

Penerapan Algoritma K-Means dalam Klasterisasi Wilayah Prioritas Penanganan BBLR dan Risiko Stunting

Salamudin^{1*}, Haris Saputro², Rusidi³, Ari Sulfani⁴

^{1,2,3} Program Studi Teknik Informatika (Kampus Kab Ogan Komering Ulu), ⁴Program Studi Informatika Universitas Mahakarya Asia

INFORMASI ARTIKEL	ABSTRAK
Kata Kunci: K-Means, Clustering, Low Birth Weight, Malnutrition, Stunting, Data Mining.	<i>This study aims to classify districts and municipalities in South Sumatra Province based on the risk levels of Low Birth Weight (LBW) and severe malnutrition using the K-Means clustering algorithm. Secondary data from 17 regions in 2024 were analyzed using two rate-based variables, namely LBW per 100 live births and severe malnutrition per 1,000 live births, in order to generate proportional risk indicators across regions. Both variables were normalized using the Min–Max method before clustering. The optimal number of clusters was determined using the elbow method and the Davies–Bouldin Index (DBI). The elbow point appeared at $k = 3$, while DBI values for $k = 2-5$ were 0.870, 0.835, 0.607, and 0.542, respectively, with the smallest value at $k = 5$. Considering cluster stability and interpretability, $k = 3$ was selected as the optimal number of clusters. The final clustering produced three risk groups consisting of nine low-risk regions, four medium-risk regions, and three high-risk regions, namely Muara Enim, Musi Banyuasin, and Palembang. These high-risk areas demonstrated the highest LBW and malnutrition rates and therefore serve as priority targets for intervention. The findings confirm that the K-Means algorithm is effective for epidemiological risk mapping and can support evidence-based decision-making in stunting prevention efforts across South Sumatra.</i>
Diserahkan Direvisi Diterima	© Author's (202x) Creative Commons Attribution-NonCommercial 4.0 International License



1. PENDAHULUAN

Stunting merupakan masalah gizi kronis yang berdampak pada pertumbuhan fisik, perkembangan kognitif, dan kesehatan jangka panjang anak. Kondisi ini umumnya dipicu oleh kekurangan gizi jangka panjang, infeksi berulang, sanitasi yang kurang memadai, serta keterbatasan layanan kesehatan ibu dan anak. Sumber Permasalahan ini semakin kompleks karena banyak anak mengalami gangguan pertumbuhan yang berawal sejak masa kandungan, salah satunya melalui kondisi Berat Badan Lahir Rendah (BBLR), yang menjadikan intervensi sejak kehamilan sangat penting.

Secara global, UNICEF melaporkan bahwa pada tahun 2020 terdapat 19,8 juta bayi (14,7%) yang lahir dengan BBLR. Temuan tersebut menunjukkan bahwa penurunan angka BBLR secara global masih berjalan lambat. [1] WHO menegaskan bahwa BBLR meningkatkan risiko gangguan pertumbuhan, rendahnya kemampuan kognitif, serta meningkatnya angka morbiditas dan mortalitas pada anak hingga dewasa, terutama karena gangguan pertumbuhan intrauterin atau kelahiran prematur [2]. Analisis tren global 2000–2020 juga menunjukkan bahwa sebagian besar wilayah dunia mengalami penurunan BBLR yang lambat dan belum mampu memenuhi target penurunan 30% pada tahun 2030 [3].

Kondisi stunting dan BBLR juga menjadi perhatian besar di Indonesia. Hasil Survei Status Gizi Indonesia (SSGI) 2024 menunjukkan penurunan prevalensi stunting dari 21,5% pada 2023 menjadi 19,8% pada 2024, menunjukkan tren positif dalam upaya nasional percepatan penurunan stunting. Meski demikian, UNICEF mencatat bahwa kesenjangan antarwilayah masih tinggi, sehingga pemetaan risiko wilayah secara presisi menjadi kebutuhan mendesak. Pemerintah menetapkan enam provinsi dengan jumlah balita stunting tertinggi Jawa Barat, Jawa

Tengah, Jawa Timur, Sumatera Utara, NTT, dan Bantensebagai wilayah prioritas percepatan penanganan stunting[4].

Di tingkat ilmiah, berbagai penelitian di Indonesia menunjukkan bahwa BBLR merupakan salah satu faktor risiko yang paling signifikan terhadap stunting, dengan hubungan yang diperkuat oleh temuan banyak studi. Sebuah tinjauan literatur terhadap sembilan penelitian menemukan bahwa enam penelitian menyatakan adanya hubungan bermakna antara BBLR dan stunting pada balita. Studi lain menegaskan bahwa bayi BBLR lebih rentan terhadap gangguan perkembangan dan pertumbuhan linear sehingga berpotensi besar menjadi stunting pada usia anak-anak. Lebih jauh lagi, penelitian di Desa Mayangrejo, Bojonegoro menunjukkan bahwa balita dengan riwayat BBLR memiliki risiko tiga kali lebih tinggi mengalami stunting dibandingkan balita yang lahir dengan berat normal [5].

Upaya pemerintah dalam menurunkan angka stunting menekankan intervensi berbasis siklus kehidupan, mulai dari prakehamilan hingga pascakelahiran, seperti pemberian tablet tambah darah untuk remaja putri, pengukuran hemoglobin ibu hamil, peningkatan cakupan ASI eksklusif, serta penguatan Posyandu dengan penyediaan alat antropometri standar. Namun, agar intervensi tersebut tepat sasaran, diperlukan pemetaan wilayah prioritas berdasarkan indikator risiko yang akurat, termasuk integrasi indikator BBLR dan gizi buruk sebagai indikator risiko stunting.

Dalam pendekatan berbasis data, algoritma K-Means telah menjadi salah satu metode klusterisasi yang paling banyak digunakan untuk mengelompokkan wilayah berdasarkan kemiripan karakteristik numerik. Columbia Public Health menjelaskan bahwa K-Means efektif dalam membentuk kelompok yang homogen menggunakan jarak Euclidean untuk menentukan pusat kluster secara iteratif [6]. Selain itu, penelitian kesehatan populasi di Amerika Serikat membuktikan bahwa K-Means dapat digunakan untuk menentukan cluster wilayah yang relevan dalam evaluasi kebijakan kesehatan dan intervensi berbasis komunitas.

Pemanfaatan K-Means dalam analisis stunting juga telah dilakukan di berbagai daerah di Indonesia. Penelitian oleh [7]berhasil membagi desa menjadi kluster rendah, sedang, dan tinggi sesuai tingkat risiko stunting sehingga memudahkan pemerintah daerah menentukan prioritas intervensi. Penelitian lain di Kabupaten Batang menghasilkan nilai Silhouette Coefficient 0.60830, menandakan kualitas kluster sangat baik untuk memetakan tingkat keparahan stunting per wilayah[8]. Studi tingkat provinsi juga mengonfirmasi bahwa metode Silhouette penting dalam menentukan jumlah kluster optimal sebelum menerapkan algoritma K-Means[9].

Namun demikian, sebagian besar penelitian klusterisasi stunting di Indonesia masih terbatas pada penggunaan indikator stunting secara langsung dan belum mengintegrasikan indikator hulu dan hilir seperti BBLR dan gizi buruk sebagai penentu risiko stunting, padahal kedua indikator ini telah terbukti berkaitan erat dengan kualitas pertumbuhan anak sejak masa kandungan hingga awal kehidupan. Penggunaan BBLR sebagai indikator hulu sangat relevan karena stunting merupakan kondisi kronis yang berawal sejak masa kehamilan, sedangkan gizi buruk mencerminkan kondisi malnutrisi akut yang dapat berkembang menjadi masalah gizi kronis apabila tidak ditangani. Dengan demikian, kombinasi kedua indikator ini dapat digunakan sebagai proksi risiko stunting yang valid berdasarkan bukti epidemiologis dan arah kausalitas biologis. Pendekatan ini selaras dengan rilis resmi SSGI 2024 yang menegaskan bahwa stunting telah dimulai sejak masa kehamilan sehingga intervensi pada faktor risiko hulu menjadi sangat penting dalam percepatan penurunan stunting. Oleh karena itu, penerapan algoritma K-Means untuk mengklusterisasi wilayah berdasarkan indikator BBLR dan gizi buruk sebagai penentu risiko stunting menjadi langkah strategis dalam pemetaan wilayah prioritas yang lebih komprehensif dan tepat sasaran

2. TEORI DAN PENELITIAN RELEVAN

Stunting merupakan masalah gizi kronis yang menyebabkan gangguan pertumbuhan linear, hambatan perkembangan kognitif, serta menurunkan kualitas kesehatan jangka panjang anak,

di mana faktor-faktor seperti defisiensi gizi, infeksi berulang, kualitas sanitasi yang buruk, dan keterbatasan layanan kesehatan berperan besar dalam meningkatkan risiko terjadinya kondisi tersebut [10]. Gangguan ini tidak hanya berkaitan dengan aspek biologis, tetapi juga dipengaruhi oleh kondisi lingkungan dan sosial ekonomi, sehingga stunting dikategorikan sebagai masalah kompleks yang menuntut penanganan multisektor (Sholihah, 2023). Stunting pada masa awal kehidupan telah terbukti berimplikasi panjang terhadap produktivitas dan kualitas sumber daya manusia, sebab anak yang mengalami stunting cenderung menghadapi hambatan dalam aspek pendidikan, perkembangan intelektual, serta risiko penyakit kronis ketika dewasa (Syahrul Khairati et al., 2024).

Dalam rantai penyebab stunting, berat badan lahir rendah (BBLR) merupakan salah satu determinan paling penting yang mencerminkan kualitas pertumbuhan sejak periode prenatal, terutama karena terkait secara langsung dengan prematuritas maupun gangguan pertumbuhan intrauterin (IUGR) [10]. Studi dalam Jurnal Kesehatan Akimal (2025) memperlihatkan bahwa bayi dengan BBLR memiliki risiko lebih tinggi mengalami gangguan metabolisme, kerentanan infeksi, dan hambatan pertumbuhan pada periode awal kehidupan, sehingga meningkatkan peluang terjadinya stunting di tahun-tahun berikutnya [12]. Hal ini sejalan dengan penelitian kasus-kontrol di Kabupaten Lamongan, yang menemukan hubungan signifikan antara BBLR dan stunting dengan nilai $p = 0,022$, menguatkan posisi BBLR sebagai prediktor kuat kejadian stunting di tingkat komunitas [10]. Selain berat lahir, indikator antropometri lain yakni panjang badan lahir rendah juga terbukti signifikan dalam meningkatkan risiko stunting, sebagaimana dilaporkan dalam penelitian di Yogyakarta bahwa panjang badan lahir rendah berhubungan erat dengan stunting usia 1–5 tahun ($p = 0,007$; $OR = 3,95$) [13].

Temuan pada tingkat nasional melalui Jurnal Ilmiah Kesehatan Masyarakat (JIKM, 2025) semakin memperkuat hubungan BBLR–stunting, di mana dari 13 studi yang dianalisis, 77% menyatakan adanya hubungan signifikan antara BBLR dan stunting, dengan risiko stunting meningkat hingga 2,19–2,42 kali pada bayi yang mengalami BBLR [14]. Konsistensi bukti ilmiah tersebut menunjukkan bahwa intervensi pencegahan stunting harus menasar fase hulu, mulai dari gizi maternal, pemantauan kehamilan, hingga kualitas layanan obstetri dan neonatal, guna mencegah terjadinya BBLR yang berimplikasi panjang terhadap kualitas pertumbuhan anak.

Selain memahami faktor risiko, penentuan wilayah prioritas untuk intervensi pencegahan stunting membutuhkan pendekatan analitis yang mampu mengelompokkan wilayah berdasarkan karakteristik yang serupa, salah satunya melalui algoritma K-Means Clustering. Penelitian oleh [7] menunjukkan bahwa K-Means efektif dalam mengelompokkan desa menjadi tiga kluster risiko stunting, yaitu rendah, sedang, dan tinggi, sehingga membantu proses perencanaan dan pengambilan keputusan berbasis data pada tingkat daerah. Keakuratan metode ini juga dibuktikan melalui studi di Kabupaten Batang yang menghasilkan nilai Silhouette Coefficient sebesar 0.60830, mengindikasikan kualitas kluster yang sangat baik dalam memetakan tingkat keparahan stunting (Pramudya et al., 2024). Pada tataran provinsi, penelitian oleh [9] dalam jurnal DjTechno menegaskan pentingnya metode Silhouette untuk menentukan jumlah kluster optimal sebelum penerapan K-Means, sehingga memastikan pemetaan risiko stunting lebih representatif dan akurat.

Penelitian-penelitian lain yang dilakukan pada isu terkait gizi dan kesehatan masyarakat juga memperkuat relevansi penggunaan K-Means. Misalnya, Jurnal Jurnal Biometrika dan Kependudukan menemukan bahwa K-Means mampu mengidentifikasi pola status gizi balita dalam beberapa kluster yang mencerminkan variasi karakteristik kesehatan antarwilayah [15]. Selain itu, Bina Insani ICT Journal juga membuktikan bahwa K-Means dapat digunakan untuk memetakan distribusi fasilitas kesehatan, yang meskipun tidak langsung mengukur stunting, tetapi sangat relevan dalam memahami akses layanan kesehatan yang turut berpengaruh pada status gizi anak (Islami et al., 2021). Secara keseluruhan, literatur nasional hingga tahun 2025 menegaskan bahwa integrasi indikator BBLR dan stunting dalam analisis klusterisasi dengan algoritma K-Means merupakan pendekatan yang kuat, komprehensif, dan berbasis bukti untuk memetakan wilayah prioritas penanganan stunting di Indonesia (Safhira et al., 2025).

Penelitian-penelitian sebelumnya telah menegaskan hubungan kuat antara BBLR dan stunting serta memanfaatkan algoritma K-Means untuk mengelompokkan wilayah berdasarkan karakteristik kesehatan. Namun, studi-studi tersebut umumnya hanya menggunakan indikator stunting secara langsung dan belum mengintegrasikan kombinasi indikator hulu (BBLR) dan indikator hilir (gizi buruk) sebagai dasar pemetaan risiko. Selain itu, belum terdapat penelitian yang menerapkan klasterisasi berbasis K-Means menggunakan kedua indikator tersebut untuk data Kabupaten/Kota di Provinsi Sumatera Selatan. Oleh karena itu, penelitian ini menawarkan pendekatan yang lebih komprehensif dengan memanfaatkan BBLR dan gizi buruk sebagai proksi risiko stunting dalam membentuk kelompok wilayah prioritas.

Selain pemilihan algoritma, implementasi proses klasterisasi dalam penelitian ini juga didukung oleh penggunaan RapidMiner Studio sebagai perangkat lunak data mining. RapidMiner menyediakan lingkungan analitis berbasis *workflow* yang memudahkan peneliti dalam melakukan tahapan penting seperti pembersihan data, normalisasi Min-Max, pembentukan atribut baru, dan penerapan algoritma K-Means secara terstruktur. Keunggulan antarmuka *draganddrop* meminimalkan kesalahan teknis dan memastikan alur pemrosesan data dapat direplikasi dengan baik. Selain itu, RapidMiner memiliki modul evaluasi seperti *Cluster Distance Performance* dan *Performance (Cluster)* yang menghasilkan nilai *Avg. within centroid distance* dan *DaviesBouldin Index*, sehingga proses penentuan jumlah klaster optimal dapat dilakukan secara objektif dan terukur. Penggunaan RapidMiner dalam penelitian ini tidak hanya mempermudah implementasi algoritma K-Means, tetapi juga memastikan bahwa analisis dilakukan mengikuti standar metodologi data mining yang baik dan transparan.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif deskriptif-analitik untuk mengelompokkan 17 Kabupaten/Kota di Provinsi Sumatera Selatan berdasarkan tingkat risiko BBLR dan gizi buruk. Proses analisis dilakukan menggunakan RapidMiner Studio, yang menyediakan alur kerja analitik data secara terstruktur mulai dari tahap ekstraksi data hingga klasterisasi menggunakan algoritma K-Means.

3.1 Jenis dan Desain Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan desain deskriptif-analitik. Tujuannya adalah mengidentifikasi pola pengelompokan wilayah melalui algoritma K-Means Clustering, sebuah metode unsupervised learning yang membagi objek menjadi beberapa kelompok berdasarkan kemiripannya. Metode ini dipilih karena sesuai untuk variabel numerik seperti BBLR dan gizi buruk serta mampu menangkap variasi risiko antarwilayah secara objektif.

3.2 Sumber Data

Data yang digunakan dalam analisis ini merupakan data sekunder tahun 2024 yang mencakup 17 Kabupaten/Kota di Provinsi Sumatera Selatan. Dataset terdiri dari tiga indikator utama, yaitu jumlah bayi lahir, jumlah bayi dengan BBLR, dan jumlah balita gizi buruk. Tabel berikut menampilkan data awal yang digunakan dalam penelitian sebelum dilakukan transformasi dan pengolahan lebih lanjut:

Tabel 1.jumlah bayi lahir, jumlah bayi dengan BBLR, dan jumlah balita gizi buruk

Kabupaten/Kota 17	Jumlah Bayi Lahir, Berat Bayi Lahir Rendah, dan Bergizi Buruk (Jiwa)		
	Bayi Lahir	Berat Badan Lahir Rendah	Gizi Buruk
	2024	2024	2024
Ogan Komering Ulu	5055	74	508
Ogan Komering Ilir	8599	105	685

Muara Enim	9511	290	249
Lahat	7353	50	365
Musi Rawas	6069	144	825
Musi Banyuasin	12746	301	247
Banyuasin	15087	94	585
Ogan Komering Ulu Selatan	4674	30	307
Ogan Komering Ulu Timur	10505	159	136
Ogan Ilir	6972	167	1319
EmpatLawang	5024	46	70
Pali	3498	230	57
Musi Rawas Utara	3320	38	280
Palembang	28244	369	663
Prabumulih	3508	99	177
Pagar Alam	2554	1	41
LubukLinggau	4409	91	235

Tabel tersebut menjadi dasar untuk proses analisis berikutnya. Sebagaimana terlihat, terdapat variasi jumlah kelahiran dan kejadian BBLR maupun gizi buruk antarwilayah. Variasi ini penting untuk dianalisis menggunakan klasterisasi karena dapat menunjukkan kelompok wilayah dengan karakteristik risiko yang mirip.

Proses pengolahan data dilakukan menggunakan aplikasi RapidMiner Studio, dengan tahapan yang dijalankan melalui *workflow operator*. Tahap pertama adalah memuat data mentah menggunakan operator Read CSV, kemudian dilakukan pengecekan awal pada panel Statistics untuk memastikan tidak ada nilai hilang ataupun kesalahan format angka

3.3 Variabel Penelitian dan Transformasi Data

Variabel asli yang digunakan terdiri dari:

- Jumlah Bayi Lahir
- Jumlah BBLR
- Jumlah Gizi Buruk

Karena jumlah kelahiran antarwilayah berbeda, data diubah menjadi rate epidemiologis untuk menghasilkan ukuran risiko yang proporsional. Transformasi dilakukan sebagai berikut:

BBLR per 100 kelahiran

$$BBLR_{per100} = \frac{\text{Jumlah BBLR}}{\text{Jumlah Bayi Lahir}} \times 100$$

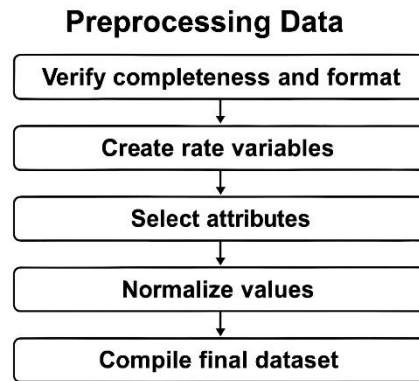
Gizi Buruk per 1000 kelahiran

$$GiziBuruk_{per1000Births} = \frac{\text{Jumlah Gizi Buruk}}{\text{Jumlah Bayi Lahir}} \times 1000$$

Transformasi ini penting agar setiap wilayah dapat dibandingkan secara adil, sesuai praktik epidemiologi dan statistik vital. Proses pembentukan variabel dilakukan melalui operator Generate Attributes pada RapidMiner Studio

3.4 Preprocessing Data

Tahapan preprocessing dilakukan untuk memastikan data siap diproses dalam algoritma klusterisasi. Tahapan tersebut meliputi:



Gambar 1. Preprocessing data

Proses preprocessing data dilakukan untuk memastikan bahwa seluruh data siap digunakan dalam analisis klusterisasi menggunakan algoritma K-Means. Tahap ini dimulai dengan memverifikasi kelengkapan dan konsistensi format data bayi lahir, BBLR, dan gizi buruk pada 17 kabupaten/kota. Setelah data dinyatakan valid, dilakukan pembuatan variabel rate berupa BBLR per 100 kelahiran dan gizi buruk per 1000 kelahiran agar perbandingan antarwilayah lebih proporsional. Selanjutnya, hanya atribut yang relevan dipilih untuk analisis, kemudian seluruh variabel dinormalisasi dengan metode Min-Max agar berada pada rentang 0–1 sehingga sesuai untuk perhitungan jarak Euclidean pada K-Means. Tahap terakhir adalah menyusun dataset akhir yang telah bersih, terstandarisasi, dan siap digunakan untuk proses klusterisasi.

3.5 Penentuan Jumlah Kluster Optimal

Penentuan jumlah kluster (k) merupakan tahap penting sebelum menerapkan algoritma K-Means karena pemilihan k yang tepat akan menghasilkan struktur kluster yang lebih representatif. Pada penelitian ini, evaluasi jumlah kluster dilakukan menggunakan dua ukuran performa yang tersedia pada RapidMiner, yaitu *Avg. within centroid distance* dan *Davies Bouldin Index*. Keduanya digunakan untuk menilai tingkat kekompakan dan pemisahan antar kluster pada berbagai nilai k .

Proses pengujian dilakukan dengan menjalankan operator K-Means dan Performance (*Cluster Distance Performance*) untuk nilai $k = 2, 3, 4$, dan 5 . Hasil pengujian dicatat pada Tabel 2 dan Tabel 3 berikut.

Tabel 2 Nilai *Avg. within centroid distance* untuk Berbagai Nilai k

k	Cluster Distance Performance
2	0.120
3	0.071
4	0.039
5	0.027

Dari Tabel 4.2 terlihat bahwa nilai *Avg. within centroid distance* menurun secara konsisten seiring bertambahnya nilai k . Penurunan relatif besar terjadi dari $k = 2$ ke $k = 3$ (selisih 0.049), kemudian menurun lebih kecil pada $k = 4$ (selisih 0.032), dan semakin kecil pada $k = 5$ (selisih 0.012). Pola ini menunjukkan bahwa peningkatan jumlah kluster setelah $k = 3$ mulai memberikan perbaikan yang semakin kecil, sehingga mengindikasikan adanya titik “*elbow*” pada $k = 3$.

Tabel 3. Nilai Davies-Bouldin Index untuk Berbagai Nilai k

k	Davies-Bouldin Index
2	0.870
3	0.835
4	0.607
5	0.542

Davies Bouldin Index (DBI) digunakan sebagai indikator kualitas pemisahan kluster, di mana nilai yang lebih kecil menunjukkan kluster yang lebih baik. Berdasarkan Tabel 4.3, nilai DBI menunjukkan pola menurun dari $k = 2$ hingga $k = 5$. Nilai terkecil terdapat pada $k = 5$ (0.542), yang secara matematis menunjukkan kualitas pemisahan kluster terbaik.

Untuk menentukan jumlah kluster final, kedua indikator dievaluasi secara bersamaan:

- Avg. within centroid distance menunjukkan pola elbow di sekitar $k = 3$, karena setelah nilai tersebut penurunan jarak menjadi relatif kecil.
- DaviesBouldin Index menunjukkan nilai terbaik pada $k = 5$.

Namun, dalam analisis klusterisasi praktis, pemilihan k tidak hanya berpatokan pada nilai DBI terkecil, tetapi juga mempertimbangkan stabilitas kluster, ukuran penurunan antar nilai, serta tujuan analisis. Pada data berjumlah kecil seperti 17 kabupaten/kota, penggunaan k yang terlalu besar (misalnya $k = 5$) dapat menghasilkan kluster yang terlalu kecil sehingga kurang informatif secara substantif. Dengan mempertimbangkan:

- Titik elbow yang jelas pada $k = 3$
- Penurunan DBI yang cukup besar dari $k = 3 \rightarrow k = 4$, tetapi tidak signifikan lagi setelah itu
- Keseimbangan antara kualitas kluster dan interpretasi praktis

maka nilai $k = 3$ dipilih sebagai jumlah kluster optimal dalam penelitian ini.

3.6 Prosedur Klusterisasi Menggunakan K-Means

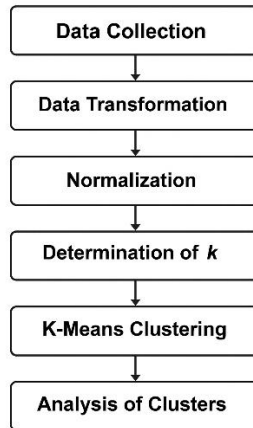
Proses klusterisasi dilakukan menggunakan algoritma K-Means pada aplikasi RapidMiner Studio. Tahap ini dimulai setelah seluruh data melewati proses preprocessing, yaitu verifikasi data, transformasi variabel menjadi bentuk rate, seleksi atribut, dan normalisasi Min-Max. Dataset yang telah distandarisasi kemudian digunakan sebagai input utama dalam proses klusterisasi.

Algoritma K-Means bekerja dengan membagi data ke dalam k kelompok berdasarkan kedekatan nilai antar-observasi menggunakan jarak *Euclidean*. Pada penelitian ini, nilai k telah ditetapkan sebanyak 3 kluster, berdasarkan hasil evaluasi pada metode *Elbow* dan *Davies-Bouldin Index*. Setelah k ditentukan, RapidMiner secara otomatis menginisialisasi centroid awal secara acak (*random initialization*) dan melakukan proses iteratif untuk memperbarui centroid hingga kondisi konvergen tercapai, yaitu ketika perpindahan centroid sudah sangat kecil atau iterasi maksimum tercapai.

3.7 Diagram Alur Penelitian

Diagram alur penelitian pada gambar berikut menjelaskan tahapan utama yang dilakukan dalam proses analisis menggunakan algoritma K-Means. Penelitian dimulai dengan pengumpulan data dari BPS Provinsi Sumatera Selatan[17], kemudian dilanjutkan dengan transformasi data untuk membentuk variabel laju BBLR per 100 kelahiran dan gizi buruk per 1000 kelahiran. Setelah itu, data dinormalisasi menggunakan metode Min-Max agar variabel

berada pada skala yang seragam. Tahap berikutnya adalah penentuan jumlah kluster optimal (k) melalui evaluasi nilai Avg. Within Centroid Distance dan Davies–Bouldin Index. Dengan $k = 3$ yang terpilih, proses klasterisasi K-Means dijalankan menggunakan RapidMiner Studio. Tahap terakhir adalah analisis kluster, yaitu menginterpretasikan nilai centroid dan distribusi wilayah dalam setiap kluster untuk menentukan tingkat risiko masing-masing daerah.



Gambar 2. Alur Penelitian

4. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil analisis yang diperoleh melalui rangkaian proses pengolahan data dan klasterisasi menggunakan algoritma K-Means pada aplikasi RapidMiner Studio. Seluruh tahapan yang telah dijelaskan pada metode penelitian mulai dari pembacaan data, pembentukan variabel berbasis rate, normalisasi, penentuan jumlah kluster optimal, hingga pengelompokan wilayah berdasarkan karakteristik risikoditampilkan secara terstruktur untuk memberikan gambaran menyeluruh mengenai pola BBLR dan gizi buruk pada 17 Kabupaten/Kota di Sumatera Selatan. Pembahasan pada bagian ini tidak hanya memaparkan hasil teknis klasterisasi, tetapi juga menafsirkan temuan tersebut dalam konteks risiko kesehatan masyarakat serta keterkaitannya dengan literatur dan penelitian sebelumnya.

4.1 Pengolahan Data

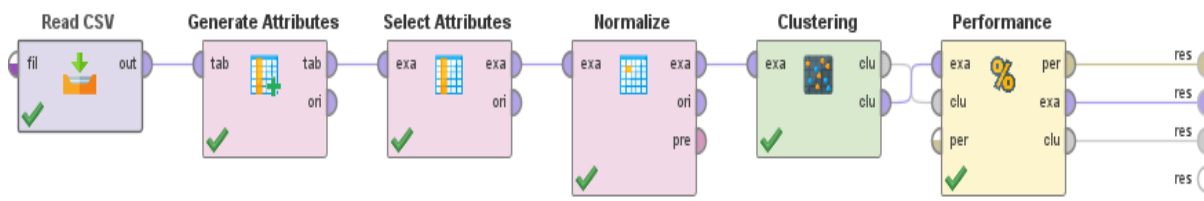
Data yang digunakan dalam analisis ini merupakan data sekunder tahun 2024 yang mencakup 17 Kabupaten/Kota di Provinsi Sumatera Selatan. Dataset terdiri dari tiga indikator utama, yaitu jumlah bayi lahir, jumlah bayi dengan BBLR, dan jumlah balita gizi buruk. Tabel berikut menampilkan data awal yang digunakan dalam penelitian sebelum dilakukan transformasi dan pengolahan lebih lanjut:

Proses pengolahan data dilakukan menggunakan aplikasi RapidMiner Studio, dengan tahapan yang dijalankan melalui *workflow operator*. Tahap pertama adalah memuat data mentah menggunakan operator Read CSV, kemudian dilakukan pengecekan awal pada panel Statistics untuk memastikan tidak ada nilai hilang ataupun kesalahan format angka.

Selanjutnya, pembentukan variabel baru dilakukan menggunakan operator Generate Attributes, yaitu: BBLR per 100 kelahiran, dan Gizi Buruk per 1000 kelahiran, sehingga perbandingan risiko antarwilayah dapat dilakukan secara proporsional. Setelah variabel rate terbentuk, operator Select Attributes digunakan untuk memilih hanya atribut yang dibutuhkan dalam proses klasterisasi.

Tahap berikutnya adalah normalisasi. Menggunakan operator Normalize (Min–Max), seluruh variabel diubah menjadi rentang nilai 0–1. Normalisasi ini diperlukan untuk mencegah bias skala dalam perhitungan jarak Euclidean pada algoritma K-Means.

Seluruh rangkaian proses tersebut dapat dilihat pada *workflow RapidMiner* berikut:



Gambar 3.workflow RapidMiner

Workflow ini menunjukkan tahapan terstruktur mulai dari pembacaan data, pembentukan atribut, pemilihan variabel, normalisasi, hingga data siap dianalisis pada tahap penentuan jumlah kluster optimal.

4.2 Hasil Klasterisasi Menggunakan K-Means

Setelah nilai k optimal ditetapkan sebanyak tiga kluster, proses klasterisasi dilakukan menggunakan operator K-Means pada RapidMiner dengan parameter $k = 3$. Proses ini menghasilkan pengelompokan 17 Kabupaten/Kota di Provinsi Sumatera Selatan berdasarkan dua indikator risiko, yaitu BBLR per 100 kelahiran dan Gizi Buruk per 1000 kelahiran, keduanya telah dinormalisasi sebelumnya. Hasil klasterisasi ini memberikan gambaran mengenai wilayah yang memiliki karakteristik risiko yang serupa.

RapidMiner menghasilkan dua keluaran utama pada tahap ini, yaitu nilai centroid untuk setiap kluster dan daftar kabupaten/kota yang tergabung dalam masing-masing kluster.

a. Centroid Kluster

Centroid menggambarkan nilai rata-rata setiap variabel dalam suatu kluster. Pada penelitian ini variabel yang digunakan ialah:

- BBLR per 100 kelahiran (setelah normalisasi Min-Max)
- Gizi Buruk per 1.000 kelahiran (setelah normalisasi Min-Max)

Secara umum, pola centroid yang terbentuk adalah sebagai berikut:

- Kluster 1 → Risiko Rendah, ditandai dengan nilai BBLR dan gizi buruk rendah.
- Kluster 2 → Risiko Sedang, menunjukkan nilai variabel yang berada pada tingkat menengah.
- Kluster 3 → Risiko Tinggi, dengan nilai kedua indikator yang paling tinggi.

Pola ini menunjukkan adanya perbedaan yang jelas antarwilayah berdasarkan tingkat risiko kesehatan yang ditinjau

b. Distribusi Wilayah dalam Setiap Kluster

Berdasarkan hasil proses K-Means pada RapidMiner, 17 Kabupaten/Kota dikelompokkan ke dalam tiga kluster. Setiap kluster mewakili kelompok wilayah dengan tingkat risiko yang serupa berdasarkan indikator BBLR dan gizi buruk.

Tabel 4. Kluster

Kluster	Anggota wilayah
Kluster 1 (Risiko Rendah)	Lahat, Ogan Komering Ulu Selatan, Ogan Komering Ulu Timur, Empat Lawang, Pali, Musi Rawas Utara, Prabumulih, Pagar Alam, Lubuk Linggau
Kluster 2 (Risiko Sedang)	Ogan Komering Ilir, Musi Rawas, Banyuasin, Ogan Ilir
Kluster 3 (Risiko Tinggi)	Muara Enim, Musi Banyuasin, Palembang

Tabel tersebut menunjukkan bahwa ketiga kluster memiliki karakteristik risiko yang berbeda berdasarkan nilai BBLR per 100 kelahiran dan gizi buruk per 1.000 kelahiran. Kluster 1 terdiri dari wilayah-wilayah dengan tingkat risiko relatif rendah yang ditandai dengan rendahnya nilai kedua indikator. Kluster 2 berisi wilayah dengan tingkat risiko sedang, di mana nilai BBLR dan gizi buruk mulai menunjukkan peningkatan. Sementara itu,

Klaster 3 merupakan kelompok wilayah dengan tingkat risiko paling tinggi dan menjadi prioritas utama intervensi karena memiliki kombinasi nilai BBLR dan gizi buruk yang jauh lebih besar dibanding dua klaster lainnya. Pembagian ini menegaskan kemampuan algoritma K-Means dalam mengelompokkan wilayah berdasarkan pola risiko yang mirip dan memberikan dasar bagi analisis lebih lanjut mengenai faktor penyebab serta implikasi kebijakan kesehatan masyarakat.

4.3 Pembahasan

Hasil klasterisasi menggunakan algoritma K-Means menghasilkan tiga klaster wilayah dengan tingkat risiko berbeda berdasarkan dua indikator utama, yaitu BBLR per 100 kelahiran dan Gizi Buruk per 1.000 kelahiran. Kedua indikator ini merupakan variabel numerik yang memungkinkan K-Means bekerja optimal karena karakteristiknya sesuai dengan kebutuhan algoritma berbasis jarak (*distance-based*). Dengan demikian, pembagian klaster yang terbentuk menggambarkan pola kemiripan antarwilayah dan memberikan dasar analitis dalam menentukan wilayah prioritas penanganan risiko stunting.

a. Interpretasi Klaster 1 (Risiko Rendah)

Klaster 1 beranggotakan wilayah: Lahat, OKU Selatan, OKU Timur, Empat Lawang, PALL, Musi Rawas Utara, Prabumulih, Pagar Alam, dan Lubuk Linggau.

Ciri utama wilayah dalam klaster ini adalah:

- Tingkat BBLR relatif rendah dibandingkan wilayah lain.
- Angka gizi buruk berada pada kategori rendah.
- Sebagian besar wilayah memiliki jumlah kelahiran yang tidak terlalu tinggi sehingga risiko lebih terkendali.

Hasil ini mengindikasikan bahwa wilayah dalam klaster 1 memiliki kondisi layanan kesehatan dasar yang relatif baik. Kemampuan klaster untuk mengelompokkan wilayah-wilayah ini membuktikan bahwa K-Means dapat menangkap pola data yang mengarah pada kelompok risiko rendah. Selain itu, karakteristik wilayah dalam klaster 1 menggambarkan bahwa intervensi yang diperlukan lebih bersifat pemeliharaan serta monitoring berkala, tanpa memerlukan penanganan khusus.

b. Interpretasi Klaster 2 (Risiko Sedang)

Wilayah dalam klaster 2 meliputi: Ogan Komering Ilir, Musi Rawas, Banyuasin, dan Ogan Ilir.

Ciri umum klaster ini adalah:

- BBLR berada pada tingkat menengah.
- Angka gizi buruk bervariasi dan menunjukkan kecenderungan meningkat.
- Beberapa wilayah memiliki cakupan geografis yang sangat luas atau jumlah penduduk besar.

Wilayah dalam klaster ini umumnya memiliki karakteristik geografis luas dan jumlah penduduk besar. Kondisi tersebut sering menimbulkan ketimpangan akses kesehatan dan variasi konsumsi gizi, sehingga indikator risiko berada pada tingkat sedang. Hasil klasterisasi pada kelompok ini menunjukkan bahwa K-Means mampu memisahkan wilayah dengan tingkat risiko menengah berdasarkan pola distribusi data. Klaster 2 memerlukan intervensi penguatan layanan, seperti peningkatan cakupan ANC dan pemerataan layanan posyandu pada wilayah yang sulit dijangkau.

c. Interpretasi Klaster 3 (Risiko Tinggi)

Klaster 3 beranggotakan wilayah: Muara Enim, Musi Banyuasin, dan Palembang.

Klaster ini memiliki karakteristik:

- Nilai BBLR tertinggi di antara semua wilayah.
- Nilai gizi buruk sangat tinggi dan menjadi indikator utama tingginya risiko.
- Terdapat wilayah berpenduduk sangat besar seperti Palembang yang menyumbang jumlah kasus gizi buruk signifikan.

Tingginya angka BBLR di wilayah-wilayah ini kemungkinan berkaitan dengan:

- Kondisi kehamilan yang kurang optimal
- Paparan risiko akibat faktor pekerjaan atau lingkungan industri
- Ketimpangan akses layanan kesehatan pada kelompok masyarakat rentan
- Beban populasi yang tinggi (khususnya Palembang)

Tingginya nilai kedua indikator menunjukkan kondisi kerentanan yang memerlukan penanganan prioritas. Klaster ini memiliki rentang data ekstrem dibanding klaster lainnya, sehingga wajar jika K-Means menempatkan wilayah-wilayah tersebut dalam kelompok risiko tinggi. Pengelompokan ini mendukung temuan literatur bahwa BBLR merupakan faktor risiko kuat terhadap terjadinya stunting, dan semakin diperparah dengan tingginya jumlah kasus gizi buruk. Secara informatika, pemisahan klaster ketiga ini memperlihatkan bahwa K-Means berhasil mendeteksi outlier kelompok risiko tinggi berdasarkan jarak centroid yang paling besar.

d. Kesesuaian Hasil dengan Tujuan dan Metode

Pembagian tiga klaster membuktikan bahwa K-Means berhasil mengidentifikasi:

- Pola persebaran risiko antarwilayah
- Perbedaan tingkatan risiko berdasarkan data numerik
- Struktur data yang secara natural terkelompok menjadi tiga bagian

Pemilihan $k = 3$ juga konsisten dengan metode *elbow* dan penurunan nilai *DaviesBouldin Index*. Secara metodologis, hasil ini menunjukkan bahwa:

- Normalisasi Min-Max yang digunakan berhasil menyetarakan skala variabel
- Penggunaan jarak Euclidean sesuai untuk dua variabel numerik
- Proses pembentukan centroid menghasilkan pengelompokan yang stabil

Dengan demikian, pembahasan ini memperkuat tujuan penelitian yaitu melakukan pemetaan wilayah prioritas penanganan BBLR dan risiko stunting berbasis algoritma K-Means, dan hasilnya menunjukkan bahwa pendekatan ini efektif digunakan untuk analisis berbasis spasial dan epidemiologis.

5. KESIMPULAN

Penelitian ini menunjukkan bahwa penerapan algoritma K-Means efektif dalam mengelompokkan 17 Kabupaten/Kota di Provinsi Sumatera Selatan berdasarkan tingkat risiko BBLR dan gizi buruk sebagai indikator penting dalam pemetaan wilayah prioritas penanganan stunting. Transformasi variabel menjadi bentuk rate, yakni BBLR per 100 kelahiran dan gizi buruk per 1.000 kelahiran, menghasilkan ukuran risiko yang lebih proporsional antarwilayah dan menghilangkan bias akibat perbedaan jumlah kelahiran. Proses normalisasi Min-Max yang diterapkan sebelum klasterisasi berhasil memastikan kedua variabel berada pada skala yang setara sehingga perhitungan jarak *Euclidean* pada algoritma K-Means dapat berjalan dengan optimal. Berdasarkan analisis *Avg. within centroid distance* dan *Davies-Bouldin Index*, jumlah klaster optimal ditetapkan sebanyak tiga klaster, karena $k = 3$ memberikan struktur klaster yang stabil, mudah diinterpretasikan, serta sesuai dengan distribusi data.

Hasil klasterisasi menunjukkan bahwa wilayah pada klaster risiko rendah memiliki nilai BBLR dan gizi buruk yang relatif kecil, sementara klaster risiko sedang menunjukkan adanya variasi yang lebih signifikan antarwilayah. Adapun klaster risiko tinggi terdiri dari wilayah yang memiliki nilai BBLR dan gizi buruk paling besar, sehingga menjadi fokus utama dalam penanganan risiko stunting. Temuan ini menegaskan bahwa algoritma K-Means mampu mengidentifikasi pola risiko berdasarkan kemiripan karakteristik data dan membedakan wilayah ke dalam kelompok yang representatif sesuai tingkat kerentanannya. Secara umum, penelitian ini membuktikan bahwa pendekatan data mining dapat dimanfaatkan untuk mendukung analisis *epidemiologis* dan menjadi dasar pengambilan keputusan berbasis data. Hasil pemetaan klaster ini diharapkan dapat memberikan acuan bagi pemerintah daerah dalam

merancang strategi intervensi yang lebih terarah dan efektif dalam upaya penurunan BBLR, gizi buruk, serta stunting di Provinsi Sumatera Selatan.

6. UCAPAN TERIMA KASIH

Penulis menyampaikan penghargaan dan terima kasih kepada Badan Pusat Statistik (BPS) Provinsi Sumatera Selatan yang telah menyediakan data resmi sebagai sumber utama dalam penelitian ini. Ucapan terima kasih juga ditujukan kepada Program Studi Teknik Informatika (Kampus Ogan Komering Ulu) D3 Universitas Mahakarya Asia atas dukungan akademik dan fasilitas yang diberikan selama proses penelitian. Penulis turut berterima kasih kepada para dosen, rekan sejawat, serta seluruh pihak yang telah memberikan bimbingan, masukan ilmiah, dan dorongan moral sehingga penelitian ini dapat terselesaikan dengan baik. Tidak lupa, apresiasi disampaikan kepada pihak-pihak lain yang tidak dapat disebutkan satu per satu, namun telah berkontribusi secara langsung maupun tidak langsung dalam kelancaran penelitian ini. Semoga segala bentuk bantuan yang diberikan menjadi amal dan bermanfaat bagi pengembangan ilmu pengetahuan, khususnya dalam bidang informatika dan analisis data kesehatan masyarakat.

7. DAFTAR PUSTAKA

- [1] UNICEF DATA, "Low birth weight," <https://www.who.int/data/nutrition/nlis/info/low-birth-weight>.
- [2] World Health Organization (WHO), "Low birth weight," <https://www.who.int/data/nutrition/nlis/info/low-birth-weight>.
- [3] Y. B. Okwaraji *et al.*, "National, regional, and global estimates of low birthweight in 2020, with trends from 2000: a systematic analysis," *The Lancet*, vol. 403, no. 10431, pp. 1071–1080, Mar. 2024, doi: 10.1016/S0140-6736(23)01198-4.
- [4] SSGI 2024, "SSGI 2024: Prevalensi Stunting Nasional Turun Menjadi 19,8%," <https://kemkes.go.id/id/ssgi-2024-prevalensi-stunting-nasional-turun-menjadi-198>.
- [5] A. Rahayu Permadi, S. Yunita Arini, P. Studi Kesehatan Masyarakat, F. Kesehatan Masyarakat, U. Airlangga, and D. Keselamatan dan Kesehatan Kerja, "FAKTOR RISIKO STUNTING PADA BALITA (0-5 TAHUN) DENGAN RIWAYAT BBLR DI DESA MAYANGREJO KABUPATEN BOJONEGORO," vol. 4, no. 3, 2023.
- [6] K-Means Cluster Analysis | Columbia Public Health, "K-Means Cluster Analysis | Columbia Public Health," <https://www.publichealth.columbia.edu/research/population-health-methods/k-means-cluster-analysis>.
- [7] Dayla May Cytry, S. Defit, and G. Nurcahyo, "Penerapan Metode K-Means dalam Klasterisasi Status Desa terhadap Keluarga Beresiko Stunting," *Jurnal KomtekInfo*, pp. 122–127, Sep. 2023, doi: 10.35134/komtekinfo.v10i3.423.
- [8] I. Pramudya, V. S. H. Sukmawati, Y. Wellyhans, and I. Darwati, "Implementasi Algoritma K-Means Pada Klasterisasi Tingkat Kasus Stunting Di Kabupaten Batang," *Jurnal Infortech*, vol. 6, no. 2, pp. 101–106, 2024.
- [9] C. Syalom, U. Khaira, and D. Arsa, "CLUSTERING DAERAH RAWAN STUNTING DI INDONESIA MENGGUNAKAN ALGORITMA K-MEANS," *Jurnal Teknologi Informasi*, vol. 6, no. 1, 2025, doi: 10.46576/djtechno.
- [10] Syahrul Khairati, Siti Maisyarah Fitri Siregar, Sri Wahyuni, and N. Nurhasanah, "Hubungan berat badan lahir rendah dengan kejadian stunting: Tinjauan Literatur," *Haga*

Journal of Public Health (HJPH), vol. 1, no. 3, pp. 105–112, Jul. 2024, doi: 10.62290/hjph.v1i3.30.

- [11] S. C. Sholihah, “HUBUNGAN BERAT BADAN LAHIR RENDAH (BBLR) TERHADAP KEJADIAN STUNTING DI WILAYAH KERJA PUSKESMAS DRADAH,” vol. 7, no. 1, 2023.
- [12] D. V. Sari, S. Damayanti, F. Apriani, A. Ardilla, and others, “BERAT BADAN LAHIR RENDAH DENGAN KEJADIAN STUNTING: TINJAUAN LITERATUR,” *Jurnal Kesehatan Akimal*, vol. 4, no. 1, pp. 75–81, 2025.
- [13] B. Waafiroh, E. Iriyani, and A. Sejati, “PENGARUH BERAT DAN PANJANG BADAN LAHIR RENDAH TERHADAP KEJADIAN STUNTING PADA ANAK USIA 1-5 TAHUN,” *Jurnal Ilmu Kebidanan*, vol. 10, no. 1, pp. 7–13, Dec. 2023, doi: 10.48092/jik.v10i1.216.
- [14] B. Safhira *et al.*, “The Association of Low Birth Weight and the Incidence of Stunting among Under-Fives in Indonesia: A Systematic Review,” *JIKM*, vol. 17, no. 1, pp. 44–51, 2025, doi: 10.52022/jikm.v17i1.744.
- [15] S. S. Nagari and L. Inayati, “IMPLEMENTATION OF CLUSTERING USING K-MEANS METHOD TO DETERMINE NUTRITIONAL STATUS,” *Jurnal Biometrika dan Kependudukan*, vol. 9, no. 1, pp. 62–68, Jun. 2020, doi: 10.20473/jbk.v9i1.2020.62-68.
- [16] B. M. Islami, C. Sukmayadi, and T. N. Padilah, “Clustering Fasilitas Kesehatan Berdasarkan Kecamatan Di Karawang Dengan Algoritma K-Means,” *Bina Insani ICT Journal*, vol. 8, no. 1, pp. 83–92, 2021.
- [17] BPS, “Jumlah Bayi Lahir, Berat Bayi Lahir Rendah, dan Bergizi Buruk - Tabel Statistik,” <https://sumsel.bps.go.id/id/statistics-table/2/MzcxIzI=/jumlah-bayi-lahir-berat-bayi-lahir-rendah-dan-bergizi-buruk.html>, 2025.