

Penerapan Algoritma K-Means Dalam Proses Klasterisasi Balita Stunting Di Kecamatan Geyer

Arindhya Ratna Hayuningtyas^{*1}, Dhika Malita Puspita Arum², Eko Supriyadi³

^{1,2,3}Program Studi Ilmu Komputer, Universitas An Nuur, Grobogan, Indonesia

e-mail: ^{*1}hayuarin@gmail.com, ²dhika.malita11@gmail.com, ³ekalaya56@gmail.com

Abstrak

Stunting merupakan permasalahan gizi kronis akibat kekurangan nutrisi jangka panjang dan infeksi berulang, yang menghambat pertumbuhan anak. Di Indonesia, angka prevalensi stunting masih berada pada tingkat yang cukup tinggi dan belum memenuhi target nasional, yaitu 14% pada tahun 2024. Jawa Tengah mencatat angka 20,7% pada tahun 2023. Kabupaten Grobogan melaporkan 9,15% balita telah terindikasi stunting. Penanganan stunting membutuhkan pendekatan berbasis data, salah satunya melalui algoritma K-Means Clustering yang mampu mengelompokkan data secara efektif. Penelitian dengan metode tersebut untuk mengelompokkan balita stunting di Kecamatan Geyer. Tiga proses dilakukan secara berurutan: klasterisasi berdasarkan usia, klasterisasi berdasarkan kombinasi usia, berat badan, dan tinggi badan, serta analisis distribusi wilayah berdasarkan hasil klaster. Hasil klaster menunjukkan bahwa balita dapat dikelompokkan ke dalam beberapa klaster dengan kemiripan karakteristik pertumbuhan. Evaluasi menggunakan Silhouette Coefficient menghasilkan nilai yang cukup optimal di setiap prosesnya. Dengan nilai evaluasi yang dihasilkan menggunakan Silhouette Coefficient, Algoritma K-Means mampu memberikan gambaran awal mengenai karakteristik balita stunting dan dapat dimanfaatkan untuk mendukung perencanaan intervensi gizi secara lebih terstruktur.

Kata kunci— k-means, klastering, silhouette coefficient, stunting

Abstract.

Stunting is a chronic nutritional problem caused by long-term nutritional deficiencies and repeated infections, which inhibit child growth. In Indonesia, the prevalence of stunting remains at a fairly high level and has not met the national target of 14% by 2024. Central Java recorded a rate of 20.7% in 2023. Grobogan Regency reported that 9.15% of toddlers were indicated to be stunted. Addressing stunting requires a data-based approach, one of which is through the K-Means Clustering algorithm, which is capable of grouping data effectively. Research using this method was conducted to group stunted toddlers in Geyer District. Three processes were carried out sequentially: clustering based on age, clustering based on a combination of age, weight, and height, and analysis of regional distribution based on cluster results. The cluster results showed that toddlers could be grouped into several clusters with similar growth characteristics. Evaluation using the Silhouette Coefficient produced fairly optimal values in each process. With the evaluation value generated using the Silhouette Coefficient, the K-Means Algorithm is able to provide an initial overview of the characteristics of stunted toddlers and can be used to support more structured nutritional intervention planning

Keywords— clustering, k-means, silhouette coefficient, stunting

1. PENDAHULUAN

Usia balita merupakan masa krusial, karena dalam tahap ini anak mengalami pertumbuhan fisik dan psikomotorik yang signifikan. Balita memerlukan perhatian khusus sebab di fase ini otak mereka tumbuh sekitar 80%. Balita harus memiliki asupan makanan yang seimbang selama periode ini untuk memastikan mereka mendapatkan gizi yang optimal [1]. Stunting adalah gangguan gizi kronis yang disebabkan oleh asupan gizi yang tidak memadai, mengakibatkan anak-anak memiliki tinggi badan di bawah standar untuk usia mereka dan mempengaruhi perkembangan motorik mereka [2], [3], [4].

Pada tahun 2022, jumlah balita stunting di seluruh dunia mencapai 148,1 juta, dan sekitar 6,3 juta di antaranya berasal dari Indonesia [5]. Meskipun prevalensi stunting nasional menunjukkan tren penurunan dari 31,8% (2020) menjadi 21,6% (2022), angka tersebut masih tinggi untuk mencapai target 14% pada tahun 2024 [6]. Data terbaru dari Survei Status Gizi Indonesia (SSGI) 2024 menunjukkan adanya kemajuan signifikan dengan prevalensi stunting nasional berhasil turun menjadi 19,8% [7].

Untuk mengatasi permasalahan yang kompleks ini, pendekatan berbasis data menjadi esensial. data mining adalah pendekatan analisis data yang bertujuan untuk menemukan pola tersembunyi dan informasi penting dari kumpulan data besar, dan saat ini banyak digunakan untuk mendukung pengambilan keputusan prediktif [8]. Berbagai teknik dalam data mining, seperti klasterisasi, memainkan peran penting dalam mengorganisasi data yang kompleks dan mengidentifikasi kelompok-kelompok individu dengan atribut serupa, sehingga memungkinkan perancangan intervensi yang lebih spesifik dan efisien [9].

Salah satu teknik yang sering digunakan adalah clustering, yaitu metode pengelompokan data ke dalam kluster homogen tanpa memerlukan label kelas [10]. Metode ini dikenal karena kesederhanaan dan kecepataannya dalam mengelompokkan data numerik. Prosesnya bersifat iteratif, di mana titik data awalnya dikelompokkan berdasarkan kedekatan geometrisnya dengan centroid kluster, dan kemudian centroid tersebut diperbarui hingga posisi kluster menjadi stabil [11].

Berbagai studi sebelumnya juga telah membuktikan bahwa K-Means merupakan metode yang efektif dalam konteks analisis data di bidang kesehatan. Misalnya, penelitian yang dilakukan oleh Sintawati pada tahun 2024 menerapkan K-Means untuk mengklasifikasikan status gizi balita di DKI Jakarta, yang hasilnya dapat digunakan untuk mengidentifikasi area yang membutuhkan perhatian lebih dan strategi intervensi yang lebih terfokus [12]. Demikian pula, penelitian oleh Stefanelli, yang menggunakan K-Means untuk mengkategorikan data balita berdasarkan tinggi dan berat badan, menunjukkan potensi algoritma ini dalam analisis status gizi [13]. Studi-studi ini menunjukkan bahwa K-Means adalah alat yang andal untuk mengubah data mentah menjadi informasi yang dapat ditindaklanjuti, membantu para pemangku kepentingan dalam merancang intervensi yang lebih tepat sasaran.

Kontribusi utama penelitian ini terletak pada penerapan algoritma K-Means dalam konteks pengelompokan balita stunting. Selain itu, penelitian ini akan memperkaya khazanah keilmuan dalam bidang *data mining* dan aplikasinya pada permasalahan kesehatan masyarakat, khususnya gizi, dengan memanfaatkan algoritma klasterisasi K-Means. Selain itu, penelitian ini juga akan menunjukkan efektivitas penggunaan metode Silhouette Coefficient sebagai alat evaluasi untuk klasterisasi data stunting. Dengan demikian, studi ini dapat membantu pemerintah atau pemangku kepentingan, khususnya Pemerintah Daerah Kabupaten Grobogan, Kecamatan Geyer, dan Dinas Kesehatan dalam menyusun perencanaan dan implementasi kebijakan yang lebih efektif untuk mengurangi stunting.

2. METODE PENELITIAN

Tahap ini menjelaskan secara detail langkah-langkah sistematis yang ditempuh dalam penelitian klasterisasi balita stunting. Metodologi ini dirancang untuk menjawab tujuan penelitian melalui pendekatan kuantitatif dan analisis data mining. Pembahasan meliputi jenis penelitian yang digunakan, metode pengumpulan data, identifikasi perangkat lunak dan perangkat keras, serta tahapan implementasi Algoritma K-Means secara terperinci.

2.1 Jenis Penelitian

Penelitian ini menggunakan metode kuantitatif. Metode pendekatan ini dipilih agar interpretasi hasil bersifat objektif, sistematis, dan dapat diuji melalui nilai-nilai statistik yang dihasilkan. Dengan demikian, pola dan distribusi balita stunting dapat diidentifikasi secara lebih terstruktur dan menyeluruh. Penelitian ini menganalisis data historis balita stunting dari bulan September hingga Desember tahun 2024.

2.2 Metode Pengumpulan Data

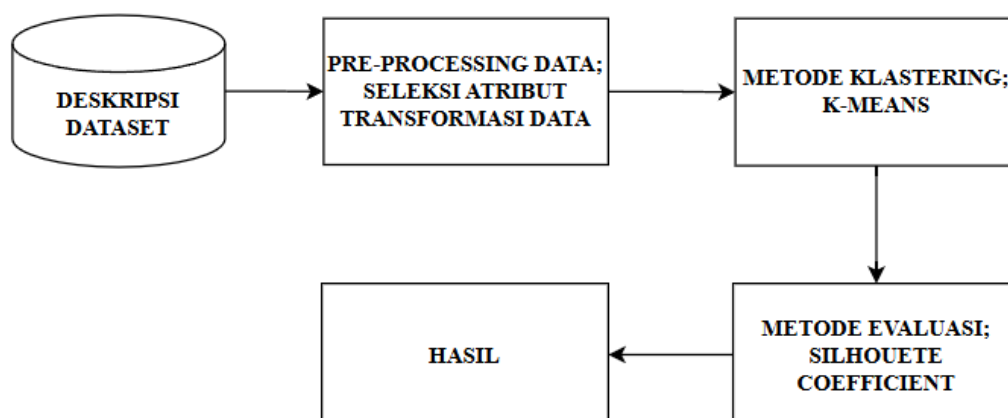
Penelitian dilakukan dengan observasi langsung pada instansi kesehatan Puskesmas Geyer 1 dan Puskesmas Geyer 2. Penelitian ini dimulai dari bulan maret-mei 2025. Data yang diperoleh kemudian disusun dalam format .xlsx untuk keperluan analisis.

2.3 Perangkat Penelitian

Penelitian ini dilakukan dengan bantuan perangkat lunak data mining Altair AI Studio versi 2025.1.1. Proses ini memanfaatkan kapabilitas *workflow* visual yang tersedia dalam *platform* tersebut untuk memuat data, melakukan pra-pemrosesan data, mengeksekusi algoritma K-Means, dan mengevaluasi hasil klasterisasi secara efisien.

2.4 Alur Penelitian

Penelitian ini dilakukan melalui beberapa tahapan yang digambarkan pada Gambar 1. Tahapan penelitian dimulai dengan pengumpulan data, diikuti oleh prapemrosesan dataset untuk memastikan data siap digunakan. Selanjutnya, dilakukan penerapan algoritma K-Means untuk proses klasterisasi, dan diakhiri dengan evaluasi hasil yang diperoleh untuk menilai kualitas dan ketepatan klaster yang terbentuk. Setiap tahap dirancang secara sistematis agar proses penelitian dapat berjalan secara konsisten dan hasil yang diperoleh dapat dipertanggungjawabkan.



Gambar 1 Alur Penelitian

4.1 Deskripsi Dataset

Penelitian ini menggunakan dataset balita stunting yang diambil dari data arsip instansi kesehatan puskesmas geyer 1 dan 2 pada bulan September 2024. Dataset terdiri dari 1470 records dengan 5 atribut. Atribut diantaranya adalah Nama, Alamat, Usia, BB, dan TB.

Tabel 1 Dataset Balita Stunting

| NO | NAMA | ALAMAT | USIA | BB | TB |
|-------|-------|------------|-------|-------|-------|
| 1 | RF | JUWORO | 54 | 14.7 | 97.2 |
| 2 | HNW | JUWORO | 46 | 11.9 | 93.5 |
| 3 | NPA | JUWORO | 28 | 8.1 | 79.8 |
| 4 | DBA | SOBO | 32 | 9.5 | 85.6 |
| 5 | NDU | SOBO | 21 | 7.6 | 75.5 |
| 6 | RTW | SOBO | 33 | 9.5 | 84.5 |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| 1464 | SPK | ASEMRUDUNG | 28 | 10.4 | 77 |
| 1465 | TIK | ASEMRUDUNG | 29 | 9 | 79 |
| 1466 | WKS | ASEMRUDUNG | 44 | 12 | 89.8 |
| 1467 | WA | ASEMRUDUNG | 21 | 9 | 77 |
| 1468 | YA | ASEMRUDUNG | 35 | 12.6 | 86.5 |
| 1469 | ZAS | ASEMRUDUNG | 48 | 12.5 | 94 |
| 1470 | ZMA | ASEMRUDUNG | 9 | 8 | 67 |

Pada penelitian ini terdapat 3 proses berdasarkan kombinasi atribut yang digunakan. Untuk mempermudah analisis dan memperkuat pembahasan dapat dilihat pada beberapa penjelasan dibawah ini:

- Proses Pertama: Klasterisasi berdasarkan atribut usia.
- Proses Kedua: Klasterisasi berdasarkan atribut usia, berat badan dan tinggi badan.
- Proses Ketiga: Analisis distribusi wilayah berdasarkan hasil klasterisasi dari proses pertama.

2.4.2 Pre-Processing Data

Dataset yang sudah dikumpulkan, kemudian akan melalui tahap pra-pemrosesan data sebelum metode K-Means klastering diterapkan. Data yang tidak sesuai dengan kebutuhan akan dieliminasi dan format data di sesuaikan agar konsisten dalam seluruh proses pengolahan.

Tabel 2 Operator dan Parameter Pre-processing

| Proses | Operator | Tujuan (Keterangan) | Parameter Kunci dan Nilai |
|----------------|----------------------------|--|--|
| Proses Pertama | <i>Select Attributes</i> | Memilih atribut USIA | <i>Attribute filter type: subset</i> |
| Proses Kedua | <i>Select Attributes</i> | Memilih atribut USIA, BB, dan TB | <i>Attribute filter type: subset</i> |
| | <i>Generate Attributes</i> | Memastikan BB dan TB bertipe data <i>real number</i> | BB : <i>round</i> (BB) TB : <i>round</i> (TB) |

- a. Import Data: membaca data dari lembar kerja (*sheet*) yang sesuai dan memastikan bahwa baris pertama digunakan sebagai nama atribut (*use first row as attribute names*).
- b. Seleksi Atribut: mempertimbangkan atribut yang akan dipilih yaitu usia, bb dan tb.
- c. Transformasi data: mengubah data yang memiliki nilai kategorikal menjadi bentuk numerik (BB, TB harus dikenali sebagai tipe data real number).

Perlu dicatat bahwa Proses Ketiga, yaitu analisis distribusi wilayah, tidak memerlukan langkah *pre-processing* data yang kompleks. Hal ini disebabkan karena Proses Ketiga merupakan tahap lanjutan yang hanya bertugas membaca file hasil klasterisasi dari Proses Pertama yang sudah mengandung label klaster, kemudian menggabungkannya dengan data alamat balita stunting. Oleh karena itu, *workflow* untuk Proses Ketiga di Altair AI Studio dimulai langsung dari operator *Read.Excel* untuk memuat data hasil dan data lokasi

2.4.3 Metode Klustering

Selanjutnya, dataset yang telah di transformasikan dalam bentuk numerik kemudian di proses untuk pengelompokkan data menggunakan algoritma K-Means Klustering.

Pada tahap ini, penerapan metode ini dilakukan menggunakan aplikasi RapidMiner. Algoritma K-Means digunakan untuk mengelompokkan data berdasarkan kemiripan atribut [14], [15].

Tahapan utama K-Means [16].

- a. Menentukan jumlah klaster
- b. Menentukan centroid
- c. Menghitung jarak titik data objek dengan centroid

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

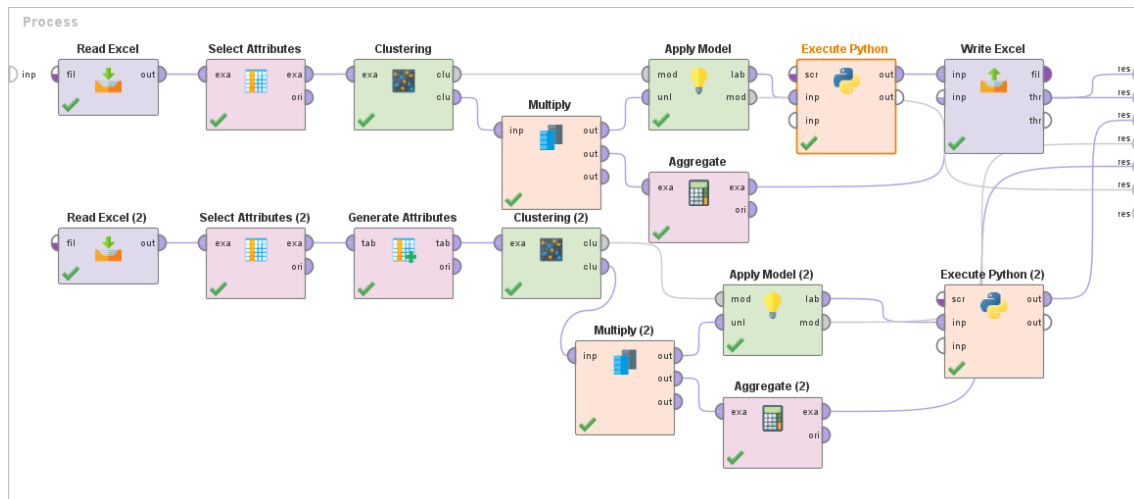
- d. Mengelompokkan data objek untuk menentukan anggota klaster berdasarkan jarak minimum
- e. Jika data untuk setiap klaster belum berhenti berubah, ulangi langkah 2-5, hingga terjadi perubahan pada anggota setiap klaster.

2.4.4 Evaluasi

Tahap Evaluasi dilakukan untuk menilai kualitas dan tingkat akurasi hasil klasterisasi yang dihasilkan oleh algoritma K-Means. Metode yang digunakan adalah *Silhouette Coefficient*, yang berfungsi mengukur seberapa baik pemisahan objek antar klaster (inter-cluster distance) dibandingkan dengan kerapatan objek di dalam klaster (intra-cluster distance). Nilai skor *Silhouette Coefficient* berkisar antara -1 hingga +1. Nilai yang mendekati +1 menunjukkan struktur klaster yang optimal dengan pemisahan yang jelas, sementara nilai mendekati 0 mengindikasikan struktur klaster yang lemah atau tumpang tindih [17].

3. HASIL DAN PEMBAHASAN

Implementasi teknis dari tahapan klasterisasi K-Means yang dijelaskan pada Bab Metodologi dilakukan menggunakan perangkat lunak Altair AI Studio versi 2025.1.1. Seluruh alur kerja yang mencakup Proses Pertama dan Proses Kedua disajikan secara visual pada gambar dibawah ini.



Gambar 2 Proses Klasterisasi K-Means

Tahapan awal implementasi Klasterisasi K-Means dimulai dengan pemuatan data dari berkas Excel menggunakan *operator* Read Excel. Data yang berhasil dimuat selanjutnya menjalani seleksi atribut untuk membatasi variabel yang relevan dalam proses klasterisasi. Data yang telah terseleksi kemudian dieksekusi menggunakan algoritma K-Means untuk membentuk klaster berdasarkan karakteristik yang dimiliki. Setelah klasterisasi selesai, evaluasi kualitas klaster dilakukan secara eksternal menggunakan *Silhouette Coefficient*. Sebagai tahapan *output* akhir, hasil klasterisasi yang telah dilengkapi dengan label kelompok disimpan kembali ke dalam berkas Excel untuk dianalisis lebih lanjut pada tahap distribusi wilayah.

Iterasi dilakukan pada beberapa nilai k (2–10) untuk memperoleh jumlah klaster optimal. Evaluasi menggunakan *Silhouette Coefficient* menunjukkan bahwa nilai $k = 2$ memberikan hasil terbaik. Oleh karena itu, jumlah klaster yang digunakan dalam penelitian ini adalah $k = 2$.

3.1 Hasil Proses Pertama (USIA)

Tabel 3 Hasil Klasterisasi Usia

| Cluster | Jumlah | Rata-rata usia | Karakteristik |
|---------|--------|----------------|--------------------|
| 0 | 667 | 19 | Balita Muda |
| 1 | 803 | 42 | Balita Pra-sekolah |

Klaster pertama yang berisi balita dengan rata-rata usia lebih muda menunjukkan bahwa kelompok usia tersebut lebih rentan terhadap kondisi stunting. Hal ini sejalan dengan pendapat [yulistianingsih,2022] bahwa periode usia 0–24 bulan merupakan fase emas pertumbuhan yang sangat dipengaruhi oleh asupan gizi.

3.1.1 Hasil Evaluasi Klaster

a. Evaluasi Global (Rata-rata)

Evaluasi kualitas klaster secara global, menggunakan rata-rata nilai *Silhouette Coefficient* dari seluruh data, menghasilkan skor 0.565. Angka ini dikategorikan sebagai struktur klaster sedang (*medium structure*), menunjukkan klaster yang terbentuk cenderung cukup baik dalam hal pemisahan dan kekompakan.

b. Evaluasi Per-klaster

Tabel 4 Hasil evaluasi proses pertama

| Cluster | Hasil Nilai Evaluasi | Keterangan |
|-----------|----------------------|-------------------------|
| Cluster_0 | 0.566 | <i>Medium Structure</i> |
| Cluster_1 | 0.546 | <i>Medium Structure</i> |

Nilai ini mengindikasikan bahwa pembentukan kluster cukup baik, meskipun pemisahan antar kluster tidak terlalu tajam.

Hasil Proses Kedua (USIA, BB, dan TB)

Tabel 5 Hasil klasterisasi Usia, BB, dan TB

| Cluster | Jumlah balita | Rata-rata Usia | Rata-rata BB (kg) | Rata-rata TB (cm) |
|---------|---------------|----------------|-------------------|-------------------|
| 0 | 640 | 19 | 8.3 | 73 |
| 1 | 830 | 42 | 11 | 89 |

Berdasarkan Tabel diatas, hasil klasterisasi membentuk dua kelompok balita. Cluster 0 merepresentasikan balita dengan usia rata-rata lebih muda (19 bulan) serta kondisi fisik relatif lebih rendah, sehingga kelompok ini lebih rentan terhadap gangguan pertumbuhan. Sementara itu, Cluster 1 berisi balita dengan rata-rata usia lebih tua (42 bulan) dan indikator pertumbuhan yang lebih baik dibandingkan dengan Cluster 0.

3.2.1 Evaluasi Hasil Kluster

a. Evaluasi Global (Rata-rata)

Hasil evaluasi global menunjukkan skor sebesar 0.535. Nilai ini dikategorikan sebagai struktur kluster sedang (*medium structure*), mengindikasikan bahwa penambahan atribut Berat Badan (BB) dan Tinggi Badan (TB) bersama Usia tetap menghasilkan kluster yang cukup kohesif dan terpisah.

b. Evaluasi Per-kluster

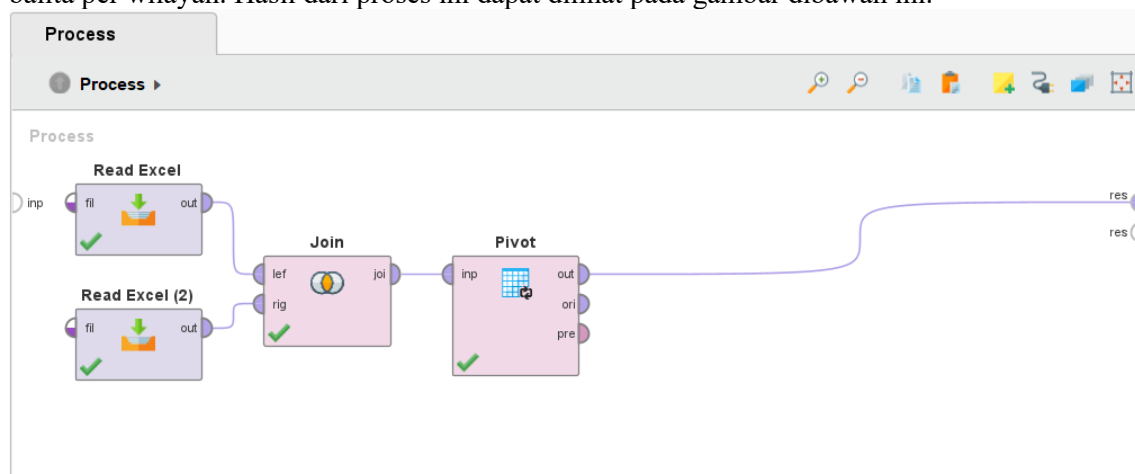
Tabel 6 Hasil evaluasi proses kedua

| Cluster | Nilai Hasil Evaluasi | Keterangan |
|-----------|----------------------|-------------------------|
| Cluster 0 | 0.496 | <i>Weak Structure</i> |
| Cluster 1 | 0.564 | <i>Medium Structure</i> |

Nilai-nilai ini menunjukkan bahwa kluster yang terbentuk mampu mengelompokkan data balita stunting dengan cukup konsisten dan tidak terdapat kluster dengan nilai evaluasi ekstrim.

3.3 Proses Distribusi Wilayah

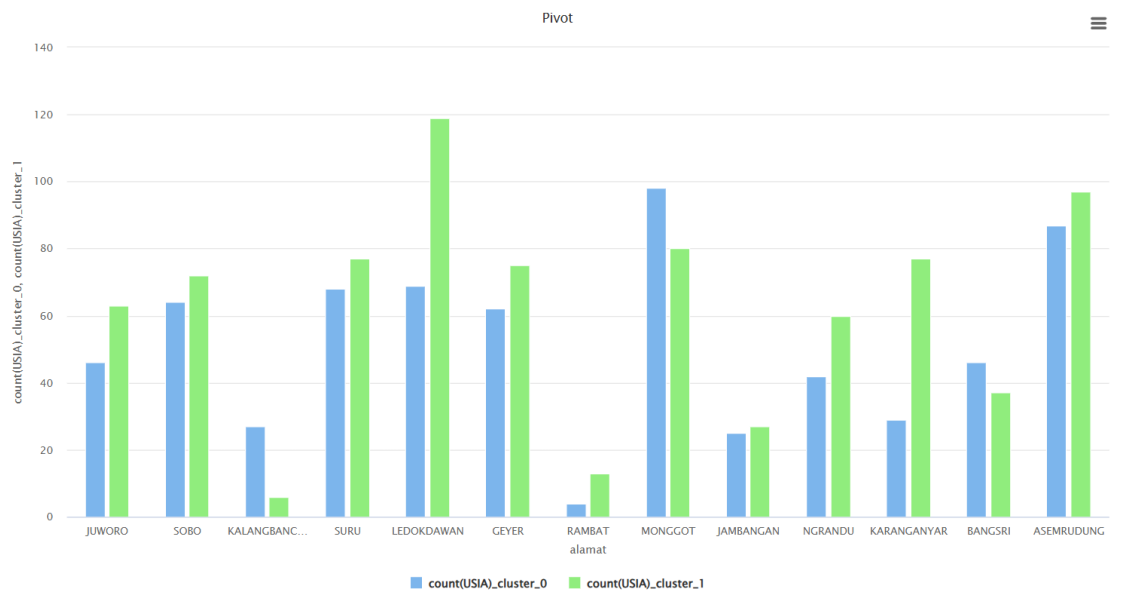
Pada proses ketiga, dilakukan pengolahan data dengan tujuan untuk memperoleh distribusi balita berdasarkan wilayah. Gambar di bawah memperlihatkan tahapan proses yang diawali dengan operator Read Excel yang digunakan untuk membaca dataset utama, kemudian digabungkan dengan data tambahan melalui operator Join. Selanjutnya, operator Pivot diterapkan untuk merubah bentuk data sehingga memudahkan dalam melihat distribusi jumlah balita per wilayah. Hasil dari proses ini dapat dilihat pada gambar dibawah ini.



Gambar 3 Proses distribusi wilayah

Hasil distribusi wilayah menunjukkan bahwa:

- Cluster_0 merepresentasikan kelompok balita usia muda yang paling banyak ditemukan di Desa Monggot dengan jumlah 98 balita.
- Cluster_1 didominasi oleh balita usia lanjut yang tercatat cukup tinggi di Desa Ledokdawan sebanyak 119 balita.



Gambar 4 Grafik visualisasi distribusi wilayah

Visualisasi grafik pada Gambar diatas memperlihatkan bahwa Desa Ledokdawan memiliki total 188 balita yang sebagian besar masuk dalam Cluster_1, Desa Asemrudung sebanyak 184 balita dengan dominasi Cluster_1, dan Desa Monggot sebanyak 178 balita dengan dominasi Cluster_0. Ketiga desa tersebut tercatat sebagai wilayah dengan jumlah balita stunting tertinggi dalam dataset penelitian ini.

Sebaliknya, wilayah dengan jumlah balita paling sedikit terdapat di Desa Rambat dengan 17 balita (dominan pada Cluster_1) dan Desa Kalangbancar dengan 33 balita (dominan pada Cluster_0).

Dengan pemetaan distribusi ini, pihak terkait dapat menentukan prioritas intervensi wilayah berdasarkan komposisi usia balita yang mendominasi. Wilayah dengan jumlah balita stunting yang tinggi serta dominasi klaster tertentu dapat menjadi fokus kajian lebih lanjut dalam penyusunan program perbaikan gizi dan kesehatan, sehingga intervensi yang diberikan lebih tepat sasaran sesuai dengan karakteristik kelompok usia balita di wilayah tersebut.

4. KESIMPULAN

Penelitian ini bertujuan untuk melakukan klasterisasi balita *stunting* di Kecamatan Geyer menggunakan Algoritma K-Means dengan tiga skenario proses, serta memvalidasi kualitas klaster menggunakan *Silhouette Coefficient*. Implementasi seluruh tahapan data mining dilakukan dengan bantuan perangkat lunak Altair AI Studio versi 2025.1.1.

Dari hasil implementasi klasterisasi, didapatkan temuan sebagai berikut:

- Proses Pertama, yang hanya menggunakan atribut Usia, berhasil membentuk dua klaster yang berkarakteristik berbeda (Usia Muda dan Usia Tua), dengan skor *Silhouette Coefficient* rata-rata sebesar 0.565, menunjukkan struktur klaster yang sedang (*medium structure*).
- Proses Kedua, yang melibatkan atribut Usia, Berat Badan (BB), dan Tinggi Badan (TB), menghasilkan dua klaster dengan skor global sedikit lebih rendah yaitu 0.535.

Meskipun demikian, analisis *per-klastr* menunjukkan variasi signifikan, di mana Cluster_1 memiliki struktur sedang (0.564), sementara Cluster_0 dikategorikan memiliki struktur lemah (0.496). Variasi ini mengindikasikan adanya tantangan pemisahan data setelah penambahan atribut multivariabel.

- c. Analisis distribusi wilayah menunjukkan bahwa Desa Ledokdawan (188 balita) dan Desa Asemrudung (184 balita) tercatat sebagai wilayah dengan kasus *stunting* tertinggi, dan keduanya didominasi oleh Cluster_1. Desa Monggot (178 balita) paling banyak ditemukan pada Cluster_0. Temuan ini memberikan dasar bagi pihak terkait untuk menetapkan prioritas intervensi wilayah dengan mempertimbangkan volume kasus dan komposisi usia balita yang mendominasi Cluster_0 atau Cluster_1. Dengan demikian, program intervensi gizi dan kesehatan dapat dirancang agar lebih tepat sasaran dan disesuaikan dengan kebutuhan spesifik kelompok usia balita di masing-masing desa.

5. SARAN

Saran untuk penelitian mendatang ditekankan pada pengembangan metode, data, dan pemanfaatan hasil. Di sisi metodologi, disarankan untuk mengeksplorasi algoritma klusterisasi lain, seperti DBSCAN atau Fuzzy C-Means, guna mengatasi keterbatasan K-Means dalam penentuan klaster dan menghasilkan kualitas yang lebih optimal. Pengujian validitas juga perlu diperkaya dengan metode pembandingan seperti Davies Bouldin Index (DBI). Dari aspek data, penelitian selanjutnya harus menambahkan atribut-atribut pendukung yang komprehensif (seperti faktor sosial-ekonomi atau kesehatan orang tua) serta memastikan data lebih lengkap dan terdistribusi merata untuk meningkatkan validitas analisis. Terakhir, secara implementasi praktis, hasil distribusi klaster ini harus segera dimanfaatkan oleh instansi kesehatan untuk merancang program intervensi yang lebih tepat sasaran, menyesuaikan strategi gizi dan kesehatan berdasarkan dominasi kelompok usia balita di setiap wilayah.

DAFTAR PUSTAKA

- [1] DHIKA MALITA PUSPITA ARUM, KARTIKA IMAM SANTOSO, ANDRI TRIYONO, EKO SUPRIYADI, AGUS SUSILO NUGROHO, and E. Widodo, "ALGORITMA RANDOM FOREST, DECISION TREE, DAN XGBOOST UNTUK KLASIFIKASI STUNTING PADA BALITA," *Jurnal Transformatika*, vol. 23, no. 1, pp. 67–76, Jul. 2025, doi: 10.26623/transformatika.v23i1.12202.
- [2] D. M. Anjani, Nurhayati Sri, and immawarti, "PENERAPAN PENDIDIKAN KESEHATAN TERHADAP PENGETAHUAN IBU TENTANG STUNTING PADA BALITA DI WILAYAH KERJA UPTD PUSKESMAS RAWAT INAP BANJARSARI METRO UTARA," 2024.
- [3] D. D. Astuti, T. W. Handayani, and D. P. Astuti, "Cigarette smoke exposure and increased risks of stunting among under-five children," *Clin Epidemiol Glob Health*, vol. 8, no. 3, pp. 943–948, Sep. 2020, doi: 10.1016/j.cegh.2020.02.029.
- [4] R. Anugrah Kahar and N. Anugra, "PENGUNAAN LABORATORIUM VIRTUAL DALAM MENINGKATKAN HASIL BELAJAR IPA UJI ZAT MAKANAN (KARBOHIDRAT)," 2023. [Online]. Available: www.olabs.edu.in.
- [5] I. Risnawati and D. Munafiah, "UPAYA PENCEGAHAN STUNTING MELALUI PEMBERIAN MAKANAN SEIMBANG DAN STUNTING MASSAGE," 2022.
- [6] S. P. Gunawan and M. Adriani, "Attribution-NonCommercial-ShareAlike license (CC BY-NC-SA 4.0). OBESITAS DAN TINGKAT STRES BERHUBUNGAN DENGAN HIPERTENSI PADA ORANG DEWASA DI KELURAHAN KLAMPIS NGASEM,

- SURABAYA,” *Media Gizi Indonesia*. 2020, vol. 15, no. 2, pp. 119–126, 2020, doi: 10.204736/mgi.v15i2.119-126.
- [7] Kementerian Kesehatan Republik Indonesia, “SSGI 2024: Prevalensi Stunting Nasional.” Accessed: Sep. 12, 2025. [Online]. Available: <https://www.badankebijakan.kemkes.go.id/en/ssgi-2024-prevalensi-stunting-nasional-turun-menjadi-198/>
- [8] M. Syahril, K. Erwansyah, and M. Yetri, “J-SISKO TECH Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD Penerapan Data Mining Untuk Menentukan Pola,” ■, vol. 118, no. 1, pp. 118–136, 2020.
- [9] M. I. Fianty, M. E. Johan, A. Aulia, and M. M. Veronica, “Application of Clustering-Based Data Mining for the Assessment of Nutritional Status in Toddlers at Community Health Centers,” *Journal of Information Systems and Informatics*, vol. 5, no. 4, pp. 1350–1362, Dec. 2023, doi: 10.51519/journalisi.v5i4.586.
- [10] N. Hendrastuty, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa,” *Jurnal Ilmiah Informatika dan Ilmu Komputer (JIMA-ILKOM)*, vol. 3, no. 1, pp. 46–56, Mar. 2024, doi: 10.58602/jima-ilkom.v3i1.26.
- [11] P. Berkhin, “Survey of Clustering Data Mining Techniques,” 2022.
- [12] I. D. Sintawati, W. Widiarina, and K. Mariskhana, “Analysis of Malnutrition Status in Toddlers Using the K-MEANS Algorithm Case Study in DKI Jakarta Province,” *sinkron*, vol. 8, no. 4, pp. 2318–2324, Oct. 2024, doi: 10.33395/sinkron.v8i4.14087.
- [13] A. Cicchella, C. Stefanelli, and M. Massaro, “Upper respiratory tract infections in sport and the immune system response. A review,” May 01, 2021, *MDPI AG*. doi: 10.3390/biology10050362.
- [14] A. Yudhistira and R. Andika, “Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering,” *Journal of Artificial Intelligence and Technology Information (JAITI)*, vol. 1, no. 1, pp. 20–28, Feb. 2023, doi: 10.58602/jaiti.v1i1.22.
- [15] N. Buslim, R. Pradono Iswara, and F. Agustian, “THE MODELING OF ‘MUSTAHIQ’ DATA USING K-MEANS CLUSTERING ALGORITHM AND BIG DATA ANALYSIS (CASE STUDY: LAZ),” *Jurnal Teknik Informatika*, vol. 13, no. 2, 2020.
- [16] A. Rohmah *et al.*, “IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING ANALYSIS UNTUK MENENTUKAN HAMBATAN PEMBELAJARAN DARING (STUDI KASUS: SMK YASPIM GEGERBITUNG),” 2021. [Online]. Available: <https://www.alfasoleh.com/2019/11/k-means-clustering-contoh>
- [17] L. E. Ekemeyong Awong and T. Zielinska, “Comparative Analysis of the Clustering Quality in Self-Organizing Maps for Human Posture Classification,” *Sensors*, vol. 23, no. 18, Sep. 2023, doi: 10.3390/s23187925.