JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 13 No. 3, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v13i3.7145

IDENTIFIKASI WILAYAH KEKURANGAN ENERGI KRONIS (KEK) PADA IBU HAMIL DI KABUPATEN KARAWANG MENGGUNAKAN K-MEANS

Bagas Aqmal Febrianto[1], Mohamad Jajuli[2], Susilawati[3]

[1,2,3]Universitas Singaperbangsa Karawang, Kabupaten Karawang

Keywords:
Kekurangan Energi Kronis (KEK*), K-Means Clustering, Elbow Method, Silhouette Coefficient, QGIS,* Kabupaten Karawang

Corespondent Email:
2110631170056@student.unsika.ac.id

Abstrak. Kekurangan Energi Kronis (KEK) pada ibu hamil merupakan permasalahan kesehatan serius di Indonesia, termasuk di Kabupaten Karawang. Kondisi ini berkontribusi terhadap berbagai komplikasi kehamilan, seperti Berat Badan Lahir Rendah (BBLR) serta peningkatan risiko morbiditas dan mortalitas ibu dan bayi. Penelitian ini bertujuan untuk mengidentifikasi dan memetakan wilayah dengan risiko tinggi KEK di Kabupaten Karawang menggunakan pendekatan K-Means Clustering. Data yang digunakan berasal dari Dinas Kesehatan Kabupaten Karawang untuk periode Januari hingga September 2024, mencakup 12.402 data ibu hamil yang dikumpulkan dari 50 Puskesmas. Proses analisis dilakukan dengan pendekatan Knowledge Discovery in Database (KDD) yang terdiri dari tahapan integrasi data, pembersihan data, seleksi data, transformasi data, data mining, evaluasi, dan visualisasi hasil. Algoritma K-Means Clustering diterapkan untuk mengelompokkan wilayah berdasarkan faktor risiko KEK, meliputi umur ibu, indeks massa tubuh (IMT), jarak kehamilan, paritas, dan status risiko kehamilan. Penentuan jumlah klaster optimal dilakukan menggunakan metode Elbow Method, yang menghasilkan nilai K = 2. Evaluasi klaster dilakukan menggunakan Silhouette Coefficient, dengan nilai 0.67, menunjukkan struktur klaster yang cukup baik. Hasil analisis menunjukkan bahwa terdapat lima wilayah dengan risiko tinggi KEK, yaitu Cikampek, Karawang Barat, Karawang Timur, Klari, dan Teluk Jambe Timur, sedangkan 25 wilayah lainnya dikategorikan sebagai tidak berisiko.

Abstract. *Chronic Energy Deficiency (CED) in pregnant women is a serious health problem in Indonesia, including in Karawang. This condition contributes to various pregnancy complications, such as Low Birth Weight (LBW) and an increased risk of morbidity and mortality for mothers and babies. This study aims to identify and map areas at high risk of CED in Karawang using the K-Means Clustering approach. The data used came from the Karawang District Office for the period January to September 2024, covering 12,402 data on pregnant women collected from 50 community health centers. The analysis process was carried out using the Knowledge Discovery in Database (KDD) approach which consists of the stages of data integration, data cleaning, data selection, data transformation, data mining, evaluation, and visualization of results. The K-Means Clustering algorithm is applied to group regions based on the risk factors of KEK, including the age of the mother, body mass index (BMI), pregnancy spacing, parity, and pregnancy risk status. The determination of the optimal number of clusters is carried out using the Elbow Method, which results in a value of K = 2. The evaluation of the cluster is carried out using the Silhouette Coefficient, with a value of 0.67, indicating a fairly good cluster structure. The results of the analysis show that there are five areas with a high risk of CED, namely Cikampek, Karawang*

Barat, Karawang Timur, Klari, and Teluk Jambe Timur, while the other 25 areas are categorized as not at risk.

1. PENDAHULUAN

Masalah gizi menjadi salah satu aspek terpenting dalam tumbuh kembang seseorang, terutama bagi ibu hamil yang memerlukan asupan nutrisi seimbang untuk mendukung kesehatan dirinya dan janin yang dikandung. Salah satu masalah gizi yang sering terjadi pada ibu hamil adalah Kekurangan Energi Kronis (KEK). KEK merupakan kondisi di mana ibu hamil mengalami kekurangan asupan makanan dalam jangka waktu yang lama atau bersifat menahun, yang dapat menyebabkan gangguan kesehatan pada ibu hamil maupun wanita usia subur (WUS) [1]. Kondisi KEK dapat memengaruhi proses kehamilan, meningkatkan risiko komplikasi saat persalinan, serta berdampak negatif pada pertumbuhan dan perkembangan janin. Ibu hamil dengan KEK berisiko lebih tinggi mengalami kematian, baik pada diri mereka sendiri maupun pada bayi yang dilahirkan. Selain risiko kematian, ibu hamil yang menderita KEK juga berisiko melahirkan bayi dengan Berat Badan Lahir Rendah (BBLR), yang dapat memicu masalah kesehatan serius bagi bayi di kemudian hari. Berdasarkan data dari Riset Kesehatan Dasar (Riskesdas) tahun 2018, prevalensi KEK pada ibu hamil di Indonesia mencapai 17,3% [2]. Tingginya angka ini mengindikasikan bahwa KEK menjadi salah satu masalah gizi yang mendesak untuk segera diatasi. KEK juga berkaitan langsung dengan prevalensi stunting, yang merupakan salah satu dampak gizi buruk yang dialami oleh ibu hamil selama kehamilan. KEK menjadi salah satu penyebab Angka Kematian Ibu (AKI) dan bayi dengan berat badan lahir rendah (BBLR) yang tergolong tinggi di Indonesia yang berada di urutan keempat dengan prevalensi KEK pada ibu hamil sebesar 35.5% [3]. Untuk menanggulangi masalah ini, Kementerian Kesehatan telah menetapkan beberapa sasaran strategis dalam Rencana Strategis (Renstra) Kementerian Kesehatan tahun 2020-2024, termasuk menurunkan prevalensi KEK pada ibu hamil sebesar 10% [2]. Berdasarkan laporan rutin tahun 2020, tercatat bahwa dari 4,6 juta ibu hamil di 34 provinsi di Indonesia, sebanyak 451.350 di antaranya memiliki Lingkar Lengan Atas (LiLA) < 23,5 cm, yang merupakan indikator utama risiko KEK [2]. Berdasarkan data tersebut, terlihat bahwa prevalensi KEK pada ibu hamil di Indonesia masih cukup tinggi, dengan didapatkannya beberapa wilayah di Indonesia yang mencatatkan angka KEK yang cukup mengkhawatirkan. Berdasarkan data pada OpenDataJabar pada tahun 2020 terkait jumlah ibu hamil KEK di Jawa Barat, Kabupaten Karawang menempati posisi ke-8 dari 27 kabupaten atau kota yang ada di Jawa Barat dengan total kasus 3.297 ibu hamil KEK. Jika melihat dari segi dampak dari ibu hamil KEK itu sendiri, berdasarkan Survei Status Gizi Indonesia (SSGI) yang dilakukan oleh Kemenkes, didapatkan bahwa prevalensi balita stunting di Jawa Barat pada tahun 2022 mencapai 20,2%. Dalam target global yang ditetapkan oleh WHO, suatu wilayah dapat dikatakan memenuhi target global jika angka stunting berada di bawah 20%. Meskipun demikian, dilansir dari data prevalensi balita stunting pada Provinsi Jawa Barat tahun 2022, Kabupaten Karawang mencatat angka stunting sebesar 14%, yang berarti masih lebih rendah dibandingkan rata-rata provinsi. Namun, meskipun angka stunting di Karawang relatif lebih rendah, tingginya jumlah kasus ibu hamil KEK menunjukkan bahwa penanganan gizi selama kehamilan tetap menjadi prioritas. Hal ini bertujuan untuk mencegah dampak jangka panjang seperti peningkatan risiko stunting pada balita. Pemerintah Kabupaten Karawang juga terus berupaya menangani permasalahan ini melalui berbagai program, termasuk Pemberian Makanan Tambahan (PMT) serta pelaksanaan survei SSGI. Kepala Bidang Pemerintah Pembangunan Manusia, Bapak Dadan Nurdiansyah, mengungkapkan bahwa survei ini dilakukan guna mendukung pencapaian target penurunan angka stunting dalam RPJMD sebesar 12,1%. Selain itu, berdasarkan data terbaru, jumlah ibu hamil KEK pada periode Januari hingga September 2024 tercatat sebanyak 1.475 kasus, yang berarti setengah dari angka tahun 2020 (3.297 kasus).

Berdasarkan permasalahan tersebut, perkembangan pengetahuan dalam bidang teknologi saat ini dapat memberikan wawasan dalam menghadapi permasalahan tersebut, salah satunya melalui Data Mining. Data mining merupakan suatu proses untuk mencari atau menemukan pola atau informasi berguna dari data yang tersedia menggunakan metode tertentu [4]. Dengan menggunakan data mining, pengelompokan wilayah-wilayah rawan KEK dapat dilakukan dengan lebih efektif, sehingga penanganan yang tepat bisa diterapkan di daerah-daerah yang membutuhkan. Salah satu algoritma yang dapat diterapkan dalam proses pengelompokan ini adalah clustering, yang bertujuan untuk membagi data ke dalam kelompok berdasarkan karakteristik tertentu. Hal ini sesuai dengan salah satu faktor keberhasilan dalam mengurangi angka prevalensi ibu hamil KEK berdasarkan Laporan Akuntabilitas Kinerja Kegiatan Pembinaan Gizi Masyarakat (2020), bahwa perlunya penguatan dalam surveilans gizi dengan meningkatkan deteksi dini masalah gizi pada ibu hamil, guna diperlukannya tindak lanjut oleh pelayanan gizi dalam memperbaiki kondisi gizi seorang ibu hamil tersebut. Surveilans gizi adalah kegiatan yang dengan cara memantau serta mengevaluasi pelaksanaan kegiatan program gizi, yang bertujuan mengumpulkan, menganalisis serta menyebarkan suatu informasi terkait dengan status gizi individu ataupun populasi tertentu [2]. Dengan demikian, penelitian ini sejalan dengan faktor keberhasilan dalam mengurangi prevalensi KEK, yang memberikan informasi terkait dengan identifikasi dini wilayah-wilayah prioritas rawan KEK. Informasi ini sangat penting untuk mendukung pengambilan keputusan dalam perencanaan intervensi gizi yang lebih tepat sasaran.

Penelitian terkait permasalahan status gizi ibu hamil telah dilakukan sebelumnya oleh [4], yang bertujuan untuk memudahkan pihak instansi dalam mengelompokkan data dengan cepat. Penelitian ini menggunakan data dari puskesmas di Kota Datar dan menerapkan algoritma K-Means dengan metode KDD (Knowledge Discovery in Database). Hasil penelitian menunjukkan terbentuknya tiga klaster berdasarkan status gizi ibu hamil. Selain itu, penelitian serupa juga dilakukan oleh Fadzly Maulana Hidayat et al. (2024) yang memfokuskan pada penggunaan algoritma K-Means dan K-Medoids untuk mengelompokkan wilayah di Jawa Barat berdasarkan permasalahan gizi buruk. Pada kedua algoritma tersebut, metode elbow digunakan untuk menentukan jumlah optimal cluster, dengan nilai k=2 sebagai titik curam yang ditentukan. Penelitian ini menghasilkan tiga klaster: klaster 1 mencakup wilayah dengan tingkat gizi buruk tinggi, klaster 2 dengan tingkat sedang, dan klaster 3 dengan tingkat rendah. Dari hasil perbandingan, algoritma K-Means terbukti memiliki performa terbaik dengan nilai Silhouette Coefficient sebesar 0.617. Penelitian oleh Neneng Sayuti Hanapiah et al. (2023) menggunakan algoritma K-Means untuk mengelompokkan kasus pemberian makanan tambahan pada anak yang mengalami stunting. Metode yang digunakan adalah KDD (Knowledge Discovery in Database) untuk memastikan bahwa setiap tahap dilakukan secara sistematis. Beberapa variabel yang digunakan dalam penelitian ini mencakup jenis makanan tambahan, status gizi, dan usia anak yang berperan dalam perkembangan kondisi stunting. Penelitian ini menghasilkan nilai Davies Bouldin Index (DBI) sebesar 0.489 dengan k=3, menunjukkan bahwa pembagian klaster valid. Terdapat tiga klaster yang dihasilkan, yaitu cluster 1, cluster 2, dan cluster 3. Selain itu, penelitian Maulana Muhammad et al. (2023) melakukan perbandingan antara K-Means dengan K-Medoids dalam mengklasifikasikan status gizi anak, dan hasilnya menunjukkan bahwa K-Means lebih unggul dalam hal accuracy, specificity, dan sensitivity pada dataset yang diuji. Berdasarkan penelitian-penelitian ini, dapat disimpulkan bahwa algoritma clustering dalam data mining memiliki potensi besar dalam mengidentifikasi wilayah-wilayah rawan KEK, terutama di Jawa Barat.

2. TINJAUAN PUSTAKA

2.1. Gizi dan Ibu Hamil

Gizi adalah proses di mana suatu organisme memanfaatkan makanan yang dikonsumsinya melalui berbagai tahapan, seperti penyerapan, pengangkutan, penyimpanan, metabolisme, serta pembuangan zat-zat yang tidak lagi

diperlukan untuk menjaga energi [6]. Gizi memegang peran penting dalam pertumbuhan dan perkembangan seseorang, terutama pada fase pertumbuhan yang sangat rentan terhadap pengaruh lingkungan. Kondisi ini dapat berkontribusi pada perubahan status gizi seseorang. Status gizi merupakan kondisi yang ditentukan oleh kebutuhan fisik terhadap energi atau zat gizi yang diperoleh dari asupan makanan, yang dipengaruhi oleh berbagai faktor, termasuk lingkungan [7]. Salah satu permasalahan gizi yang dihadapi di Indonesia adalah gizi selama masa kehamilan [8]. Status gizi sangat berpengaruh terhadap tumbuh kembang ibu hamil, di mana asupan zat gizi yang dikonsumsi oleh ibu hamil berdampak pada perkembangan janin yang dikandung.

2.2. Kekurangan Energi Kronis (KEK)

Kekurangan Energi Kronis (KEK) merupakan kondisi yang dialami oleh ibu hamil akibat kekurangan asupan protein dan energi selama kehamilan, yang berdampak negatif pada kesehatan ibu dan janin yang dikandungnya [8]. Ibu hamil yang mengalami KEK memiliki risiko lebih tinggi melahirkan bayi dengan Berat Badan Lahir Rendah (BBLR), menghadapi komplikasi seperti kematian saat persalinan, serta mengalami pendarahan berlebih. Selain itu, bayi yang lahir dari ibu dengan KEK lebih rentan terhadap infeksi, risiko keguguran (*abortus*), dan gangguan pada perkembangan otak janin. Kondisi ini dapat mempengaruhi kesehatan bayi dalam jangka panjang [2]).

2.3. Data Mining

Data mining adalah metode yang memungkinkan pengguna untuk mengakses data dalam jumlah besar dengan cepat. Teknik ini digunakan untuk mengungkap pola-pola tersembunyi dalam data yang telah dikumpulkan sebelumnya. Secara khusus, data mining diartikan sebagai alat atau aplikasi yang menerapkan analisis pada data [9] *Data mining* juga dikenal sebagai proses analitis yang digunakan untuk mengeksplorasi data dalam jumlah besar guna memperoleh informasi atau pengetahuan yang berharga [10].

2.4. Clustering

Clustering adalah teknik dalam *data mining* yang bertujuan untuk mengelompokkan objek atau data ke dalam beberapa *cluster*, sehingga data yang memiliki kesamaan dapat ditempatkan dalam *cluster* yang sama. Pengelompokan ini dilakukan dengan tujuan untuk meminimalkan variasi di dalam satu *cluster* dan memaksimalkan perbedaan antar *cluster*. Algoritma *clustering* merupakan teknik yang digunakan suatu perusahaan dalam melakukan segmentasi pelanggan, dengan tujuan meningkatkan penjualan. Algoritma ini mengelompokkan populasi dengan karakteristik serupa ke dalam beberapa kelompok kecil [11].

2.5. K-Means Clustering

Algoritma K-Means adalah metode klasterisasi data non hirarki yang menggunakan pengukuran jarak untuk membagi data menjadi beberapa kelompok (cluster) berdasarkan atribut numerik. Algoritma ini mengelompokkan data ke dalam cluster sedemikian rupa sehingga data dengan karakteristik serupa ditempatkan dalam cluster yang sama, sedangkan data dengan karakteristik yang berbeda dikelompokkan dalam cluster yang lain [12]. K-Means tergolong dalam algoritma partitional karena prosesnya didasarkan pada penentuan awal jumlah kelompok dengan mendefinisikan centroid awal. Untuk menjalankannya, algoritma ini memerlukan jumlah cluster awal sebagai input dan menghasilkan jumlah cluster akhir sebagai output [20].

2.6. Elbow Method

Tujuan dari metode elbow adalah untuk menentukan nilai k yang paling optimal dengan cara meminimalkan within-cluster sum of squares (withinss). Jumlah cluster optimal ditentukan dengan mengamati grafik yang menunjukkan titik di mana terbentuk sudut tajam atau "siku." Semakin besar nilai k (jumlah cluster), nilai SSE (Sum of Squared Errors) akan terus berkurang secara signifikan [8].

2.7. Silhouette Coefficient

Silhouette Coefficient adalah metode yang digunakan untuk mengukur seberapa baik kualitas pembentukan sebuah cluster. Metode ini mengkombinasikan dua konsep, yaitu kohesi dan separation. Kohesi mengukur kedekatan antar objek dalam satu cluster, sedangkan separation mengukur seberapa jauh cluster tersebut dipisahkan dari cluster lainnya. Semakin tinggi nilai kohesi dan separation, semakin baik kualitas cluster yang terbentuk [13].

2.8. Geographic Information System (GIS)

Geographic Information System (GIS) adalah perubahan dalam memahami atau mengelola suatu informasi geografis. Yang memiliki kemampuan untuk menggabungkan, mengolah, menganalisis, menggambarkan atau memvisualisasikan data yang georeferensi atau data yang berhubungan dengan lokasi di permukaan bumi. SIG melakukan penggabungan dari beberapa prinsip, seperti kartografi, statistik ataupun komputasi [13].

2.9. Inisialisasi Faktor Penyebab

Inisialisasi faktor penyebab diperoleh dari hasil tinjauan literatur mengenai Kekurangan Energi Kronis (KEK) pada ibu hamil. Faktor-faktor ini mencakup berbagai penyebab yang dapat mempengaruhi terjadinya kejadian KEK pada ibu hamil. Rincian faktor penyebab yang telah diidentifikasi dapat dilihat sebagai berikut :

1. Faktor Umur

Usia ibu hamil mempengaruhi risiko KEK, dengan ibu <20 tahun dan >35 tahun berisiko lebih tinggi mengalami KEK dibandingkan dengan usia 20-35 tahun [14][15].

2. Faktor Riskan

Penyakit seperti anemia meningkatkan risiko KEK pada ibu hamil, karena mengganggu kemampuan tubuh untuk menyerap energi [16].

3. Faktor Indeks Massa Tubuh (IMT)

IMT yang rendah (kurang dari 18,5) meningkatkan risiko KEK dan Berat Badan Lahir Rendah (BBLR) pada bayi [17].

4. Faktor Paritas

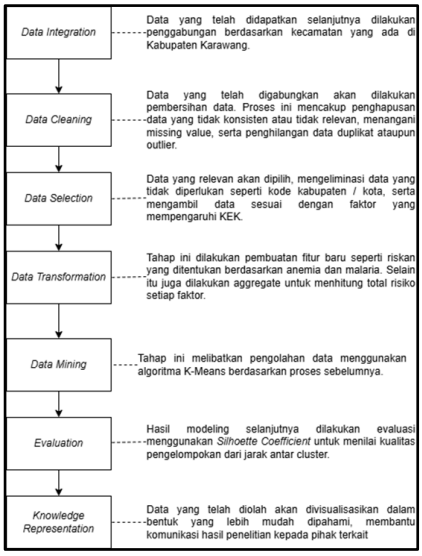
Paritas tinggi dan jarak kehamilan yang dekat meningkatkan risiko KEK pada ibu hamil. Wanita dengan paritas tinggi memiliki risiko 25,51 kali lebih besar mengalami KEK[18].

5. Faktor Jarak Kehamilan

Jarak kehamilan yang terlalu dekat meningkatkan risiko KEK karena ibu tidak memiliki cukup waktu untuk pemulihan tubuh [19].

3. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini mengikuti tahapan *Knowledge Discovery in Database* (KDD) sebagai pendekatan utama dalam proses *data mining*. Berikut adalah tahapan dari penelitian ini.



Gambar 1. Metodelogi Penelitian

3.1. Rancangan Penelitian

Metode yang digunakan dalam penelitian ini adalah Knowledge Discovery in Database (KDD). Berikut adalah rancangan yang dilakukan pada proses data mining:

3.1.1. Data Integration

Data yang diperoleh berjumlah 50 file dalam format Excel, masing-masing mewakili data ibu hamil dari berbagai wilayah. Selanjutnya, seluruh file tersebut akan digabungkan menjadi satu kesatuan dataset. Proses penggabungan ini bertujuan untuk mengelompokkan data berdasarkan wilayah administratif, khususnya pada tingkat kecamatan. Langkah ini dilakukan untuk mempermudah proses eksplorasi dan analisis data pada tahap selanjutnya.

3.1.2. Data Cleaning

Pada tahap ini, data yang telah dikumpulkan akan dilakukan pembersihan data. Proses ini mencakup penghapusan data yang tidak valid, penanganan missing value menggunakan teknik imputasi, penghilangan nilai duplikat, serta

identifikasi outlier menggunakan IQR (Interquartile Range) dan penanganan outlier menggunakan teknik winsorizing. Dengan demikian, data yang tersisa akan lebih siap digunakan untuk tahap selanjutnya.

3.1.3. Data Selection

Tahap ini merupakan proses pemilihan data yang relevan untuk digunakan, sehingga tidak semua data akan diikutsertakan. Salah satu contohnya adalah nama ibu hamil. Proses seleksi atribut ini dilakukan untuk memilih data yang sesuai dengan analisis yang akan dilakukan.

3.1.4. Data Transformation

Tahap ini mencakup proses pembuatan fitur baru, seperti menambahkan kolom riskan, yang ditentukan berdasarkan kasus anemia dan malaria. Selain itu, tahap ini juga melibatkan proses agregasi berdasarkan jumlah kasus untuk setiap resiko.

3.1.5. Data Mining

Tahap Data Mining melibatkan pengolahan data berdasarkan langkah-langkah yang telah dilakukan sebelumnya, menggunakan algoritma K-Means. Sebelum algoritma ini diterapkan, nilai k (jumlah cluster) ditentukan menggunakan metode Elbow. Metode ini menghitung Sum of Squared Errors (SSE) untuk setiap nilai k. Prosesnya dimulai dari k=2, kemudian nilai k ditingkatkan secara bertahap. Hasil SSE untuk setiap k di-plot pada grafik, dan titik Elbow di mana penurunan SSE mulai melambat dan membentuk "sudut siku tajam" dipilih sebagai nilai k yang optimal.

Setelah nilai k ditentukan, algoritma K-Means diterapkan. Pada tahap awal, centroid awal untuk setiap cluster dipilih secara acak. Selanjutnya, setiap data dalam dataset dihitung jaraknya terhadap centroid menggunakan rumus Jarak Euclidean dengan mempertimbangkan jarak antara data dan centroid pada setiap variabel. Data kemudian diklasifikasikan ke dalam cluster yang memiliki jarak terdekat dari centroid. Setelah pengelompokan awal, centroid diperbarui dengan menghitung rata-rata nilai dalam setiap cluster. Proses penghitungan jarak dan pembaruan centroid ini diulang hingga tidak ada perubahan signifikan pada anggota cluster, atau hingga nilai centroid menjadi stabil.

3.1.6. Evaluation

Setelah menyelesaikan proses data mining, langkah berikutnya adalah melakukan evaluasi terhadap hasil yang diperoleh dengan menggunakan indeks validasi Silhouette. Dengan metode ini, kita dapat menilai seberapa baik pengelompokan yang telah dilakukan oleh algoritma dan mengukur seberapa terpisah setiap cluster satu sama lain.

3.1.7. Knowledge Presentation

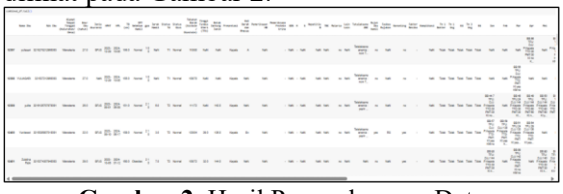
Pada tahap akhir, data yang telah diolah akan disajikan atau divisualisasikan dalam bentuk yang lebih mudah dipahami. Visualisasi ini berfungsi untuk mempermudah komunikasi hasil penelitian kepada pihak-pihak terkait, sehingga informasi yang disampaikan dapat dipahami dengan lebih baik.

4. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data rekam medis ibu hamil dari Dinas Kesehatan Kabupaten Karawang, yang mencakup 50 dataset dari masing-masing puskesmas, dengan total sekitar 12.000 data dan 70 atribut, yang dikumpulkan antara Januari hingga September 2024. Algoritma K-Means diterapkan untuk mengklasterisasi data dan menganalisis persebaran wilayah yang berpotensi mengalami kekurangan energi kronis (KEK). Hasil klasterisasi ini diharapkan dapat mengidentifikasi wilayah dengan risiko KEK tinggi sebagai dasar perencanaan intervensi kesehatan yang lebih tepat sasaran.

4.1 Data Integration

Data Integration merupakan tahap di mana dilakukan penggabungan data untuk membentuk satu dataset yang lebih terstruktur dan siap digunakan dalam proses analisis. Setiap dataset berisi rekaman data ibu hamil dari masing-masing Puskesmas yang ada di Kabupaten Karawang, dengan total sebanyak 50 Puskesmas atau setara dengan 50 file data berformat Excel. Hasil penggabungan ini dapat dilihat pada Gambar 2.



Gambar 2. Hasil Penggabungan Data

4.2 Data Cleaning

Pada tahap ini selanjutnya data yang sudah digabungkan akan dilakukan pengecekan kualitas dan perbaikan data.

Hal ini dilakukan agar memastikan bahwa data memiliki kualitas yang bagus dan dapat digunakan untuk analisis lebih lanjut. Langkah-langkah yang dilakukan mencakup identifikasi serta penghapusan data duplikat guna menghindari bias, penanganan data kosong dengan metode imputasi yang sesuai, serta perbaikan label yang tidak relevan agar lebih konsisten. Selain itu, dilakukan deteksi serta penanganan outlier menggunakan metode statistik seperti IQR atau *Z-score* untuk mengidentifikasi nilai ekstrim yang dapat mempengaruhi hasil analisis. Dengan serangkaian proses ini, data yang digunakan menjadi lebih akurat, bersih, dan siap untuk diolah pada tahap selanjutnya.

```
combined_df.describe()
```

	Umur Ibu (tahun)	TB (cm)	Jarak Hamil	Tinggi Fundus Uteri (TFU)	Detak Jantung Janin	Pemeriksaan HB	Des	Tatalaksana PNC	Keterangan
count	11979.000000	11214.000000	4010.000000	3816.000000	4715.000000	2626.000000	0.0	0.0	0.0
mean	29.719927	157.150089	7.758105	25.540304	144.652810	12.064699	NaN	NaN	NaN
std	57.948637	153.406040	63.417260	56.611245	189.553071	23.478451	NaN	NaN	NaN
min	0.000000	-1168.000000	0.000000	0.000000	-7.000000	-10.000000	NaN	NaN	NaN
25%	23.000000	151.000000	3.000000	20.000000	138.000000	11.000000	NaN	NaN	NaN
50%	27.000000	155.000000	5.000000	26.000000	142.000000	11.500000	NaN	NaN	NaN
75%	32.000000	158.000000	8.000000	30.000000	146.000000	12.000000	NaN	NaN	NaN
max	2023.000000	15668.000000	2020.000000	2635.000000	13080.000000	1208.000000	NaN	NaN	NaN

Gambar 3. Analisis Statistik

Berdasarkan analisis statistik pada Gambar 3 menggunakan fungsi *describe*(), tampak terdapat nilai yang tidak masuk akal di beberapa kolom.Hal ini dapat disebabkan oleh kesalahan input data atau *human error.*

4.2.1 Menangani Data Duplikat

Pengecekan nilai duplikat pada data berdasarkan pada kolom “Nik Ibu”, nilai yang tidak unik atau dapat dikatakan duplikat akan dilakukan pengecekan dengan fungsi *duplicated()*. Setelah dilakukan pengecekan terhadap data duplikat, langkah selanjutnya adalah proses pembersihan data dengan menggunakan fungsi drop_duplicates().

4.2.2 Menangani Missing Value

Setelah penanganan data duplikat, data yang dianggap valid sebagai perwakilan data atau dapat dikatakan unik selanjutnya akan dilakukan pengecekan dan penanganan missing value atau nilai yang bersifat NaN (*Not a Number*). Pengecekan ini dilakukan pada beberapa kolom yang memiliki peran penting dalam analisis, terutama kolom yang berkaitan dengan indikator penyebab KEK

4.2.3 Pemeriksaan Outlier

Pada tahap ini pemeriksaan dan penanganan outlier dilakukan, mengingat bahwa algoritma K-Means sangat sensitif terhadap outlier. Penanganan outlier dilakukan pada tabel bertipe numerik, seperti kolom “Umur Ibu (Tahun)”, “Jarak Hamil”, dan “Pemeriksaan HB”.

4.1 Data Selection.

Pada tahap data selection, Pemilihan data dilakukan pada data yang relevan, dalam artian tidak semua data atau atribut digunakan. Atribut yang digunakan berdasarkan hasil Literatur Review yang dilakukan oleh Nurannisa Fitria (2021) terkait Faktor-faktor yang berhubungan dengan Kekurangan Energi Kronik pada Ibu Hamil, antara lain Paritas, Jarak Kehamilan, dan IMT yang menjadi variabel paling berpengaruh, sedangkan variabel umur, anemia dan pekerjaan sebagai variabel perancu. Data yang tidak digunakan atau dipilih akan dilakukan penghapusan, sehingga hasil data yang sudah diseleksi dapat dilihat pada Gambar 4.

```
df_clean.head(100)
```

	Umur Ibu (tahun)	IMT Sebelum Hamil	Jarak Hamil	Pemeriksaan HB	Malaria	Kecamatan	Partus
0	22.0	Normal	5.11646	11.5758	0	Karawang Barat	0
3	28.0	Normal	1.00000	11.5758	0	Karawang Barat	2
4	33.0	Normal	5.11646	11.5758	0	Karawang Barat	2
5	26.0	Normal	5.11646	11.5758	0	Karawang Barat	0
6	27.0	Normal	5.11646	11.5758	0	Karawang Barat	0
...	...	...	...	...	...	...	...
97	24.0	Normal	5.11646	11.5758	0	Karawang Barat	0
98	33.0	Gemuk	5.11646	11.0000	0	Karawang Barat	2
99	28.0	Normal	0.00000	11.5758	0	Karawang Barat	0
100	26.0	Normal	5.11646	11.5758	0	Karawang Barat	0
101	30.0	Gemuk	5.00000	11.5758	0	Karawang Barat	1

100 rows × 7 columns

Gambar 4. Hasil Data Selection

4.2 Data Transformation

Pada tahap ini dilakukan perubahan terhadap bentuk pada data yang sudah dipilih sebelumnya, perubahan bentuk ini berdasarkan pada nilai indikator yang menyebabkan seorang ibu hamil dapat dikatakan beresiko, aturan perubahan didasarkan pada Bab 2.9 tentang Faktor Penyebab.

Gambar 5. Hasil Data Transformation

Data yang telah diidentifikasi sebagai berisiko selanjutnya akan dianalisis lebih lanjut dengan melakukan agregasi menggunakan fungsi *agg*(). Proses ini bertujuan untuk menghitung jumlah kasus berisiko berdasarkan setiap desa. Hasil

agregasi dan perhitungan jumlah kasus berisiko dapat dilihat pada Gambar 6.

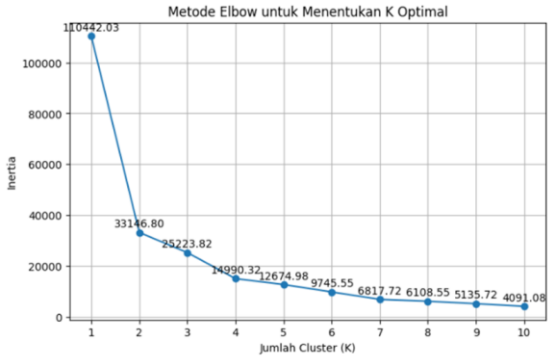
	Kecamatan	Risiko Umur	Risiko IMT	Risiko Paritas	Risiko Jarak Hamil	Riskan
0	Banyusari	16	16	3	31	15
1	Batujaya	29	35	7	72	21
2	Ciampel	18	24	5	18	9
3	Cibuaya	13	10	2	34	0
4	Cikampek	28	49	17	127	39

Gambar 6. Hasil *Data Aggregation*

4.3 Data Mining

Pada tahap data mining, akan dilakukan pemrosesan data guna menemukan pola atau informasi dengan menggunakan suatu teknik tertentu. Teknik ini memberikan informasi untuk menggolongkan daerah yang memiliki potensi beresiko KEK. Pada penelitian ini, teknik yang digunakan adalah clustering dengan menggunakan K-Means.

4.3.1 Metode Elbow



Gambar 7. Pencarian Nilai K menggunakan Metode Elbow

Berdasarkan Gambar 7, dapat dilihat bahwa sudut yang membentuk siku berada pada nilai K = 2. Penentuan nilai K = 2 sebagai nilai paling optimal didasarkan pada titik di mana inertia mengalami penurunan signifikan sebelum akhirnya melambat. Hal ini menunjukkan bahwa setelah K = 2, penurunan inertia tidak lagi terlalu drastis, sehingga menambah jumlah klaster tidak memberikan perbedaan yang signifikan dalam segmentasi data.

4.3.2 K-Means

Setelah nilai K optimal diperoleh melalui metode Elbow, langkah selanjutnya adalah melakukan clustering menggunakan algoritma K-Means. Proses ini menggunakan modul KMeans dari library Scikit-Learn untuk mengelompokkan data berdasarkan karakteristik yang telah ditentukan. Algoritma K-Means akan mengelompokkan data ke dalam K klaster, di mana setiap data akan dikelompokkan ke cluster dengan centroid terdekat

Tabel 1. Hasil *Clustering*

Cluster	Kecamatan	Risiko Umur	Risiko IMT	Risiko Paritas	Risiko Jarak Kehamilan	Riskan
0	Cikampek	28.0	49.0	17.0	127.0	39.0
	Karawang Barat	48.0	66.0	16.0	153.0	34.0
	Karawang Timur	20.0	43.0	7.0	192.0	11.0
	Klari	26.0	71.0	11.0	252.0	91.0
	Teluk Jambe Timur	22.0	72.0	15.0	150.0	51.0
1	Banyusari	16.0	16.0	3.0	31.0	15.0
	Batujaya	29.0	35.0	7.0	72.0	21.0
	Ciampel	18.0	24.0	5.0	18.0	9.0
	…	…	…	…	...	…

	Tirtajaya	36.0	32.0	5.0	41.0	45.0

Hasil pemrosesan K-Means ini dapat dilihat pada Tabel 4.4, yang menunjukkan distribusi data ke dalam masing-masing klaster berdasarkan fitur-fitur yang dianalisis.

4.4 Evaluation

Setelah tahap data mining selesai dilakukan, langkah selanjutnya adalah evaluasi cluster guna mengetahui kualitas dari klaster yang terbentuk. Pada penelitian ini, evaluasi cluster dilakukan menggunakan Silhouette Coefficient, yang bertujuan untuk mengukur seberapa baik setiap data dalam cluster dikelompokkan dan seberapa jelas pemisahan antar klaster. Evaluasi dilakukan dengan memanfaatkan modul silhouette_score dari library Scikit-Learn, yang akan menghitung nilai Silhouette Coefficient berdasarkan jarak rata-rata antara satu data dengan data lain dalam klaster yang sama, serta jarak dengan cluster terdekat. Hasil evaluasi cluster menggunakan Silhouette Coefficient dengan nilai k=2 dapat dilihat pada Gambar 4.29.

```
[ ] from sklearn.metrics import silhouette_score

    # Menghitung Silhouette Score untuk mengevaluasi hasil clustering
    silhouette_avg = silhouette_score(X, agg_kec['Cluster'])
    print(f'Silhouette Score: {silhouette_avg}')

    Silhouette Score: 0.6717267293371431
```

Gambar 8. Hasil Evaluasi menggunakan *Silhoette Coeffisien*

Gambar di atas menunjukkan bahwa hasil evaluasi dengan menggunakan Silhouette Coefficient pada nilai k=2 menghasilkan skor sebesar 0.68, nilai ini menunjukkan bahwa cluster yang terbentuk memiliki struktur yang cukup baik dan dapat diandalkan dalam mengelompokkan data. Skor Silhouette Coefficient yang mendekati 1 menunjukkan bahwa objek dalam satu cluster lebih mirip satu sama lain dibandingkan dengan objek di cluster lain.

4.5 Knowledge Presentation

Pada tahap *Knowledge Presentation*, dilakukan pemetaan hasil dari pengolahan data yang telah dilakukan sebelumnya. Berdasarkan Tabel 4.4, diperoleh hasil bahwa jumlah cluster yang terbentuk adalah dua, yaitu Cluster 0 dan Cluster 1. Kedua cluster ini mengelompokkan daerah ke dalam kategori berisiko dan tidak berisiko berdasarkan karakteristik tertentu. Untuk menentukan cluster mana yang termasuk kategori berisiko, dilakukan perhitungan dengan menganalisis rata-rata nilai risiko dari setiap cluster.

Tabel 2. Rata-Rata Dari Setiap Cluster Berdasarkan Indikator Penyebab KEK

Cluster	Risiko Umur	Risiko IMT	Risiko Paritas	Risiko Jarak Hamil	Riskan
0	28,8	60,2	13,2	140	45,2
1	22	25,0	25.04	34.04	14,08

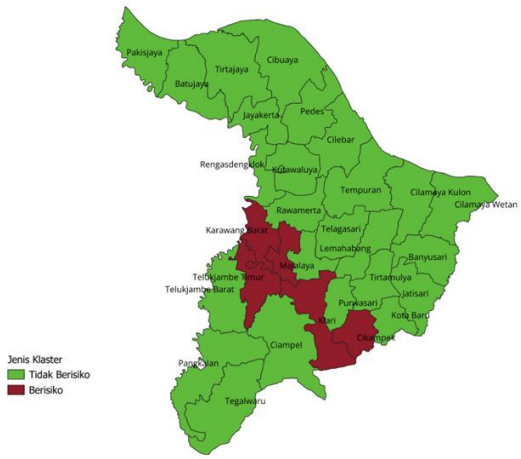
Dari Tabel diatas, dapat disimpulkan bahwa cluster dengan rata-rata risiko lebih tinggi dikategorikan sebagai daerah berisiko, sedangkan cluster dengan rata-rata risiko lebih rendah dikategorikan sebagai daerah tidak berisiko. Daerah hasil klasterisasi dapat dilihat pada Tabel 4.6.

Tabel 3. Pelabelan Cluster

Nama Kecamatan	
Beresiko	**Tidak Beresiko**
Cikampek	Banyusari
Karawang Barat	Batujaya
Karawang Timur	Ciampel
Klari	Cibuaya
Teluk Jambe Timur	Cikampek Utara

	Cilamaya Kulon
	Cilamaya Wetan
	Dan kecamatan lainnya.

Setelah label pada *clsuster* sudah ditentukan, selanjutnya adalah proses dilakukannya visualisasi pemetaan dengan menggunakan *tools* QGIS. Berdasarkan pada Gambar 4.30 daerah yang dinyatakan tidak berisiko KEK ditandai dengan warna hijau, sedangkan daerah berisiko KEK diwarnai dengan warna merah.



Gambar 9. Hasil Pemetaan Daerah Risiko dan Tidak Berisiko KEK Kabupaten Karawang

KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan sebelumnya, didapatkan kesimpulan sebagai berikut :

a. Penerapan algoritma K-Means dalam mengidentifikasi wilayah berisiko KEK di Kabupaten Karawang dilakukan dengan menggunakan dataset yang diperoleh dari Dinas Kesehatan Karawang. Dataset ini berupa rekam data ibu hamil dari periode Januari hingga September yang dikumpulkan dari 50 Puskesmas di Kabupaten Karawang. Data dari masing-masing Puskesmas kemudian digabungkan menjadi satu kesatuan, menghasilkan total 12.402 data rekam. Setelah proses penggabungan, data diolah menggunakan metode KDD (Knowledge Discovery in Databases), yang mencakup tahapan data integration, data cleaning, data transformation, data mining, evaluasi, dan knowledge representation.
b. Hasil evaluasi menggunakan Silhouette Coefficient dengan nilai K = 2 menghasilkan skor sebesar 0.67, yang menunjukkan bahwa struktur clustering yang terbentuk memiliki tingkat pemisahan yang cukup baik atau dapat dikategorikan sebagai medium structure. Untuk menentukan wilayah yang termasuk dalam masing-masing cluster, dilakukan analisis berdasarkan rata-rata jumlah daerah dengan tingkat risiko tinggi.
c. Dari hasil pengelompokan, diperoleh bahwa Cluster 0 terdiri dari 5 daerah yang teridentifikasi berisiko tinggi, yaitu Cikampek, Karawang Barat, Karawang Timur, Klari, dan Teluk Jambe Timur. Sementara itu, Cluster 1 mencakup wilayah yang dikategorikan sebagai tidak berisiko, yaitu Banyusari, Batujaya, Ciampel, Cibuaya, Cikampek Utara, Cilamaya Kulon, Cilamaya Wetan, Cilebar, Jatisari, Jayakerta, Kotabaru, Kutawaluya, Lemah Abang, Majalaya, Pakisjaya, Pangkalan, Pedes, Rawamerta, Rengasdengklok, Talagasari, Tegalwaru, Teluk Jambe Barat, Tempuran, Tirta Mulya, dan Tirtajaya.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

[1]Hevrialni, R., & Sartika, Y. (2021). Intervensi Pendampingan Kurang Energi Kronik (KEK) Pada Ibu Hamil Dengan Pendekatan Continuity Of Midwifery Care (COMC) Sebagai Upaya Pencegahan Stunting. Jurnal Riset Kesehatan Poltekkes Depkes Bandung, 13(2), 310-318.

[2]Kemenkes RI. (2021). Laporan Kinerja Kementrian Kesehatan Tahun 2020. In Kementerian Kesehatan RI.

[3]Putra, M. G. S., & Dewi, M. (2020). Faktor risiko kurang energi kronis (KEK) pada ibu hamil di Cikembar Kabupaten Sukabumi. ARTERI: Jurnal Ilmu Kesehatan, 1(4), 319-332.

[4]Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner. JBASE-Journal of Business and Audit Information Systems, 4(1).

[5]Hidayat, F. M., Rohana, T., Nurlaelasari, E., & Masruriyah, A. F. N. (2024). KLASTERISASI KABUPATEN DAN KOTA DI JAWA BARAT DALAM KASUS GIZI BURUK MENGGUNAKAN ALGORITMA K-MEANS DAN K-MEDOIDS. Jurnal Tekinkom (Teknik Informasi dan Komputer), 7(1), 251-261.

[6]Hanapiah, N. S., Suarna, N., & Prihartono, W. (2023). Meningkatkan Penanganan Stunting Pada Anak Melalui Klasifikasi Pemberian Makanan Tambahan Berdasarkan Usia Dengan Metode K-Means Di Desa Cintarasa. JATI (Jurnal Mahasiswa Teknik Informatika), 7(6), 3484-3493.

[7]Muhammad, M., Mahmudi, A., & Auliasari, K. (2023). Perbandingan Metode K-Means dan K-Medoids untuk Klasifikasi Status Gizi Anak. JATI (Jurnal Mahasiswa Teknik Informatika), 7(4), 2122-2129

[8]Kanah, P. (2020). Hubungan pengetahuan dan pola konsumsi dengan status gizi pada mahasiswa kesehatan. Medical Technology and Public Health Journal, 4(2), 203-211.

[9]Prasetio, A., Hasibuan, M. H., & Sitompul, P. (2021). Simulasi Penerapan Metode Decision Tree (C4. 5) Pada Penentuan Status Gizi Balita. Jurnal Nasional Komputasi Dan Teknologi Informasi, 4(3).

[10]Diningsih, R. F., Wiratmo, P. A., & Lubis, E. (2021). Hubungan Tingkat Pengetahuan tentang Gizi Terhadap Kejadian Kekurangan Energi Kronik (KEK) pada Ibu Hamil. Binawan Student Journal, 3(3), 8-15.

[11]Zai, C. (2022). Implementasi Data Mining Sebagai Pengolahan Data. Jurnal Portal Data, 2(3).

[12]Nugroho, B. I., Ma'arif, Z., & Arif, Z. (2022). Tinjauan Pustaka Sistematis: Penerapan Data Mining Metode Klasifikasi Untuk Menganalisa Penyalahgunaan Sosial Media: Array. Jurnal Sistem Informasi dan Teknologi Peradaban, 3(2), 46-51.

[13]Zuhal, N. K. (2022, February). Study Comparison K-Means Clustering Dengan Algoritma Hierarchical Clustering: AHC, K-Means Clustering, Study Comparison. In Seminar Nasional Teknologi & Sains (Vol. 1, No. 1, pp. 200-205).

[14]Julyantari, N. K. S., Budiarta, I. K., & Putri, N. M. D. K. (2021). Implementasi K-Means Untuk Pengelompokan Status Gizi Balita (Studi Kasus Banjar Titih). Jurnal Janitra Informatika Dan Sistem Informasi, 1(2), 92-101.

[15]Chandra, M. D., Irawan, E., Saragih, I. S., Windarto, A. P., & Suhendro, D. (2021). Penerapan Algoritma K-Means dalam Mengelompokkan Balita yang Mengalami Gizi Buruk Menurut Provinsi. BIOS: Jurnal Teknologi Informasi Dan Rekayasa Komputer, 2(1), 30-38.

[16]Rochman, E. M. S., & Suprajitno, H. (2022). Comparison of clustering in tuberculosis using fuzzy c-means and k-means methods. Commun. Math. Biol. Neurosci., 2022, Article-ID.

[17]Fatimah, S., & Fatmasanti, A. U. (2019). Hubungan Antara Umur, Gravida dan Usia Kehamilan Terhadap Resiko Kurang Energi Kronis (KEK) pada Ibu Hamil. Jurnal Ilmiah Kesehatan Diagnosis, 14(3), 271-274.

[18]Rosita, U., & Rusmimpong, R. (2022). Hubungan paritas dan umur ibu hamil dengan kejadian kekurangan energi kronik di Desa Simpang Limbur Wilayah Kerja Puskesmas Simpang Limbur. Nursing Care and Health Technology Journal (NCHAT), 2(2), 78-86.

[19]Sari, W. K., & Deltu, S. N. (2021). Hubungan Tingkat Pengetahuan Gizi, Anemia, Dan Tingkat Konsumsi Makanan Dengan Kejadian Kek Pada Ibu Hamil Di Desa Muara Madras Kabupaten Merangin Jambi. Jurnal Kesehatan Lentera'aisyiyah, 4(1), 434-439.

[20]Febriansyah, F. (2024). PENERAPAN ALGORITMA K-MEANS CLUSTERING DATA GIZI BALITA PADA UPTD PUSKESMAS BUMI AGUNG. Jurnal Informatika dan Teknik Elektro Terapan, 12(3).