

## Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering Di RSUD Muhammadiyah Bantul

Alfian Aziz Ardiansyah<sup>1</sup>, Ezar Atha Khaeroshi<sup>2</sup>, Raihan Rais Wicaksono<sup>3</sup>

<sup>1,2,3</sup>Informatika, Fakultas Teknologi Industri, Universitas Ahmad Dahlan, Jln. Ring Road Selatan, Tamanan Banguntapan Bantul Yogyakarta, 55166

[tedy.setiady@tif.uad.ac.id](mailto:tedy.setiady@tif.uad.ac.id)<sup>1\*</sup>, [alfianardiansyah295@gmail.com](mailto:alfianardiansyah295@gmail.com)<sup>2</sup>

### Abstrak

*Pengelolaan data rekam medis merupakan salah satu aspek penting dalam mendukung efisiensi operasional rumah sakit. Data ini menyimpan informasi strategis yang dapat dianalisis menggunakan teknik data mining untuk mengungkap pola dan struktur tersembunyi. Penelitian ini menggunakan metode K-Means Clustering untuk mengelompokkan pasien berdasarkan karakteristik tertentu dan mengidentifikasi pola penyakit yang dominan. Data penelitian diambil dari pasien rawat inap RSUD Muhammadiyah Bantul pada periode Oktober hingga Desember 2023, dengan total 871 pasien. Variabel yang dianalisis meliputi jenis kelamin, usia, alamat, diagnosis, durasi rawat inap, dan metode pembayaran. Hasil klasterisasi menghasilkan empat kelompok utama: Cluster 1 dengan 116 pasien (13%), Cluster 2 dengan 136 pasien (16%), Cluster 3 dengan 498 pasien (57%), dan Cluster 4 dengan 121 pasien (14%). Setiap klaster menggambarkan pola prevalensi penyakit tertentu, seperti demam, diabetes, dan penyakit sistemik lainnya. Penelitian ini diharapkan menjadi dasar untuk menyusun strategi peningkatan pelayanan kesehatan yang lebih efisien dan berbasis data di RSUD Muhammadiyah Bantul.*

**Kata kunci:** K-Means Clustering, Klasterisasi Pasien, Rekam Medis, Rumah Sakit

### A. Pendahuluan

Rekam medis adalah dokumen penting yang mencakup informasi tentang identitas pasien, tindakan medis, pengobatan, pemeriksaan, serta berbagai layanan kesehatan lainnya. Dokumen ini memiliki peran sentral dalam mendukung pelayanan rumah sakit. Pengelolaan rekam medis yang baik dapat memfasilitasi pengambilan keputusan klinis maupun administratif. Namun, tantangan pengelolaan data yang kurang efektif masih menjadi kendala di banyak rumah sakit. Data rekam medis yang terus bertambah sering kali tidak dimanfaatkan secara optimal, menyebabkan informasi historis pasien menjadi kurang terintegrasi dalam proses pelayanan.

Kemajuan teknologi informasi telah memberikan solusi dalam pengelolaan data, salah satunya melalui pengembangan sistem basis data yang memungkinkan pengelolaan informasi secara digital. Dalam konteks ini, data mining menjadi salah satu metode yang digunakan untuk menganalisis data besar. Kamber dan Han (2006) menyatakan bahwa data mining adalah konsep yang sangat penting dalam menganalisis data besar untuk mengungkap pola yang tidak terlihat sebelumnya.

Data mining memungkinkan peneliti untuk mengekstrak pengetahuan dari data melalui pendekatan kecerdasan buatan, statistik, dan matematika. Salah satu teknik utama dalam data mining adalah clustering, yang bertujuan untuk mengelompokkan data ke dalam beberapa kategori berdasarkan karakteristik tertentu.

Metode K-Means Clustering, sebagai salah satu teknik clustering yang populer, bekerja dengan mempartisi dataset ke dalam kelompok berdasarkan kesamaan fitur. Menurut Ordila et al. (2020), penerapan algoritma clustering seperti K-Means dapat digunakan untuk mengelompokkan data rekam medis pasien berdasarkan jenis penyakit. Pendekatan ini telah banyak digunakan dalam penelitian untuk menganalisis data medis, seperti identifikasi pola penyakit berdasarkan wilayah, jenis penyakit, dan usia pasien. Misalnya, penelitian sebelumnya menggunakan K-Means untuk mengelompokkan data rekam medis di berbagai fasilitas kesehatan, menghasilkan informasi yang mendukung pengambilan keputusan strategis. Purba et al. (2023) menunjukkan bahwa penggunaan K-Means Clustering dalam pengelolaan data rekam medis dapat memberikan wawasan yang lebih baik untuk pengambilan keputusan.

Dalam penelitian ini, pengelompokan data rekam medis pasien dilakukan berdasarkan variabel seperti jenis kelamin, usia, alamat, diagnosis, lama rawat inap, dan cara pembayaran. Proses klasterisasi dilakukan menggunakan aplikasi RStudio dengan algoritma K-Means Clustering. Penelitian ini bertujuan untuk memberikan wawasan baru tentang pola penyebaran penyakit dan memberikan rekomendasi strategis untuk meningkatkan kualitas pelayanan kesehatan di RSUD Muhammadiyah Bantul.

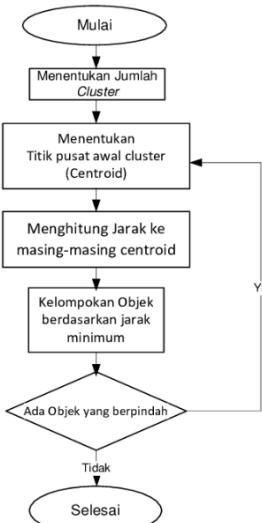
## B. Metode

Metode Penelitian ini menggunakan pendekatan deskriptif kuantitatif untuk mengolah data secara numerik dan menghasilkan gambaran yang jelas tentang pola serta karakteristik data rekam medis pasien di RSUD Muhammadiyah Bantul. Pendekatan ini diterapkan dengan rancangan cross-sectional, yang memungkinkan pengumpulan data pada satu periode waktu tertentu sehingga memberikan informasi yang terfokus pada kondisi saat ini.

Langkah pertama dalam penelitian ini adalah pengumpulan data rekam medis pasien rawat inap yang diperoleh dari Rumah Sakit Umum Daerah Muhammadiyah Bantul. Data tersebut kemudian melewati tahap pembersihan untuk memastikan hanya data yang lengkap dan relevan yang digunakan dalam analisis. Selanjutnya, data diproses melalui tahapan pra-pemrosesan yang mencakup transformasi dan pengkodean ke dalam format spreadsheet untuk mempermudah analisis lebih lanjut.

Tahap utama dalam penelitian ini adalah klasterisasi menggunakan metode K-Means Clustering, sebuah algoritma yang digunakan untuk mengelompokkan data ke dalam beberapa cluster berdasarkan tingkat kesamaan tertentu. Data yang memiliki karakteristik serupa akan dikelompokkan dalam satu cluster, sementara data dengan karakteristik berbeda akan dipisahkan ke dalam cluster lain. Nusa Putra & Anisa (2021) menambahkan bahwa pengelompokan data rekam medis menggunakan K-Means Clustering membantu mengidentifikasi karakteristik pasien berdasarkan diagnosa penyakit dan usia. Proses klasterisasi ini bertujuan untuk mengidentifikasi pola-pola yang mendasari karakteristik pasien, seperti jenis kelamin, usia, dan diagnosis penyakit. Setelah klasterisasi selesai, hasil analisis akan diinterpretasikan untuk menemukan pola atau informasi bermanfaat yang dapat digunakan sebagai dasar rekomendasi dalam meningkatkan pelayanan kesehatan di rumah sakit.

Pendekatan data mining diterapkan untuk menggali informasi yang lebih dalam dari kumpulan data besar ini. Teknik ini memanfaatkan metode kecerdasan buatan, statistik, dan pembelajaran mesin untuk mengidentifikasi hubungan, pola, atau informasi tersembunyi dalam data rekam medis. Dengan pendekatan ini, diharapkan dapat diperoleh wawasan yang lebih mendalam tentang distribusi penyakit dan karakteristik pasien, yang pada akhirnya dapat membantu pihak rumah sakit dalam merancang strategi peningkatan layanan yang lebih efisien dan berbasis data.



Gambar 1. Tahapan Algoritma K-Means

## C. Hasil dan Pembahasan

### 1. Identifikasi Data

Data yang ada sebanyak 871 dengan 6 variabel yaitu diagnosa penyakit, kecamatan, usia, jenis kelamin, lama dirawat dan cara bayar.

1	Jenis Kelamin	Umur	Alamat	Diagnosa	Lama Rawat Inap	Cara Pembayaran
2	Laki-Laki	42	Bantul	Stroke	9	Umum
3	Perempuan	50	Srandakan	Demam Berdarah	10	BPJS
4	Perempuan	23	Srandakan	Penyakit Jantung	4	Umum
5	Perempuan	25	Sewon	Diabetes	6	BPJS
6	Laki-Laki	48	Sewon	Diabetes	1	Umum
7	Laki-Laki	44	Kasihan	Radang Tenggorokan	5	BPJS
8	Perempuan	32	Bantul	Stroke	5	BPJS
9	Laki-Laki	49	Sanden	Diabetes	6	BPJS
10	Perempuan	48	Piyungan	Radang Tenggorokan	5	BPJS
11	Perempuan	26	Srandakan	Radang Tenggorokan	5	Umum
12	Perempuan	46	Kretek	Hipertensi	1	BPJS
13	Perempuan	37	Pleret	Hipertensi	5	BPJS
14	Perempuan	35	Sanden	Vertigo	4	BPJS
15	Laki-Laki	32	Sedayu	Radang Tenggorokan	9	BPJS
16	Laki-Laki	21	Kasihan	Demam Berdarah	3	Umum
17	Laki-Laki	22	Pleret	Gastritis	7	Umum
18	Perempuan	38	Imogiri	Radang Tenggorokan	8	BPJS
19	Perempuan	30	Sedayu	Asma	9	BPJS
20	Laki-Laki	40	Sewon	Gastritis	8	BPJS
21	Laki-Laki	48	Jatis	Asma	6	BPJS
22	Laki-Laki	54	Pleret	Hipertensi	5	BPJS

Gambar 1. Identifikasi Data

### 2. Tahapan Pada Data mining

#### a. Pemeriksaan missing values

Pemeriksaan missing values atau seleksi data adalah langkah penting untuk menghilangkan data yang tidak lengkap atau mengandung noise. Tahap ini bertujuan agar proses eksplorasi data dalam Knowledge Discovery in Database (KDD) tidak terganggu. Proses diawali dengan memastikan data yang akan digunakan dalam data mining telah disimpan dengan benar dalam sebuah file.

b. Normalisasi data

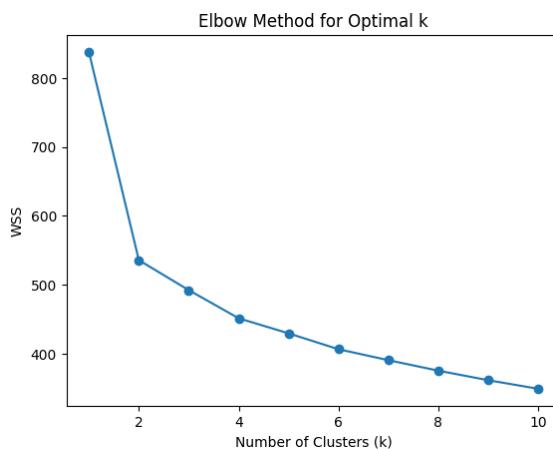
Normalisasi data dilakukan untuk menyelaraskan skala data numerik dengan menggunakan metode scaling. Pada tahap ini, dilakukan penskalaan nilai pada data frame menggunakan Min-Max Scaler, yang mengubah nilai ke dalam rentang  $[0, 1]$ .

c. Transform data

Transformasi data adalah langkah untuk mengubah format atau struktur data agar lebih mudah dianalisis atau diproses. Biasanya, proses ini melibatkan konversi variabel kategori menjadi tipe data faktor atau bentuk lain yang sesuai.

d. Penentuan Jumlah Cluster dalam K-Prototype Clustering

Jumlah cluster ditentukan menggunakan metode elbow. Dalam metode ini, dilakukan analisis hubungan antara WSS (Within-Cluster Sum of Squares) dan jumlah cluster. Grafik hasil analisis akan menunjukkan titik optimal yang merepresentasikan jumlah cluster terbaik untuk dataset yang digunakan.



Gambar 3. Grafik Elbow

e. Membuat Centeroid

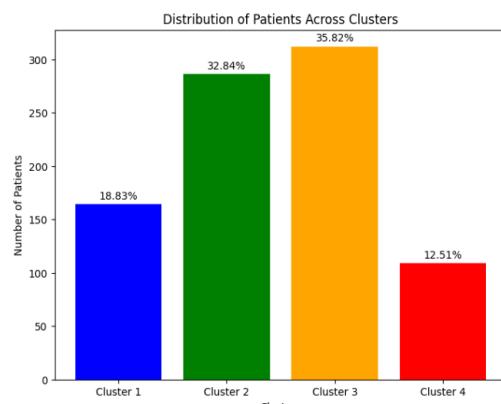
Pada tahap ini, centroid memiliki peran penting dalam menentukan posisi pusat dari setiap klaster yang terbentuk selama proses pembuatan model K-Prototypes. Untuk memperoleh hasil klasterisasi, digunakan fungsi *enumerate()* untuk mengakses indeks dan koordinat pusat dari *cluster\_centers*. Dari implementasi tersebut, diperoleh koordinat centroid untuk setiap klaster yang dihasilkan.

	Umur	Lama Rawat Inap	Jenis Kelamin	Diagnosa	Alamat	Cara Pembayaran
0	0.7899529964747367	0.8345410628019316	Laki-Laki	Diabetes	Sanden	BPJS
1	0.24409672830725443	0.2444444444444444	Perempuan	Vertigo	Kasihan	Umum
2	0.8026398491514776	0.1950904392764858	Laki-Laki	Gastritis	Jetis	Umum
3	0.5462033462033462	0.2941798941798944	Perempuan	Penyakit Jantung	Piyungan	BPJS
4	0.35699878983461064	0.751243781094527	Laki-Laki	Vertigo	Jetis	BPJS
5	0.15720720720720688	0.8648148148148144	Laki-Laki	Gastritis	Sewon	BPJS
6	0.7069701280227596	0.2309941520467836	Laki-Laki	Hipertensi	Banguntapan	BPJS
7	0.176164115923152	0.18473895582329275	Laki-Laki	Asma	Jetis	BPJS
8	0.7072972972972973	0.7655555555555555	Perempuan	Vertigo	Kretek	Umum
9	0.29047739328113154	0.6479750778816193	Laki-Laki	Diabetes	Sewon	Umum

**Gambar 4. Hasil Klasterisasi**

**f. Memvisualisasi Hasil Analisis dan Pembagian Data Cluster**

Memvisualisasi Hasil Analisis dan Pembagian Data Cluster Visualisasi hasil klaster dilakukan menggunakan bar chart agar terlihat jelas perbedaan jumlah anggota dalam masing-masing klaster. Dari gambar dapat dilihat bahwa klaster 1 terdiri dari 164 pasien (18,82%), klaster 2 terdiri dari 286 pasien (32,83%), klaster 3 terdiri dari 312 pasien (35,82 %), dan klaster 4 terdiri dari 109 pasien (12,51%).



**Gambar 5. Hasil Visualisasi Klasterisasi**

**g. Hasil Analisa Data Setiap Cluster**

Keterangan	Kluster 1	Kluster 2	Kluster 3	Kluster 4
Jumlah Data	164	286	312	109
Jenis Kelamin	Laki-laki = 83 Perempuan = 81	Laki-laki = 96 Perempuan = 190	Laki-laki = 180 Perempuan = 132	Laki-laki = 50 Perempuan = 59
Alamat	Bantul strandakan sewon kasihan sanden piyungan kretek pleret sedaya imogiri jetis dlingo pundong banguntapan pandak bambanglipuro	Bantul strandakan sewon kasihan sanden piyungan kretek pleret sedaya imogiri jetis dlingo pundong banguntapan pandak bambanglipuro	Bantul strandakan sewon kasihan sanden piyungan kretek pleret sedaya imogiri jetis dlingo pundong banguntapan pandak bambanglipuro	Bantul strandakan sewon kasihan sanden piyungan kretek pleret sedaya imogiri jetis dlingo pundong banguntapan pandak bambanglipuro
Lama Rawat Inap	1 - 10	1 - 10	1 - 10	1 - 10
Cara Pembayaran	BPJS = 98 Umum = 66	BPJS = 144 Umum = 142	BPJS = 150 Umum = 162	BPJS = 51 Umum = 58

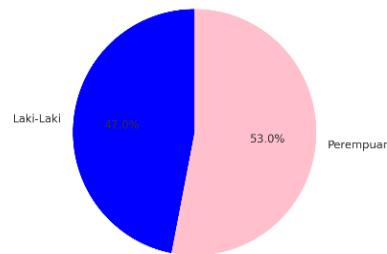
**Gambar 6. Hasil Klasterisasi**

### 3. Proses Pengelolahan Kontingensi antar Variabel

#### a. Pengelompokan Variable Jenis Kelamin

Faktor jenis kelamin dan gender menjadi salah satu aspek yang memengaruhi penggunaan fasilitas pelayanan kesehatan. Preferensi pasien laki-laki dan perempuan dalam memilih layanan kesehatan dapat dipengaruhi oleh kondisi ekonomi, tingkat pendidikan, fasilitas yang tersedia, kenyamanan, serta faktor lainnya. Selain itu, data jenis kelamin pasien di rumah sakit juga dapat dimanfaatkan untuk menyusun tabulasi jumlah kunjungan pasien di bangsal berdasarkan jenis kelamin. Dari hasil pengelompokan jumlah kunjungan pasien rasio jumlah kelamin laki-laki lebih banyak daripada perempuan dengan jumlah 409 dan 462 dari total kunjungan 871 pasien.

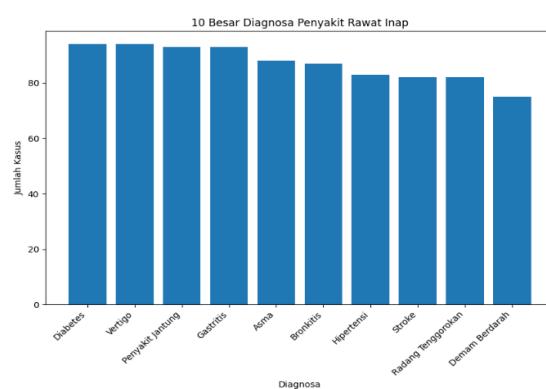
Distribusi Laki-Laki dan Perempuan



Gambar 7. Hasil Pengelompokan Variabel Jenis Kelamin

#### b. Pengelompokan 10 Besar Diagnosa Penyakit

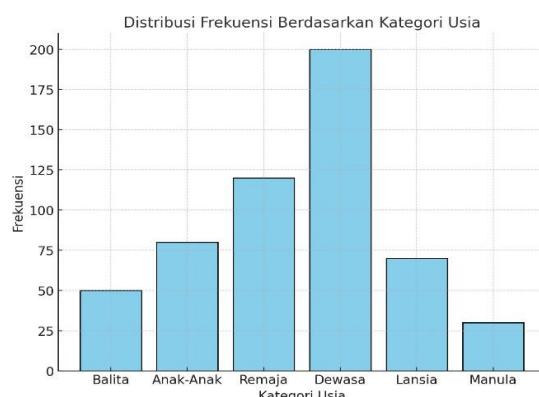
Pembangunan kesehatan bertujuan untuk meningkatkan kesadaran, kemauan, kemampuan hidup sehat bagi setiap orang agar terwujud derajat kesehatan yang optimal. Kondisi kesehatan masyarakat di suatu wilayah dapat dilihat dari jumlah kasus penyakit yang berkunjung ke Rumah Sakit. Dari data hasil kunjungan pasien rawat inap di Rumah Sakit Umum Muhammadiyah Bantul dapat dilihat pada gambar di bawah ini.



Gambar 8. Hasil Pengelompokan Diagnosa Penyakit

c. Pengelompokan Berdasarkan Kategori Usia

Dalam konsep dasar epidemiologi, hubungan antara tiga faktor utama yang mempengaruhi terjadinya penyakit, salah satunya adalah host. Salah satu faktor host yang penting adalah usia. Seperti terlihat pada bar chart, usia dapat dikelompokkan ke dalam beberapa kategori, yaitu balita (0-5 tahun), anak-anak (6-10 tahun), remaja (11-15 tahun), dewasa (16-40 tahun), lansia (41-60 tahun), dan manula (>60 tahun). Kategori ini memberikan gambaran tentang distribusi frekuensi berdasarkan rentang usia.



Gambar 9. Hasil Pengelompokan Kategori Usia

D. Kesimpulan

Dari hasil penelitian ini dapat disimpulkan bahwa penerapan data mining melalui *K-Means Clustering* pada pengelompokan data rekam medis pasien berdasarkan variabel jenis kelamin, usia, alamat, diagnosis, lama rawat inap, dan metode pembayaran dapat diterapkan menggunakan aplikasi RStudio. Sumber data studi ini adalah data sekunder yang diambil dari pasien rawat inap di RSUD Muhammadiyah Bantul pada periode Oktober hingga Desember 2023, dengan jumlah data sebanyak 871 pasien. Dari hasil pengelompokan, diperoleh empat klaster: klaster 1 terdiri dari 116 pasien (13%), klaster 2 terdiri dari 136 pasien (16%), klaster 3 terdiri dari 498 pasien (57%), dan klaster 4 terdiri dari 121 pasien (14%). Kasus prevalensi penyakit yang paling banyak diderita pasien rawat inap berdasarkan hasil klasterisasi ini adalah demam, diabetes, dan penyakit sistemik lainnya. Pengelompokan data rekam medis pasien dari proses data mining ini dapat digunakan sebagai dasar untuk manajemen statistik dalam pengambilan keputusan di rumah sakit, terkait dengan peningkatan pelayanan atau fasilitas rawat inap, seperti ketersediaan ruang perawatan dan tempat tidur, penyediaan stok obat, ataupun kebutuhan dokter spesialis. Selain itu, pengelompokan penyakit yang dihasilkan dari data mining ini memberikan informasi baru mengenai penyebaran penyakit di setiap klaster yang terbentuk. Hal ini dapat menjadi acuan bagi pihak rumah sakit untuk meningkatkan kualitas pelayanan kesehatan serta bagi pemerintah daerah, terutama dinas kesehatan, untuk melakukan sosialisasi dan pencegahan terhadap penyakit yang banyak diderita oleh pasien, berdasarkan analisis usia, jenis kelamin, dan diagnosis penyakit yang ada.

## Daftar Pustaka

M. Kamber and J. Han, *Data Mining: Concepts and Techniques : Concepts and Techniques*. 2006.

R. Ordila, R. Wahyuni, Y. Irawan, and M. Yulia Sari, "Penerapan Data Mining Untuk Pengelompokan Data Rekam Medis Pasien Berdasarkan Jenis Penyakit Dengan Algoritma Clustering (Studi Kasus : Poli Klinik PT.Inecda)," *J. Ilmu Komput.*, vol. 9, no. 2, pp. 148–153, 2020, doi: 10.33060/jik/2020/vol9.iss2.181.

W. Purba et al., "Penerapan Data Mining Untuk Pengelolaan Data Rekam Medis Menggunakan Metode K-means Clustering Pada Rumah Sakit Royal Prima Medan," *J. TEKINKOM*, vol. 6, no. 1, pp. 158–168, 2023, doi: 10.37600/tekinkom.v6i1.857

M. D. Chandra, E. Irawan, I. S. Saragih, A. P. Windarto, and D. Suhendro, "Penerapan Algoritma K- Means dalam Mengelompokkan Balita yang Mengalami Gizi Buruk Menurut Provinsi," *BIOS J. Teknol. Inf. dan Rekayasa Komput.*, vol. 2, no. 1, pp. 30–38, 2021, doi: 10.37148/bios.v2i1.19.

H. Nusa Putra and D. Putri Anisa, "Klasterisasi Data Rekam Medis Pada Diagnosa Penyakit Berdasarkan Usia Pasien Menggunakan Algoritma K-Means Di Puskesmas Lubuk Alung," *Ensiklopedia J.*, vol. 3, no. 5, pp. 128–133, 2021

Yullia Agustin and Adiati Trihastuti, "Pengaruh Kualitas Pelayanan Terhadap Kepuasan Wajib Pajak Kendaraan Bermotor (Di Kantor Samsat Kabupaten Lamongan)," *J. Student Res.*, vol. 1, no. 2, pp. 248–263, 2023, doi: 10.55606/jsr.v1i2.994.

T. Amalina, D. Bima, A. Pramana, and B. N. Sari, "Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food," *J. Ilm. Wahana Pendidik.*, vol. 8, no. 15, pp. 574–583, 2022.

Dilawati, H., Widianto, H., & Kuswiadji, A. (2024). Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering Di Rumah Sakit Widodo Ngawi. *BIOS: Jurnal Teknologi Informasi dan Rekayasa Komputer*, 5(2), 139-147.

Idham, I., Rosika, H., & Yuliadi, Y. (2024). IMPLEMENTASI RAPIDMINER UNTUK CLESTERING DATA PENJUALAN PAKAIAN MENGGUNAKAN METODE K-MEANS. *JUTECH: Journal Education and Technology*, 5(1), 221-231.

Azkar, A., & Kusrini, K. (2024). Klasterisasi Pasien Rawat Jalan di Puskesmas dengan Menggunakan Metode Algoritma Clustering K-Means. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 4(5), 2628-2636.

Hendrik, J. (2024). Perancangan Sistem Informasi Pengklasifikasian Rumah Sakit Menggunakan Algoritma K-Means Clustering. *Bulletin of Computer Science Research*, 4(2), 225-232.

