80–92, 2023, [Online]. Available: <https://ojs.unkriswina.ac.id/index.php/semnas-FST/article/view/587/324>
- [12] S. Narulita, P. Prihati, A. Tigor Oktaga, and A. Eka Widyantoro, “Performansi Algoritma Clustering K-Means untuk Penentuan Status Malnutrisi pada Balita,” *J. Informasi, Sains, dan Teknol.*, vol. 06, no. 1, Dec. 2023.
- [13] Ua *et al.*, “Penggunaan Bahasa Pemrograman Python Dalam Analisis Faktor Penyebab Kanker Paru-Paru,” *J. Publ. Tek. Inform.*, vol. 2, no. 2, pp. 88–99, Jul. 2023, doi: 10.55606/jupti.v2i2.1742.
- [14] V. Syaputri, D. Hartama, F. Angraini, M. Safii, and R. Dewi, “PENERAPAN ALGORITMA K-MEANS UNTUK MENENTUKAN STATUS GIZI BALITA (STUDI KASUS: PUSKESMAS KECAMATAN JAWA MARAJA BAH JAMBI),” 2022.