

IMPLEMENTASI METODE *K-MEANS CLUSTERING* UNTUK MENGLASIFIKASI STATUS GIZI BALITA BERDASARKAN WILAYAH KERJA PUSKESMAS KARANG REJO

Ayunda Pratiwi¹, Hotler Manurung², Melda Pita Uli Sitompul³

Sistem Informasi, STMIK Kaputama, Binjai

E-mail: ayundapratiwi205@gmail.com¹

ABSTRAK

Status gizi balita merupakan indikator vital dalam kesehatan masyarakat, namun seringkali data yang tersebar di fasilitas kesehatan, seperti puskesmas, belum dianalisis secara optimal. Penelitian ini bertujuan untuk menerapkan metode *K-Means Clustering* dalam mengklasifikasi status gizi balita di wilayah kerja Puskesmas Karang Rejo. Data yang digunakan mencakup usia, wilayah, dan status gizi balita pada tahun 2024. Melalui proses *Knowledge Discovery in Database* (KDD), data mentah diubah menjadi informasi yang terstruktur dan bermakna. Hasil klasifikasi menunjukkan bahwa pengelompokan menjadi 3 *Cluster* adalah yang paling optimal dengan nilai *Cluster Variance* terendah sebesar 9,106. *Cluster* ini merepresentasikan balita usia 2 tahun di Kwala Begumit dengan gizi buruk, balita usia 4 tahun di Ara Condong dengan gizi baik, dan balita usia 2 tahun di Perdamaian dengan gizi buruk. Analisis ini membuktikan efektivitas algoritma *K-Means* dalam menyediakan informasi yang akurat dan terstruktur untuk mendukung pengambilan keputusan yang lebih cepat, tepat, dan terarah dalam upaya pencegahan dan penanganan masalah gizi balita.

Kata kunci

Data Mining, *K-Means Clustering*, Status Gizi, Balita, Puskesmas

ABSTRACT

Toddler nutritional status is a vital indicator of public health, but data scattered across healthcare facilities, such as community health centers, are often not optimally analyzed. This study aims to apply the K-Means Clustering method to classify the nutritional status of toddlers in the working area of Karang Rejo Community Health Center. The data used include the age, region, and nutritional status of toddlers in 2024. Through the Knowledge Discovery in Database (KDD) process, raw data are transformed into structured and meaningful information. The classification results indicate that grouping into 3 Clusters is the most optimal, with the lowest Cluster Variance value of 9.106. These Clusters represent toddlers aged 2 in Kwala Begumit with poor nutrition, toddlers aged 4 in Ara Condong with good nutrition, and toddlers aged 2 in Perdamaian with poor nutrition. This analysis proves the effectiveness of the K-Means algorithm in providing accurate and structured information to support faster, more precise, and targeted decision-making in the prevention and management of toddler malnutrition.

Keywords

Data Mining, *K-Means Clustering*, Nutritional Status, Toddlers, Community Health Center.

1. PENDAHULUAN

Status gizi balita menjadi cerminan keberhasilan kesehatan sebuah wilayah, yang secara langsung berkaitan dengan asupan nutrisi dan keseimbangan pertumbuhan (Raudotul Janah, Kurniawan, & Suprpti, 2024). Di Puskesmas Karang Rejo, data status gizi balita mencatat bahwa meskipun mayoritas balita memiliki status gizi baik (65,69%), prevalensi gizi buruk (14,54%) dan gizi kurang (19,77%) masih tergolong tinggi. Pengelolaan data secara manual yang saat ini dilakukan oleh pihak puskesmas menyulitkan identifikasi pola dan penentuan prioritas intervensi gizi secara spesifik per

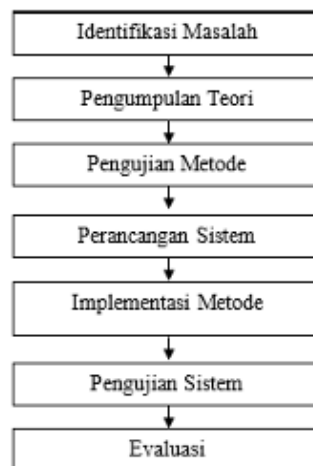
wilayah. Kondisi ini menuntut sebuah pendekatan sistematis untuk mengubah data menjadi informasi yang dapat ditindaklanjuti.

Penelitian terdahulu menunjukkan bahwa metode data *mining*, khususnya algoritma *K-Means Clustering*, efektif dalam mengelompokkan data berdasarkan kemiripan karakteristik. Penelitian oleh Julyantari et al. (2021) dan Janah et al. (2023) berhasil mengimplementasikan *K-Means* untuk mengelompokkan status gizi balita. Penelitian ini juga relevan dengan studi Surbakti (2021) tentang pengelompokan pasien dan penelitian Dewi et al. (2022) tentang pengelompokan siswa berdasarkan keaktifan belajar. Selain itu, penggunaan data *mining* juga telah diterapkan dalam berbagai sektor lain, seperti pada data bencana dan perikanan (Setiadi et al., 2023).

Berdasarkan permasalahan dan relevansi penelitian sebelumnya, penelitian ini mengusulkan penerapan *K-Means Clustering* untuk mengklasifikasi status gizi balita di wilayah kerja Puskesmas Karang Rejo. Tujuan penelitian ini adalah untuk: (1) menerapkan metode *K-Means Clustering* guna mendukung proses pengambilan keputusan; (2) mengklasifikasi balita berdasarkan usia, wilayah, dan status gizi; dan (3) menyediakan informasi terstruktur sebagai dasar evaluasi program kesehatan. Dengan demikian, diharapkan hasil penelitian ini dapat memberikan kontribusi nyata dalam upaya pencegahan dan penanganan masalah gizi balita secara lebih efisien dan tepat sasaran.

2. METODE PENELITIAN

Metodologi penelitian merupakan langkah ilmiah yang digunakan untuk memperoleh data sesuai tujuan penelitian. Dalam penelitian ini, penulis mengikuti beberapa tahapan metodologi sebagai berikut:



Gambar 1. Alur Kerja Penelitian

Proses *K-Means Clustering* dilakukan dengan langkah-langkah berikut:

1. Menentukan jumlah *Cluster* (k) yang diinginkan.
2. Menginisialisasi k buah centroid awal secara acak.
3. Menghitung jarak Euclidean dari setiap data ke setiap centroid.
4. Mengelompokkan setiap data ke dalam *Cluster* dengan jarak terdekat.
5. Memperbarui posisi centroid dengan menghitung rata-rata dari semua data dalam setiap *Cluster*.
6. Mengulangi langkah 3-5 hingga tidak ada perubahan posisi centroid yang signifikan

3. HASIL DAN PEMBAHASAN

3.1 Data Input

Penelitian ini menggunakan data balita yang bersumber dari Puskesmas Karang Rejo. Data tersebut pada tahap awal disimpan dalam format Microsoft Excel sebagai media penyimpanan, kemudian dilakukan proses transformasi data berdasarkan variabel penelitian yang ditetapkan. Rincian data yang digunakan dapat dijelaskan sebagai berikut:

- a. Data Masukan
 - a. Nama file : Data Balita
 - b. Jumlah data : 1740 entri
 - c. Variabel penelitian:
 - X = Usia
 - Y = Wilayah
 - Z = Status Gizi
- b. *Cluster* pengelompokan : 3 *Cluster*
- c. Untuk mempermudah pengolahan, setiap variabel diberikan kode numerik sebagai berikut:

Tabel 1. Inisialisasi Kriteria Usia

Kode	Usia Anak
1	1 Tahun
2	2 Tahun
3	3 Tahun
4	4 Tahun
5	5 Tahun

Tabel 2. Inisialisasi Wilayah

Kode	Keterangan
1	Mangga
2	Kwala Begumit
3	Ara Condong
4	Karang rejo
5	Perdamaian
6	Dendang

Tabel 3. Inisialisasi Status Gizi

Kode	Status Gizi
1	Gizi Baik
2	Gizi Buruk
3	Gizi Kurang

3.2 Implementasi

Implementasi pada tahap ini merupakan proses akhir dari penerapan metode *Clustering* sesuai dengan rancangan yang telah disusun. Proses implementasi dilakukan menggunakan aplikasi MATLAB dengan menerapkan algoritma *K-Means Clustering* pada kode program, sehingga sistem mampu mengelompokkan data balita sesuai tahapan yang telah ditentukan. Berdasarkan hasil perancangan, sistem berhasil mengklasifikasikan data balita ke dalam tiga *Cluster* berdasarkan variabel usia, wilayah, dan status gizi.

Dengan adanya sistem ini, Puskesmas Karang Rejo dapat melakukan analisis status gizi balita secara lebih efektif dan efisien. Penerapan metode *Clustering* memudahkan

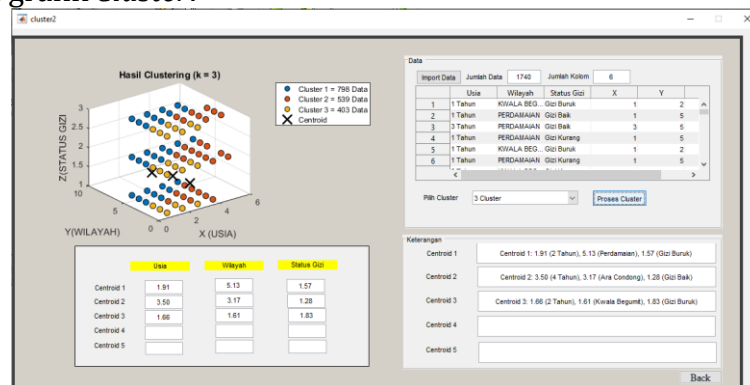
dalam mengidentifikasi pola tersembunyi pada data, sehingga dapat diketahui wilayah yang memiliki tingkat gizi buruk maupun gizi kurang lebih tinggi. Informasi tersebut diharapkan dapat menjadi dasar bagi pihak terkait dalam merumuskan strategi yang tepat guna meningkatkan kualitas gizi balita di wilayah kerja Puskesmas Karang Rejo.

3.3 Hasil

Hasil pengujian tersebut menjadi dasar analisis yang akan digunakan dalam penyusunan kesimpulan. Berikut disajikan uraian dari setiap tahap pengujian:

a. Hasil Pengujian 3 Cluster

Hasil pengujian pertama pada proses *Clustering* ditampilkan dalam bentuk tabel serta visualisasi grafik *Cluster*.



Gambar 2. Hasil Pengujian 3 Cluster

Berdasarkan hasil penerapan metode *Clustering* terhadap 1.740 data status gizi balita, diperoleh tiga kelompok (*Cluster*) dengan rincian, yaitu *Cluster* 1 terdiri dari 798 data, *Cluster* 2 terdiri dari 539 data, dan *Cluster* 3 terdiri dari 403 data.

a. Centroid 1 : 2(1.91) 5(5.13) 2(1.57)

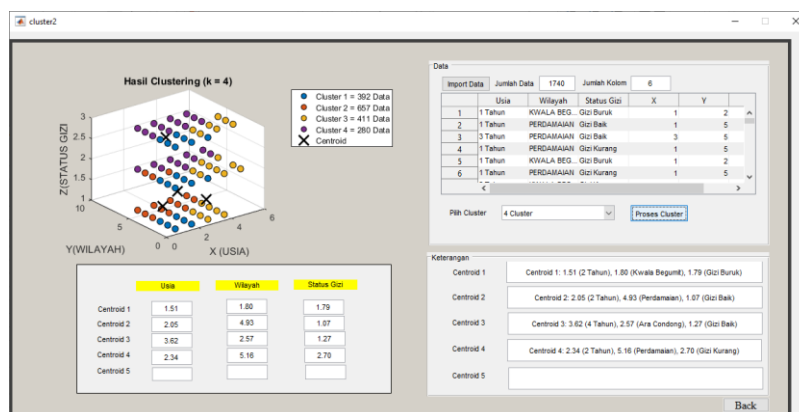
Dapat diketahui bahwa centroid pertama merepresentasikan balita berusia 2 tahun yang berada di wilayah Perdamaian dengan status gizi Buruk.

b. Centroid 2 : 4(3.50) 3(3.17) 1(1.28)

Dapat diketahui bahwa centroid kedua merepresentasikan balita berusia 4 tahun yang berada di wilayah Ara Condong dengan status gizi Baik.

c. Centroid 3 : 2(1.66) 2(1.61) 2(1.83)

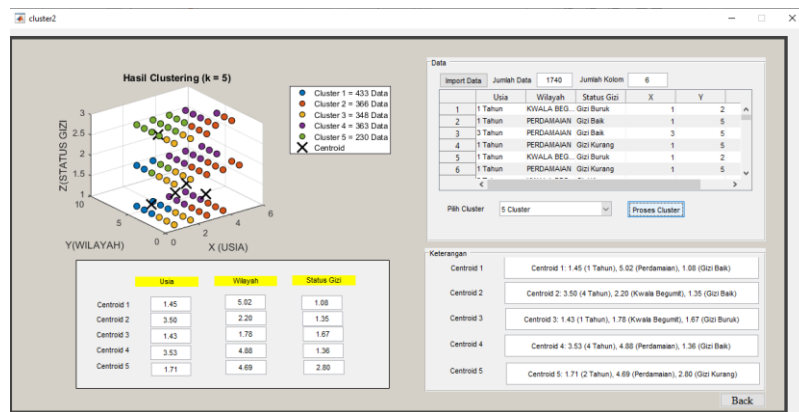
Dapat diketahui bahwa centroid ketiga merepresentasikan balita berusia 2 tahun yang berada di wilayah Kwala Begumit dengan status gizi Buruk.



Gambar 3. Hasil Pengujian 4 Cluster

Berdasarkan hasil penerapan metode *Clustering* terhadap 1.740 data status gizi balita, diperoleh empat kelompok (*Cluster*) dengan rincian, yaitu *Cluster 1* terdiri dari 392 data, *Cluster 2* terdiri dari 657 data, *Cluster 3* terdiri dari 411 data, dan *Cluster 4* terdiri dari 280 data.

- b. Centroid 1 : 2(1.51) 2(1.80) 2(1.79)
Dapat diketahui bahwa centroid pertama merepresentasikan balita berusia 2 tahun yang berada di wilayah Kwala Begumit dengan status gizi Buruk.
- c. Centroid 2 : 2(2.05) 5(4.93) 1(1.07)
Dapat diketahui bahwa centroid kedua merepresentasikan balita berusia 2 tahun yang berada di wilayah Perdamaian dengan status gizi Baik.
- d. Centroid 3 : 4(3.62) 3(2.57) 2(1.27)
Dapat diketahui bahwa centroid ketiga merepresentasikan balita berusia 4 tahun yang berada di wilayah Ara Condong dengan status gizi Buruk.
- e. Centroid 4 : 2(2.34) 5(5.16) 3(2.70)
Dapat diketahui bahwa centroid keempat merepresentasikan balita berusia 2 tahun yang berada di wilayah Perdamaian dengan status gizi Kurang.



Gambar 3. Hasil Pengujian 5 Cluster

Berdasarkan hasil penerapan metode *Clustering* terhadap 1.740 data status gizi balita, diperoleh lima kelompok (*Cluster*) dengan rincian, yaitu *Cluster 1* terdiri dari 433 data, *Cluster 2* terdiri dari 366 data, *Cluster 3* terdiri dari 348 data, *Cluster 4* terdiri dari 363 data, dan *Cluster 5* terdiri dari 230 data.

- a. Centroid 1 : 1(1.45) 5(5.02) 1(1.08)
Dapat diketahui bahwa centroid pertama merepresentasikan balita berusia 1 tahun yang berada di wilayah Perdamaian dengan status gizi Baik.
- b. Centroid 2 : 4(3.50) 2(2.20) 1(1.35)
Dapat diketahui bahwa centroid kedua merepresentasikan balita berusia 4 tahun yang berada di wilayah Kwala Begumit dengan status gizi Baik.
- c. Centroid 3 : 1(1.43) 2(1.78) 2(1.67)
Dapat diketahui bahwa centroid ketiga merepresentasikan balita berusia 1 tahun yang berada di wilayah Kwala Begumit dengan status gizi Buruk.
- d. Centroid 4 : 4(3.53) 5(4.88) 1(1.36)
Dapat diketahui bahwa centroid keempat merepresentasikan balita berusia 4 tahun yang berada di wilayah Perdamaian dengan status gizi Baik.
- e. Centroid 5 : 2(1.71) 5(4.69) 3(2.80)
Dapat diketahui bahwa centroid kelima merepresentasikan balita berusia 2 tahun yang berada di wilayah Perdamaian dengan status gizi Kurang.

3.4 Evaluasi Hasil

Untuk menentukan jumlah *Cluster* yang paling optimal, dilakukan evaluasi dengan mengukur nilai varians (*Variance*), varians minimum (V_{min}), varians maksimum (V_{max}), dan *Cluster Variance*.

a. Pengujian Hasil 3 Cluster

- 1) *Cluster 1* (1,91; 5,13; 1,57) Dengan 798 Data
 $Variance\ C1 = 1/798 (1,0824 + 1,2318 + 1,7047 + + 1,0824) = 1,386$
 $V_{min} = min (1,0824 + 1,2318 + 1,7047 + + 1,0824) = 0,4602$
 $V_{max} = max (1,0824 + 1,2318 + 1,7047 + + 1,0824) = 3,2365$
- 2) *Cluster 2* (3,50; 3,17; 1,28) Dengan 539 Data
 $Variance\ C1 = 1/539 (1,9172 + 0,8904 + 0,8904 + + 1,3039) = 12,390$
 $V_{min} = min (1,9172 + 0,8904 + 0,8904 + + 1,3039) = 0,5989$
 $V_{max} = max (1,9172 + 0,8904 + 0,8904 + + 1,3039) = 2,7334$
- 3) *Cluster 3* (1,66; 1,61; 1,83) Dengan 403 Data
 $Variance\ C1 = 1/403 (0,7813 + 0,7813 + 1,2778 + + 1,3939) = 13,542$
 $V_{min} = min (0,7813 + 0,7813 + 1,2778 + + 1,3939) = 0,5480$
 $V_{max} = max (0,7813 + 0,7813 + 1,2778 + + 1,3939) = 1,9305$

b. Pengujian Hasil 4 Cluster

- 1) *Cluster 1* (1,51; 1,80; 1,79) Dengan 392 Data
 $Variance\ C1 = 1/392 (0,5867 + 0,5867 + 1,3257 + + 1,3314) = 13,270$
 $V_{min} = min (0,5867 + 0,5867 + 1,3257 + + 1,3314) = 0,5735$
 $V_{max} = max (0,5867 + 0,5867 + 1,3257 + + 1,3314) = 2,0830$
- 2) *Cluster 2* (2,05; 4,93; 1,07) Dengan 657 Data
 $Variance\ C1 = 1/657 (1,0515 + 0,9576 + 0,1068 + + 1,0515) = 12,028$
 $V_{min} = min (1,0515 + 0,9576 + 0,1068 + + 1,0515) = 0,1068$
 $V_{max} = max (1,0515 + 0,9576 + 0,1068 + + 1,0515) = 2,2256$
- 3) *Cluster 3* (3,62; 2,57; 1,27) Dengan 411 Data
 $Variance\ C1 = 1/411 (1,0517 + 1,0517 + 0,8860 + + 0,7343) = 11,386$
 $V_{min} = min (1,0517 + 1,0517 + 0,8860 + + 0,7343) = 0,6329$
 $V_{max} = max (1,0517 + 1,0517 + 0,8860 + + 0,7343) = 2,6336$
- 4) *Cluster 4* (2,34; 5,16; 2,70) Dengan 280 Data
 $Variance\ C1 = 1/280 (1,3857 + 1,3857 + 1,7296 + + 0,7408) = 13,459$
 $V_{min} = min (1,3857 + 1,3857 + 1,7296 + + 0,7408) = 0,4842$
 $V_{max} = max (1,3857 + 1,3857 + 1,7296 + + 0,7408) = 2,8720$

c. Pengujian Hasil 5 Cluster

- 1) *Cluster 1* (1,45; 5,02; 1,08) Dengan 433 Data
 $Variance\ C1 = 1/433 (0,4533 + 0,5604 + 0,5604 + + 0,4533) = 8,812$
 $V_{min} = min (0,4533 + 0,5604 + 0,5604 + + 0,4533) = 0,4533$
 $V_{max} = max (0,4533 + 0,5604 + 0,5604 + + 0,4533) = 1,4544$
- 2) *Cluster 2* (3,50; 2,20; 1,35) Dengan 366 Data
 $Variance\ C1 = 1/366 (1,1420 + 1,1420 + 0,6472 + + 0,6430) = 11,014$
 $V_{min} = min (1,1420 + 1,1420 + 0,6472 + + 0,6430) = 0,6430$
 $V_{max} = max (1,1420 + 1,1420 + 0,6472 + + 0,6430) = 2,3639$
- 3) *Cluster 3* (1,43; 1,78; 1,67) Dengan 348 Data
 $Variance\ C1 = 1/348 (0,5859 + 0,5859 + 1,4675 + + 1,4177) = 12,482$
 $V_{min} = min (0,5859 + 0,5859 + 1,4675 + + 1,4177) = 0,5859$
 $V_{max} = max (0,5859 + 0,5859 + 1,4675 + + 1,4177) = 1,6465$
- 4) *Cluster 4* (3,53; 4,88; 1,36) Dengan 363 Data
 $Variance\ C1 = 1/363 (0,6576 + 0,6030 + 1,3707 + + 1,0588) = 11,327$

$$V_{min} = \min (0,6576 + 0,6030 + 1,3707 + \dots + 1,0588) = 0,6030$$

$$V_{max} = \max (0,6576 + 0,6030 + 1,3707 + \dots + 1,0588) = 2,3659$$

5) Cluster 5 (1,71; 4,69; 2,80) Dengan 230 Data

$$\text{Variance C1} = 1/230 (0,7995 + 0,7995 + 1,5022 + \dots + 1,3433) = 11,897$$

$$V_{min} = \min (0,7995 + 0,7995 + 1,5022 + \dots + 1,3433) = 0,4710$$

$$V_{max} = \max (0,7995 + 0,7995 + 1,5022 + \dots + 1,3433) = 1,8498$$

Tabel 1. Hasil Pengujian 4 Cluster

Cluster	Centroid	Variance	V_{min}	V_{max}	Cluster Variance
3	1,91; 5,13; 1,57	1,386	0,4602	3,2365	9,106
	3,50; 3,17; 1,28	12,39	0,5989	2,7334	
	1,66; 1,61; 1,83	13,542	0,548	1,9305	
4	1,51; 1,80; 1,79	13,27	0,5735	2,083	12,648
	2,05; 4,93; 1,07	12,028	0,1068	2,2256	
	3,62; 2,57; 1,27	11,386	0,6329	2,6336	
	2,34; 5,16; 2,70	13,459	0,4842	2,872	
5	1,45; 5,02; 1,08	8,812	0,4533	1,4544	11,106
	3,50; 2,20; 1,35	11,014	0,643	2,3639	
	1,43; 1,78; 1,67	12,482	0,5859	1,6465	
	3,53; 4,88; 1,36	11,327	0,603	2,3659	
	1,71; 4,69; 2,80	11,897	0,471	1,8498	

Berdasarkan hasil evaluasi, pengelompokan dengan 3 cluster adalah yang paling optimal karena memiliki nilai Cluster Variance terendah, yang menandakan bahwa data lebih terkonsentrasi dan representatif. Hasil ini dapat menjadi dasar bagi Puskesmas Karang Rejo untuk memprioritaskan intervensi gizi di wilayah dengan kasus gizi buruk yang tinggi secara lebih efisien.

4. KESIMPULAN

Penerapan algoritma *K-Means* Clustering berhasil mengklasifikasi status gizi balita di Puskesmas Karang Rejo. Pengelompokan data menjadi 3 cluster terbukti paling optimal, yang ditandai dengan nilai Cluster Variance terendah sebesar 9,106. Hasil ini memungkinkan Puskesmas untuk mengidentifikasi kelompok balita berisiko, khususnya di wilayah Kwala Begumit dan Perdamaian, yang membutuhkan intervensi gizi segera. Klasifikasi yang dihasilkan dapat mendukung upaya pencegahan dan penanganan gizi buruk secara lebih cepat, tepat, dan terarah.

5. DAFTAR PUSTAKA

- Abdillah, R. (2021). Pemodelan Uml Untuk Sistem Informasi Persewaan Alat Pesta. *Jurnal Fasilkom*, 11(2), 79–86.
- Fithri, P., Muluk, A., & Rayhanda, R. H. (2024). Perancangan User Interface (UI) Dan User Experience (UX) Pada Sistem Informasi PT. XYZ. *Jurnal Nasional Teknologi dan*

- Sistem Informasi, 9(3), 280–289.
- Halimah, N. N. (2024). Karakterisasi Sensor Hy-Srf05 Dan Load Cell Single-Point Sebagai Parameter Pengukuran Antropometri Pada Sistem Pemantauan Status Gizi Bayi. *Jurnal Ilmiah Teknik Elektro*, 12, 63–72.
- Janah, F. R., Kurniawan, R., & Suprpti, T. (2023). Analisis Dataset Status Gizi Pada Balita Menggunakan Algoritma K-Means Clustering. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3602–3609.
- 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.
- Kristian, S., Gultom, S., & Samosir, S. (2023). Analisis Data Keterlambatan Pesawat Menggunakan Algoritma K-Means. *Jurnal Ilmu Komputer dan Teknologi Informasi*, 1(1), 1–11.
- Nuzula, R. F., Arfan, N. A., & Ningrum, S. (2023). Peran Kader Terhadap Upaya Peningkatan Status Gizi Balita Di Posyandu. *Jurnal Kesehatan Samodra Ilmu*, 14(01), 18–21.
- Qomarudin, M., Robby, R., & Akbarita, R. (2022). Pelatihan Matlab Guna Meningkatkan Kemampuan Komputasi. *AMONG Jurnal Pengabdian Masyarakat*, 4(1), 33–38.
- Raudotul Janah, F., Kurniawan, R., & Suprpti, T. (2024). Analisis Dataset Status Gizi Pada Balita Menggunakan Algoritma K-Means Clustering. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3602–3609.
- Sekar Setyaningtyas, B. I. N., & Arif, Z. (2022). Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means. *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, 10(2), 52–61.
- Sinaga, E. S., Rasyid, I. A., Mubarak, M. R., Sudharma, N. I., & Nolia, H. (2023). Pemantauan Konsumsi Pemberian Makanan Tambahan (PMT) Dalam Meningkatkan Berat Badan Balita Dengan Masalah Gizi. *ABDI MOESTOPO: Jurnal Pengabdian Pada Masyarakat*, 6(1), 1–8.
- Surbakti, N. K. (2021). Data Mining Pengelompokan Pasien Rawat Inap Peserta BPJS Menggunakan Metode Clustering (Studi Kasus: RSUD. Bangkatan). *Journal of Information and Technology*, 1(2), 47–53.
- Wondal, R., Mahmud, N., Purba, N., Budiarti, E., Arfa, U., & Oktaviani, W. (2023). Deskripsi Status Gizi Balita, Serta Partisipasi Orang Tua Pada Masa Pandemi Covid-19. *Jurnal Obsesi: Jurnal Pendidikan Anak Usia Dini*, 7(1), 345–357.
- Queyrut, S., Bromberg, Y.-D., & Schiavoni, V. (2022). *Pelta*. 12(1), 37–43.
- Setyo, W. N., & Wardhana, S. (2021). Implementasi Data Mining Pada Penjualan Produk Di Cv Cahaya Setya Menggunakan Algoritma Fp-Growth. *Petir*, 12(1), 54–63.