PENERAPAN *ALGORITMA K-MEANS CLUSTERING* UNTUK ANALISIS FAKTOR PENYEBAB *STUNTING*

Khonsa Nida Mujahidah, April Lia Hananto, Bayu Priyatna, Shofa Shofia Hilabi
Sistem Informasi, Universitas Buana Perjuangan Karawang
Jalan HS. Ronggo Waluyo, Telukjambe Timur, Puseurjaya, Telukjambe Timur
Kabupaten Karawang, Jawa Barat 41361, Indonesia
khonsamujahidah@mhs.ubpkarawang.ac.id

ABSTRAK

Stunting masih menjadi salah satu permasalahan utama dalam kesehatan masyarakat di Indonesia karena berdampak pada pertumbuhan fisik anak, perkembangan kognitif, serta kualitas sumber daya manusia di masa depan. Kondisi ini umumnya disebabkan oleh kekurangan gizi kronis yang terjadi dalam jangka panjang dan dipengaruhi oleh berbagai faktor, seperti sanitasi lingkungan, akses layanan kesehatan, serta kondisi sosial ekonomi. Meskipun berbagai upaya telah dilakukan, analisis faktor penyebab *stunting*, khususnya di tingkat desa, masih belum dilakukan secara terstruktur dan berbasis data. Hal ini menyebabkan intervensi yang diberikan sering kali kurang tepat sasaran. Penelitian ini bertujuan untuk mengelompokkan wilayah desa berdasarkan faktor risiko stunting guna mendukung perumusan kebijakan yang lebih efektif. Metode yang digunakan adalah *Knowledge Discovery in Databases* (KDD) dengan algoritma *K-Means Clustering* yang memanfaatkan tujuh variabel indikator. Proses analisis dilakukan melalui tahapan preprocessing, transformasi data, serta proses clustering. Hasil penelitian menunjukkan bahwa jumlah *cluster* optimal adalah tiga *cluster* dengan nilai *silhouette* sebesar 0,36, yang terdiri dari *cluster* risiko tinggi, sedang, dan rendah. Temuan ini diharapkan dapat menjadi dasar dalam pengambilan keputusan dan perencanaan program penanganan stunting yang lebih terarah dan berbasis data.

Kata kunci : Stunting, Data Mining, K-Means Clustering, KDD *(Knowledge Discovery in Database)*

1. PENDAHULUAN

Stunting masih menjadi salah satu permasalahan utama dalam bidang kesehatan masyarakat di Indonesia karena berdampak pada pertumbuhan fisik, perkembangan kognitif anak, serta kualitas sumber daya manusia di masa depan. Kondisi ini umumnya disebabkan oleh malnutrisi kronis yang berlangsung dalam jangka panjang, serta dipengaruhi oleh berbagai faktor seperti sanitasi lingkungan, akses layanan kesehatan, dan kondisi sosial ekonomi masyarakat. Oleh karena itu, penanganan stunting menjadi prioritas penting dalam upaya peningkatan kualitas kesehatan nasional [1].

Meskipun berbagai program intervensi telah dilakukan, distribusi kasus stunting di Indonesia masih menunjukkan variasi yang cukup signifikan antar wilayah. Pada wilayah penelitian, prevalensi stunting pada tahun 2025 tercatat sebesar 17,4%, yang menunjukkan bahwa permasalahan ini masih memerlukan perhatian serius. Tingginya angka tersebut mengindikasikan bahwa faktor-faktor penyebab stunting belum teridentifikasi secara optimal, khususnya pada tingkat wilayah yang lebih spesifik seperti desa. Kondisi ini menyebabkan intervensi yang dilakukan sering kali belum tepat sasaran karena belum didasarkan pada pemetaan risiko yang jelas [2].

Sejumlah penelitian sebelumnya telah memanfaatkan metode data mining, khususnya algoritma K-Means Clustering, untuk mengelompokkan data stunting berdasarkan karakteristik tertentu. Penelitian oleh Cahya Syalom menunjukkan bahwa metode K-Means mampu mengelompokkan wilayah berdasarkan tingkat stunting secara efektif [3]. Selain itu, Fadillah Luthfiyah juga berhasil mengidentifikasi faktor pemicu stunting melalui pendekatan clustering pada tingkat kabupaten/kota [4]. Penelitian lain oleh Yulia Trisanti Dendo menunjukkan bahwa teknik clustering dapat membantu dalam pemetaan status gizi anak secara lebih sistematis [5]. Namun, sebagian besar penelitian tersebut masih berfokus pada tingkat wilayah yang lebih luas dan belum secara spesifik mengkaji pengelompokan faktor risiko stunting pada tingkat desa dengan variabel yang lebih komprehensif.

Berdasarkan permasalahan tersebut, diperlukan suatu pendekatan berbasis data yang mampu mengelompokkan wilayah secara objektif berdasarkan kemiripan karakteristik faktor risiko stunting. Pendekatan data mining melalui algoritma K-Means Clustering dipilih karena mampu mengelompokkan data ke dalam beberapa cluster berdasarkan tingkat kemiripan antar variabel, sehingga dapat mengungkap pola tersembunyi dalam data[6]. Dengan memanfaatkan pendekatan ini, diharapkan dapat diperoleh gambaran yang lebih jelas mengenai karakteristik wilayah berdasarkan tingkat risiko stunting.

Penelitian ini bertujuan untuk mengelompokkan wilayah administratif tingkat desa berdasarkan faktor-faktor penyebab stunting menggunakan algoritma K-Means Clustering. Hasil dari penelitian ini diharapkan dapat memberikan informasi yang lebih terstruktur mengenai pola distribusi risiko stunting, sehingga dapat menjadi dasar dalam penyusunan kebijakan dan

strategi intervensi yang lebih tepat sasaran dan berbasis data.

2. TINJAUAN PUSTAKA

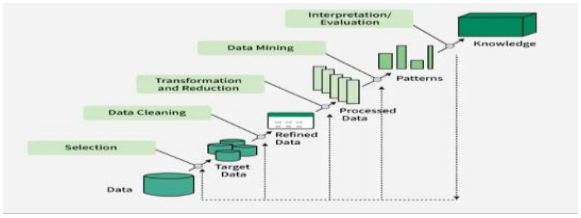
2.1. *Stunting*

Gangguan pertumbuhan pada anak-anak usia dini yang disebabkan oleh malnutrisi kronis dan jangka panjang disebut sebagai stunting. Produktivitas, perkembangan fisik, dan kemajuan kognitif generasi mendatang dapat terganggu akibat hal ini. Gizi ibu selama kehamilan dan menyusui, ketersediaan air bersih, kualitas sanitasi, pendidikan ibu, dan status ekonomi keluarga merupakan penyebab utama stunting di Indonesia [7].

Malnutrisi jangka panjang, infeksi, kesehatan ibu selama kehamilan, persalinan, dan periode pasca persalinan, wanita dengan postur tubuh pendek, pemberian makanan pendamping sebelum usia enam bulan, dan tidak menyusui secara eksklusif merupakan penyebab langsung pertumbuhan terhambat pada balita [8]. Posisi ekonomi yang rendah, yang mempengaruhi keamanan pangan keluarga, pengaruh sosial terhadap gaya hidup komunitas, budaya, gaya pengasuhan, kebiasaan makan, kesehatan keluarga, dan akses ke layanan kesehatan merupakan contoh penyebab tidak langsung [9].

2.2. *Knowledge Discovery in Database (KDD)*

Penemuan Pengetahuan dalam Basis Data (KDD) adalah metode terorganisir untuk menarik kesimpulan yang bermakna dari kumpulan data yang besar. Menurut Fayyad, KDD melibatkan serangkaian langkah yang saling terkait, termasuk pemilihan data, pra-pemrosesan, transformasi, penambangan data, dan evaluasi hasil. Tujuan pendekatan ini adalah mengubah data mentah menjadi pengetahuan yang berguna yang memudahkan pengambilan keputusan [10].



Gambar 1. Tahapan *knowladge discovery in database (kdd)*
Sumber *: Google*

Dalam studi ini, pendekatan KDD dipilih karena mampu menangani data kompleks dan multidimensi yang sering dijumpai dalam bidang kesehatan masyarakat, khususnya data faktor penyebab *stunting*. Proses KDD membantu peneliti menyiapkan data dengan benar sebelum dilakukan analisis menggunakan algoritma seperti *K-Means Clustering* [11].

2.3. *K-Means Clustering*

Metode pembelajaran tanpa pengawasan yang disebut K-Means Clustering digunakan untuk mengelompokkan data berdasarkan kemiripan. Setiap titik data ditugaskan ke kluster dengan pusat kluster terdekat setelah teknik ini menghitung jumlah kluster, k. K-Means sering digunakan dalam bidang seperti pemasaran, analisis perilaku, segmentasi pengguna, dan pengelompokan kinerja akademik karena kemudahan penggunaannya dan efektivitasnya[12].

Biasanya, k titik pusat awal dipilih secara acak untuk memulai proses. Jarak antara setiap titik data dan titik pusat kemudian ditentukan, biasanya menggunakan jarak Euclidean. Titik pusat kemudian dihitung ulang sebagai rata-rata dari semua titik dalam setiap kluster setelah titik data dialokasikan ke titik pusat terdekat. Proses iteratif ini diulang hingga titik pusat stabil atau perubahan menjadi tidak signifikan. Rumus Euclidean digunakan untuk menghitung jarak, dan prosedur ini diulang hingga pusat kluster menunjukkan kluster yang stabil dan tidak lagi berubah secara signifikan [13], [14].

Teknik menghitung jarak antara titik data dan pusat kluster menggunakan rumus Euclidean adalah sebagai berikut:

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

Sumber : Hastie, T., Tibshirani, R., & Friedman, J.

2.4. *Elbow Method*

Metode siku membantu menentukan jumlah klaster optimal (k) dengan meminimalkan jumlah kuadrat dalam klaster (withinss) sebanyak mungkin. Nilai k yang optimal diperoleh melalui pengamatan grafik, yaitu pada titik di mana kurva mulai membentuk lekukan tajam menyerupai “siku”. Seiring bertambahnya jumlah cluster, nilai *Sum of Squared Errors* (SSE) cenderung mengalami penurunan yang cukup signifikan[15].

2.5. *Silhoutte Method*

Sebuah teknik untuk mengevaluasi kualitas hasil pengelompokan data adalah Koefisien Siluet. Kohesi dan pemisahan adalah dua metrik utama yang digunakan dalam metode ini. Derajat kedekatan antara titik data dalam kluster yang sama ditunjukkan oleh kohesi., sedangkan separasi menggambarkan jarak atau perbedaan antar kelompok yang berbeda. Nilai kohesi dan separasi yang tinggi mencerminkan struktur cluster yang semakin jelas dan berkualitas[16].

2.6. Penelitian Terdahulu

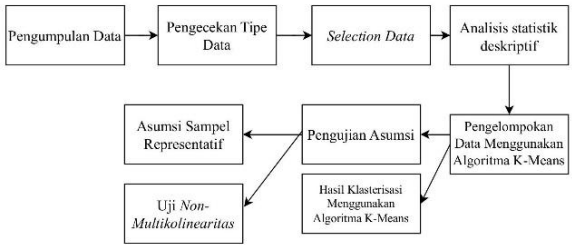
Penelitian terdahulu digunakan sebagai referensi dan bahan perbandingan dalam penelitian ini untuk mengetahui metode yang digunakan, hasil penelitian yang diperoleh, serta keterbatasan dari penelitian sebelumnya yang berkaitan dengan topik yang diteliti disajikan pada tabel 1.

Tabel 1. Penelitian terdahulu

No	Peneliti	Fokus Peneliti	Temuan Utama	Gap
1	[17]	Menganalisis pengelompokan kota berdasarkan faktor penyebab stunting menggunakan algoritma K-Means.	Hasil penelitian menunjukkan bahwa jumlah cluster optimal adalah 3 cluster. Cluster tersebut menggambarkan karakteristik wilayah dengan kondisi faktor penyebab stunting yang berbeda	Penelitian hanya menggunakan beberapa variabel penyebab stunting sehingga kemungkinan masih terdapat faktor lain yang mempengaruhi
2	[2]	Mengidentifikasi wilayah dengan risiko Kekurangan Energi Kronis (KEK) pada ibu hamil di	Hasil analisis menunjukkan bahwa jumlah cluster optimal adalah 2 cluster berdasarkan metode Elbow	Penelitian hanya berfokus pada pengelompokan wilayah KEK pada ibu hamil.
3	[18]	Menganalisis cluster wilayah kecamatan berdasarkan faktor penyebab stunting	Hasil penelitian menunjukkan terbentuk 3 cluster wilayah dengan karakteristik penyebab stunting yang berbeda.	Penelitian menggunakan metode hierarchical clustering
4	[19]	Mengelompokkan wilayah kota di Provinsi Jawa Barat berdasarkan faktor pemicu stunting menggunakan algoritma K-Means	Hasil penelitian menunjukkan bahwa metode K-Means dapat mengelompokkan data wilayah berdasarkan faktor pemicu stunting	Penelitian lebih berfokus pada wilayah Jawa Barat.
5	[20]	Mengelompokan kepadatan penduduk di Provinsi DKI Jakarta menggunakan algoritma K-Means Clustering	Metode K-Means mampu mengelompokkan wilayah berdasarkan tingkat kepadatan penduduk sehingga mempermudah analisis distribusi	Penelitian berfokus pada data kependudukan dan belum mengkaji faktor kesehatan lainnya.

3. METODE PENELITIAN

Untuk menemukan pola data dalam bentuk kluster, penelitian ini menggunakan alat Google Colab dan *Algoritma K-Means Clustering*. Klaster-klaster ini kemudian akan disajikan secara visual melalui diagram. Langkah-langkah berikut menguraikan proses untuk memperoleh informasi yang diperlukan.



Gambar 2. Langkah-Langkah Penelitian

3.1. Pengumpulan Data

Penelitian ini dimulai dengan pengumpulan data sekunder, termasuk variabel-variabel yang didefinisikan dalam studi. Data tersebut diambil dari sumber-sumber yang relevan dan dipilih secara cermat sesuai kebutuhan analisis. Proses ini menjamin bahwa informasi yang diperoleh mencerminkan kondisi yang sedang diteliti secara akurat dan mendukung tahap pengolahan data selanjutnya.

3.2. Pengecekan Tipe Data

Tahap selanjutnya berupa tahap Pengecekan Tipe Data, yaitu untuk memastikan bahwa setiap variabel yang digunakan telah memenuhi format yang sesuai untuk analisis. Proses ini dilakukan dengan mengidentifikasi jenis data pada setiap atribut, seperti data numerik maupun kategorikal, sehingga dapat dipastikan bahwa data tersebut dapat diolah dengan metode yang digunakan. Pemeriksaan tipe data juga bertujuan untuk menghindari kesalahan dalam proses pengolahan dan memastikan bahwa seluruh variabel dapat dianalisis secara tepat pada tahap berikutnya.

3.3. *Selection* Data

Langkah berikutnya adalah pemilihan data, yang melibatkan pemilihan data dan variabel yang relevan berdasarkan tujuan penelitian. Pada tahap ini, penyaringan diterapkan pada dataset dengan mengevaluasi pentingnya setiap variabel untuk analisis. Tujuannya adalah memilih data yang benar-benar membantu proses pengelompokan, sehingga analisis menjadi lebih terfokus dan menghasilkan pola pengelompokan yang lebih akurat.

3.4. Analisis Statistik Deskriptif

Analisis statistik deskriptif dilakukan untuk merangkum karakteristik data. Hal ini mencakup rata-rata perhitungan, nilai minimum dan maksimum, serta data sebaran. Analisis semacam ini membantu peneliti mengidentifikasi pola awal dan potensi pencilan, serta memberikan gambaran awal sebelum melanjutkan ke analisis yang lebih mendetail.

3.5. Pengelompokan Data Menggunakan *Algoritma K-Means*

Langkah berikutnya melibatkan melakukan pengelompokan data dengan algoritma *K-Means*. Metode ini, yang umum digunakan dalam penambangan data, mengelompokkan data ke dalam klaster berdasarkan kemiripan fitur-fiturnya. Proses pengelompokan dimulai dengan asumsi awal.

3.6. Pengujian Asumsi

Langkah pertama dalam pengelompokan data menggunakan clustering k-means adalah pengujian asumsi. Untuk memastikan data memenuhi persyaratan analisis, uji asumsi dilakukan. Uji ini meliputi uji non-multikolinearitas untuk memastikan tidak ada korelasi tinggi antara variabel menggunakan nilai FIV (Faktor Inflasi Varians) dan uji asumsi sampel representatif untuk memastikan sampel mewakili populasi menggunakan uji KMO (*Kaiser-Meyer-Olkin*).

3.7. Hasil Klasterisasi Menggunakan *K-Means*

Tahap akhir dari penelitian ini melibatkan hasil pengelompokan melalui algoritma K-Means. Di sini, data yang telah diproses dan diuji asumsi akan dikelompokkan ke dalam klaster berdasarkan kemiripan fitur-fiturnya. Setiap klaster kemudian akan dianalisis untuk mengidentifikasi pola dan mendefinisikan karakteristik dari masing-masing kelompok. Hasil pengelompokan ini menjadi dasar untuk memahami pola data dan dapat mendukung rekomendasi atau kesimpulan yang sesuai dengan tujuan penelitian.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Pengumpulan data merupakan tahap awal penelitian untuk memperoleh informasi yang diperlukan untuk analisis. Data yang digunakan dalam studi ini adalah data sekunder yang diperoleh dari Dinas Kesehatan. Data ini mencakup berbagai indikator yang menjadi faktor penyebab masalah kesehatan, seperti rata-rata kebersihan yang buruk, rata-rata kurangnya imunisasi, rata-rata fasilitas toilet, rata-rata infestasi cacing, rata-rata merokok, rata-rata riwayat malnutrisi ibu dan bayi (BKB), dan rata-rata penyakit.

4.2. Pengecekan Tipe Data

Pada tahap ini, data sekunder yang telah dikumpulkan oleh peneliti akan diperiksa terlebih dahulu untuk mengidentifikasi tipe data pada masing-masing variabel. Uraian mengenai karakteristik data tersebut disajikan secara lebih rinci pada Tabel 2 berikut.

Table 2. Deskripsi Data

Variabel	Tipe Data
Desa	Objek
X1(Air Tidak Bersih)	Float
X2(Kecacingan)	Float
X3(Jamban)	Float
X4(Tidak Imunisasi)	Float
X5(Merokok)	Float
X6(Riwayat Ibu KEK)	Float
X7(Penyakit)	Float

Berdasarkan tabel 2, variabel yang digunakan dalam penelitian terdiri dari satu variabel identitas wilayah dan beberapa variabel numerik yang merepresentasikan faktor-faktor yang berkaitan dengan risiko *stunting*. Variabel Desa memiliki tipe data objek yang berfungsi sebagai penanda atau identitas setiap wilayah pengamatan. Sementara itu, variabel X1 hingga X7, yang meliputi Air Tidak Bersih, Kecacingan, Jamban, Tidak Imunisasi, Merokok, Riwayat Ibu KEK, dan Penyakit, bertipe data *float* karena memuat nilai numerik berbentuk persentase atau rasio yang dapat memiliki nilai desimal. Struktur tipe data ini memungkinkan seluruh variabel numerik dapat diolah lebih. lanjut dalam proses analisis clustering menggunakan algoritma *K-Means.*

4.3. *Selection* Data

Table 3. Dataset Penyebab Stunting

No	Desa	X1	X2	X3	X4	X5	X6	X7
1	CIPTA	3	2	4	2	4	2	4
2	MEKAR	9	12	8	10	9	7	9
3	SARI	10	13	9	14	5	13	7
4	MEDAL	8	7	7	8	8	8	9
5	JATI	14	9	4	13	6	8	13
6	ASIH	9	2	2	10	15	1	7
8	…	…	…	…	…	…	…	…
9	…	…	…	…	…	. …	…	…
300	TELUK	9	2	1	9	5	6	3

Selama tahap verifikasi data ini, dilakukan pemilihan data untuk membuat sebuah dataset yang siap untuk dianalisis. Dataset yang digunakan dalam studi ini mencakup 300 *record* data yang merepresentasikan data pada tingkat desa. Pada Tabel 2 ditampilkan contoh dataset penyebab yang berkaitan dengan risiko *stunting* yang terdiri dari 7 variabel penelitian. Variabel X1 merepresentasikan persentase rumah tangga dengan akses air tidak bersih, X2 menunjukkan persentase kejadian kecacingan, X3 menggambarkan persentase jamban yang tidak layak, X4 merepresentasikan persentase balita yang tidak mendapatkan imunisasi, X5 menunjukkan persentase kebiasaan merokok dalam rumah tangga, X6 menggambarkan persentase riwayat ibu dengan kondisi KEK dan X7 menunjukkan persentase kejadian penyakit yang berkaitan dengan kondisi kesehatan balita. Dataset tersebut selanjutnya digunakan sebagai dasar dalam proses analisis *clustering* menggunakan algoritma *K-Means.*

4.4. Analisis statistik deskriptif

Analisis statistik deskriptif digunakan dalam penelitian ini untuk menentukan karakteristik data. Bagian ini menyajikan jumlah data, nilai minimum dan maksimum, rata-rata, serta simpangan baku untuk setiap variabel. Hasil terkait penyebab stunting pada balita ditampilkan dalam Tabel 4.

Table 4. Statistik Deskriptif Data Penelitian

Variabel	N	Minimum	Maximum	Mean	Standar Deviasi
X1	308	1	14	5.93	2.88
X2	308	1	14	5.44	3.05
X3	308	1	15	5.64	2.76
X4	308	1	14	5.91	2.90
X5	308	1	15	5.50	3.02
X6	308	1	15	5.79	2.89
X7	308	1	14	5.68	3.05

Berdasarkan hasil analisis statistik deskriptif pada Tabel tersebut, diketahui bahwa seluruh variabel penelitian memiliki jumlah data (N) sebanyak 308. Nilai minimum dari setiap variabel berada pada rentang 1, sedangkan nilai maksimum berkisar antara 14 hingga 15. Nilai rata-rata (*mean*) dari masing-masing variabel berada pada kisaran 5,44 hingga 5,93.

Nilai standar deviasi dari seluruh variabel berada pada kisaran 2,76 hingga 3,05 yang menunjukkan bahwa tingkat variasi data relatif tidak terlalu besar. Hal ini mengindikasikan bahwa penyebaran data pada masing-masing variabel masih berada di sekitar nilai rata-ratanya. Temuan analisis ini memberikan ringkasan awal mengenai sifat-sifat data sebelum pendekatan K-Means Clustering diterapkan.

4.5. Pengujian Asumsi

Langkah pertama dalam pengelompokan data dengan *Algoritma K-Means Clustering* adalah memeriksa asumsi untuk memastikan bahwa variabel-variabel yang digunakan sudah sesuai dan layak untuk diterapkan dianalisis.

a. Asumsi Sampel Representatif

Dalam studi yang menggunakan pendekatan Klasterisasi K-Means, pengujian asumsi harus dilakukan sebelum pengelompokan untuk memastikan kecocokan sampel dalam mewakili populasi data. Uji Kaiser-Meyer-Olkin (KMO) adalah salah satu metode untuk menilai apakah sampel cukup mewakili populasi. Hasil pengujian asumsi tersebut, yang dilakukan untuk memverifikasi representativitas sampel, ditampilkan di bawah ini.

Table 5. KMO Test

KMO Test	
KMO (Kaiser Meyer Olkin)	0.533

Nilai Kaiser-Meyer-Olkin (KMO) yang dihasilkan adalah 0,533. Skor ini, yang lebih besar dari 0,5, menunjukkan bahwa sampel dalam studi ini cukup representatif terhadap populasi, sehingga data tersebut cocok untuk penelitian lebih lanjut.

b. Uji *Non-Multikolinearitas*

Uji *non-multikolinearitas* dilakukan untuk mengetahui apakah terdapat hubungan atau korelasi antar variabel independen yang digunakan dalam penelitian ini. Asumsi ini dinyatakan terpenuhi apabila tidak ditemukan gejala multikolinearitas. Ringkasan nilai VIF (Variance Inflation Factor) dari setiap variabel penelitian disajikan sebagai berikut.

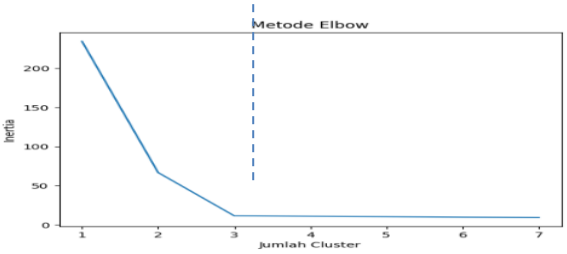
Table 6. Uji *Non-Multikolinearitas*

Variabel	Nilai VIF
X1	4.82
X2	3.98
X3	4.36
X4	5.03
X5	4.07
X6	4.92
X7	3.99

Menurut hasil uji non-multikolinearitas dalam tabel, semua variabel memiliki nilai VIF antara 3,98 dan 5,03. Karena nilai-nilai ini kurang dari 10, yang menunjukkan tidak adanya multikolinearitas di antara variabel independen, nilai-nilai tersebut berada dalam batas yang dapat diterima. Dengan demikian, variabel-variabel yang dipilih sudah sesuai untuk melanjutkan ke tahap analisis berikutnya.

4.6. Hasil Klasterisasi Menggunakan *Algoritma K-Means*

Dalam penerapan algoritma *K-Means Clustering*, penentuan jumlah cluster (k) yang optimal merupakan langkah awal yang krusial sebelum proses pengelompokan dilakukan. Pada penelitian ini, penentuan jumlah cluster optimal dilakukan menggunakan dua pendekatan yang saling melengkapi, yaitu *Elbow Method* dan *Silhouette Coefficient*. Kedua metode ini digunakan secara bersamaan agar hasil penentuan jumlah cluster optimal lebih valid dan dapat dipertanggungjawabkan secara metodologis.



Gambar 3. Jumlah Cluster Optimum

Gambar 3 menampilkan hasil pengujian menggunakan Elbow Method dengan rentang jumlah cluster k = 1 hingga k = 7. Sumbu horizontal menunjukkan jumlah cluster yang diuji, sedangkan sumbu vertikal menunjukkan nilai inertia, yaitu jumlah kuadrat jarak setiap titik data terhadap pusat cluster-nya. Semakin kecil nilai inertia, semakin rapat data dalam cluster tersebut. Berdasarkan grafik, terlihat bahwa nilai inertia mengalami penurunan yang sangat tajam dari k = 1 ke k = 2, kemudian dilanjutkan dengan penurunan yang masih cukup signifikan dari k = 2 ke k = 3. Setelah k = 3, grafik mulai melandai dan perubahan nilai inertia menjadi sangat kecil hingga k = 7. Titik di mana grafik mulai melandai inilah yang disebut sebagai titik "siku" (*elbow*), dan pada penelitian ini titik tersebut berada pada k = 3. Hal ini mengindikasikan bahwa penambahan jumlah cluster di atas 3 tidak lagi memberikan peningkatan kualitas

pengelompokan yang berarti, sehingga k = 3 dipilih sebagai jumlah cluster yang optimal berdasarkan Elbow.

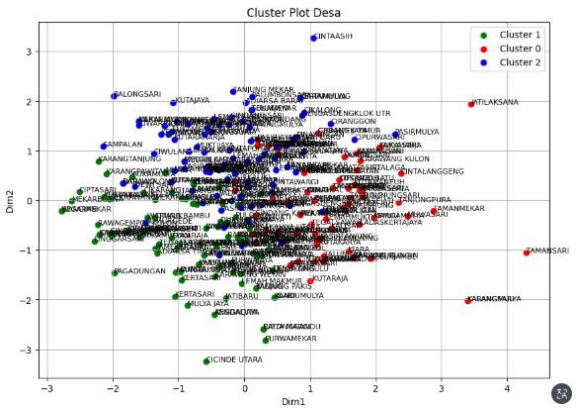
```
k=2, silhouette score=0.6354597186527698
k=3, silhouette score=0.7579467781694962
k=4, silhouette score=0.5535241596862023
k=5, silhouette score=0.3511794721049242
```

Gambar 4. Hasil Nilai Silhoutte

Gambar 4 menampilkan nilai *Silhouette Coefficient* tertinggi diperoleh pada k = 3 dengan nilai sebesar 0,757. Nilai ini lebih tinggi dibandingkan k = 2 (0,635), k = 4 (0,553), maupun k = 5 (0,351), sehingga k = 3 dipilih sebagai jumlah *cluster* yang optimal. Berdasarkan pedoman interpretasi *Silhouette Coefficient* menurut Kaufman & Rousseeuw (1990), nilai 0,757 termasuk dalam kategori struktur *cluster* sangat kuat (*strong structure*) karena berada pada rentang 0,71–1,00. Hal ini menunjukkan bahwa desa-desa yang berada dalam satu *cluster* memiliki tingkat kemiripan yang tinggi satu sama lain, sekaligus memiliki perbedaan yang jelas dengan desa-desa di *cluster* lain. Dengan demikian, hasil pengelompokan yang terbentuk dapat diandalkan sebagai dasar analisis dan pengambilan kebijakan.

Table 7. Cluster perDesa

Cluster	Desa
0	TAMANMEKAR, TAMANSARI, MEDALSARI, JATILAKSANA.
1	CIPTASARI, KERTASARI, MEKARBUANA, KUTALANGGEN.
2	CINTAASIH, MULANGSARI, CINTAWARGI, CIPURWASAR.



Gambar 5. Cluster Plot Desa

Plot klaster dalam Gambar 5 menunjukkan bahwa data desa terbagi menjadi tiga kelompok, yang diwakili oleh warna titik yang berbeda. Setiap titik mewakili sebuah desa, dengan posisi pada sumbu Dim1 dan Dim2 ditentukan oleh Analisis Komponen Utama (PCA) dari berbagai variabel indikator. Reduksi dimensi ini memungkinkan visualisasi data yang kompleks dalam dua dimensi, sehingga pola pengelompokan menjadi lebih mudah dipahami.

Table 8. Karakteristik Setiap *Cluster*

Cluster	X1	X2	X3	X4	X5	X6	X7
0	7.42	8.11	5.40	7.58	7.07	7.21	6.05
1	3.22	6.05	5.34	4.68	4.23	5.45	4.81
2	7.45	2.86	6.11	5.86	5.55	5.05	6.26

Berdasarkan analisis pengelompokan K-Means, muncul tiga kelompok yang berbeda, masing-masing menunjukkan profil yang berbeda di seluruh variabel indikator. Profil-profil ini dirangkum dalam Tabel 7, yang menyajikan nilai rata-rata dari setiap variabel untuk setiap klaster. Klaster 0 menunjukkan nilai rata-rata yang relatif tinggi pada sebagian besar indikator, seperti X1 (Air Tidak Bersih), X2 (Cacing Tambang), dan X5 (Merokok), menunjukkan bahwa desa-desa dalam kelompok ini menghadapi tingkat faktor risiko yang lebih tinggi dibandingkan yang lain. Klaster 1 umumnya menunjukkan nilai rata-rata yang lebih rendah pada beberapa variabel dibandingkan Klaster 0, mewakili desa-desa dengan risiko yang relatif lebih rendah. Sementara itu, klaster 2 menampilkan rentang nilai rata-rata di berbagai indikator, menunjukkan bahwa desa-desa ini memiliki karakteristik yang berada di antara Klaster 0 dan 1.

5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa *Algoritma K-Means Clustering* mampu mengelompokkan wilayah administratif tingkat desa berdasarkan faktor-faktor yang berkaitan dengan risiko stunting. Hasil analisis menunjukkan bahwa jumlah klaster yang optimal adalah tiga klaster, di mana setiap klaster merepresentasikan tingkat risiko yang berbeda. Klaster pertama terdiri dari desa dengan tingkat faktor risiko stunting yang tinggi, klaster kedua menunjukkan desa dengan tingkat risiko relatif rendah, sedangkan klaster ketiga menggambarkan desa dengan tingkat risiko sedang. Hasil pengelompokan ini memberikan gambaran mengenai pola distribusi faktor penyebab stunting pada wilayah penelitian sehingga dapat menjadi dasar dalam penyusunan strategi intervensi yang lebih terarah dan berbasis data.

Untuk penelitian selanjutnya, disarankan untuk menambahkan variabel yang lebih beragam serta mempertimbangkan penggunaan metode klasterisasi lain guna memperoleh hasil analisis yang lebih komprehensif dan akurat.

DAFTAR PUSTAKA

[1] Deden Renhad Sudrajat, S. Shofia Hilabi, B. Huda, and A. L. Hananto, "Implementasi Algoritma K-Means Untuk Klasterisasi Data Stunting," *INNOVATIVE: Journal Of Social Science Research*, vol. 4, no. 2, pp. 363–373, 2024, [Online]. Available: https://j-innovative.org/index.php/Innovative

[2] Bagas Aqmal Febrianto, "IDENTIFIKASI WILAYAH KEKURANGAN ENERGI KRONIS (KEK) PADA IBU HAMIL DI KABUPATEN KARAWANG

MENGGUNAKAN K-MEANS," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 3, pp. 1049–1059, Jul. 2025, doi: 10.23960/jitet.v13i3.7145.

[3] C. Syalom, U. Khaira, and D. Arsa, "CLUSTERING DAERAH RAWAN STUNTING DI INDONESIA MENGGUNAKAN ALGORITMA K-MEANS," *Jurnal Teknologi Informasi*, vol. 6, no. 1, pp. 389–401, 2025, doi: 10.46576/djtechno.

[4] F. Luthfiyah, R. Amelia, and I. Fahria, "ANALISIS CLUSTER WILAYAH KECAMATAN BERDASARKAN FAKTOR PENYEBAB STUNTING DI PROVINSI KEPULAUAN BANGKA BELITUNG," *Jurnal Ilmiah Matematika*, vol. 13, no. 02, pp. 124–136, 2025.

[5] Y. T. Dendo *et al.*, "ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN NILAI GIZI BALITA K-MEANS CLUSTERING ALGORITHM TO DETERMINE NUTRITIONAL VALUE FOR TODDLER 1)," *Jurnal Sains dan Sistem Teknologi Informasi (SANDI) CCS*, vol. 5, no. 2, pp. 243–247, 2023.

[6] A. Subayu, "PENERAPAN METODE K-MEANS UNTUK ANALISIS STUNTING GIZI PADA BALITA: SYSTEMATIC REVIEW," *Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi*, vol. 2, no. 1, pp. 42–50, Sep. 2022, doi: 10.20885/snati.v2i1.18.

[7] E. Y. Andini, "PEMETAAN WILAYAH KECAMATAN DI KABUPATEN BANYUWANGI BERDASARKAN FAKTOR PENYEBAB STUNTING TAHUN 2023 MENGGUNAKAN ANALISIS BIPLOT," *Prosiding Seminar Nasional Sains dan Teknologi Seri III Fakultas Sains dan Teknologi*, vol. 2, no. 1, pp. 430–444, 2025.

[8] F. T. Wahyuni, N. Shofie Kirana, N. Syafila 'ashifa, and E. Nazila, "Kategorisasi Tingkat Gizi Balita Menggunakan K-Means: Studi Kasus Puskesmas Desa Karangsambung," *Seminar Nasional Sains Data*, vol. 2024, no. Senada, pp. 863–873, 2024, [Online]. Available: https://temansigizi.com/.

[9] I. S. Tinendung and I. Zufria, "Pengelompokan Status Stunting Pada Anak Menggunakan Metode K-Means Clustering," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, no. 4, pp. 2014–2023, Oct. 2023, doi: 10.30865/mib.v7i4.6908.

[10] M. Handayani, M. Fitri, L. Sibuea, S. Tinggi, M. Informatika, and D. K. Royal, "OPTIMASI MODEL CLUSTERING DALAM PEMETAAN STUNTING DI KABUPATEN ASAHAN," *Journal of Science and Social Research*, vol. 7, no. 4, pp. 2192–2197, 2024, [Online]. Available: http://jurnal.goretanpena.com/index.php/JSSR

[11] F. A. Yuwono, A. L. Hananto, F. Nurapriani, and B. Huda, "Implementasi Algoritma K-means untuk Klasterisasai Data Stunting di Kabupaten Bekasi," *Remik: Riset dan E-Jurnal Manajemen Informatika Komputer*, vol. 9, no. 2, pp. 626–633, 2025, doi: 10.33395/remik.v9i2.14748.

[12] A. Lia Hananto *et al.*, "Analysis of Drug Data Mining with Clustering Technique Using K-Means Algorithm," *J. Phys. Conf. Ser.*, vol. 1908, no. 1, pp. 1–8, Jul. 2021, doi: 10.1088/1742-6596/1908/1/012024.

[13] K. Rizky Aditya Nugroho, M. Tholabah, R. Aditya Nugroho, and S. Mu, "Artikel Nusantara Computer and Design Review Penerapan Metode K-Means dalam Pengelompokkan Status Gizi Anak dan Remaja," *NCDR*, vol. 2, no. 1, pp. 7–14, 2024, [Online]. Available: https://journal.unusida.ac.id/index.php/ncdr/

[14] B. Hadi Prakoso *et al.*, "Klasterisasi Puskesmas dengan K-Means Berdasarkan Data Kualitas Kesehatan Keluarga dan Gizi Masyarakat," *Jurnal Buana Informatika*, vol. 14, no. 1, pp. 60–68, 2023.

[15] N. F. Rahman *et al.*, "PENERAPAN ALGORITMA K-MEANS UNTUK KLASTERISASI FILM PADA PLATFORM AMAZON PRIME VIDEO," *Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 5, pp. 8200–8206, 2025.

[16] Muhammad Raqib Syahkur, D. Hartama, and S. Solikhun, "Evaluasi Jumlah Cluster pada Algoritma K-Means++ Menggunakan Silhouette dan Elbow dengan Validasi Nilai DBI dalam Mengelompokkan Gizi Balita," *JST (Jurnal Sains dan Teknologi)*, vol. 13, no. 3, pp. 487–496, Oct. 2024, doi: 10.23887/jstundiksha.v13i3.86419.

[17] R. Andarini Hatti Imanni, E. Sulistianingsih, and H. Perdana INTISARI, "ANALISIS CLUSTER MENGGUNAKAN ALGORITMA K-MEANS BERDASARKAN FAKTOR PENYEBAB STUNTING PADA PROVINSI KALIMANTAN BARAT," *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)*, vol. 12, no. 3, pp. 301–308, 2023.

[18] W. Juliantika Br Sembiring Depari and D. Aditiya Nugroho, "Optimasi Cluster Menggunakan Algoritma Fuzzy C-Means Untuk Menentukan Tingkat Stunting Pada Balita Di Kabupaten Karo," *JURNAL SISTEM INFORMASI TGD* , vol. 4, no. 6, pp. 1368–1375, 2025, [Online]. Available: https://ojs.trigunadharma.ac.id/index.php/jsi

[19] M. F. Amalia and D. B. Arianto, "Implementasi Algoritma K-Means Clustering Dalam Klasterisasi Kabupaten/Kota Provinsi Jawa Barat Berdasarkan Faktor Pemicu Stunting Pada Balita," *SIMKOM*, vol. 9, no. 1, pp. 36–46, Jan. 2024, doi: 10.51717/simkom.v9i1.356.

[20] N. Oktaviany, N. Suarna, and W. Prihartono, "IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING DALAM MENGELOMPOKKAN KEPADATAN PENDUDUK DI PROVINSI DKI JAKARTA," *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 1, pp. 119–125, 2024

.