

# Implementasi Algoritma K-Medoids dan Aplikasi RapidMiner dalam Pengelompokan Kasus BALITA Stunting

Dian Permata Sari

Program Studi Sistem Informasi, STMIK Jayanusa, Indonesia

E-mail: langitbiru621@gmail.com

## Abstract

*Stunting is a condition where a child's growth is disrupted, one of which is characterized by the child's growth being abnormally tall, so that it is lower than children his age. Genetic factors have a very small influence on stunting cases compared to environmental factors and health services. In the Tarab River area itself, the biggest factors causing stunting in children are poor nutrition, economics and latrines. In this study the author tried to group stunting cases in the Sungai Tarab II community health center using the k-medoids method, the variables used were age, economic status, latrine conditions and complete or incomplete immunization. The results of research using the k-medoids method which carried out manual calculations obtained 2 clusters, namely cluster 1 and cluster 2 and the results could be obtained in iteration 1 where the difference was 1.112854. As for cluster 1, children are categorized as not being in the stunting category, while cluster 2 children are categorized as stunting.*

**Keywords:** Data mining, K-Medoids, Toddlers, Stunting

## Abstrak

*Stunting merupakan suatu kondisi dimana terganggunya pertumbuhan anak salah satunya ditandai dengan pertumbuhan tinggi anak yang tidak normal, sehingga lebih rendah dibandingkan anak-anak seusianya. Factor genetic memiliki pengaruh yang sangat kecil untuk kasus stunting dibandingkan dengan factor lingkungan dan pelayanan kesehatan. Di daerah Sungai Tarab sendiri factor terbesar penyebab anak stunting adalah gizi buruk, ekonomi dan jamban. Pada penelitian ini penulis mencoba mengelompokkan kasus stunting di puskesmas sungai tarab II menggunakan metode k-medoids, adapun variabel yang digunakan adalah usia, status ekonomi, kondisi jamban dan imunisasi lengkap atau tidak.lengkap. Hasil penelitian dengan metode k-medoids yang melakukan perhitungan secara maunal diperoleh 2 cluster, yaitu cluster 1 dan cluster 2 dan hasil sudah dapat diperoleh pada iterasi 1 dimana selisih nya 1,112854. Adapun cluster 1 dikategorikan anak tidak masuk dalam kategori stunting sedangkan cluster 2 anak dikategorikan stunting.*

**Kunci :** Data mining, K-Medoids, Balita, Stunting

## 1. Pendahuluan

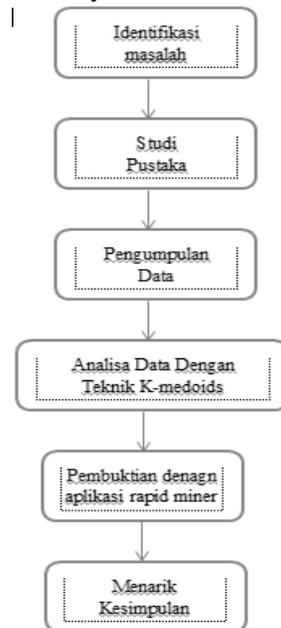
Kasus stunting di Indonesia menjadi prioritas utama kementerian kesehatan untuk mengatasinya hal ini dikarenakan masih tingginya kasus stunting di Indonesia[1]. Penyebab utama dari stunting adalah gizi buruk sehingga anak tumbuh menjadi pendek dibandingkan usia normalnya [2]. Pada usia 1000 hari diawal kehidupan anak merupakan masa yang sangat rentan dengan gizi buruj jika suapan makanannya kurang terpenuhi akibatnya anak rentan terserang penyakit [3] serta berpotensi terhambat perkembangan tumbuh secara fisik, perkembangan kognitif yang berefek kepada perkembangan kecerdasan anak dimasa mendatang[4]. K-medoids merupakan salah satu teknik data mining dimana menggunakan

objek perwakilan yang disebut medoids yang merupakan titik pusat centroid [5]. K-medoids adalah metode partisi yang menggabungkan  $n$  objek kedalam  $k$  cluster [6] dan melakukan partisi dengan cara meminimalkan ketidakmiripan antar setiap objek. Teknik K-medoids merupakan metode pengelompokan pengembangan dari metode k-means [7]. Perbedaan daro kedua teknik ini adalah k-medoids menggunakan objek sebagai perwakilan (medoid) sebagai cluster dari untuk detisp cluster sedangkan k-means menggunakan nilai rata-rata sebagai pusat cluster. Teknik K-medoids sudah banyak digunakan dalam penelitian, diantaranya yang berjudul Analisis metode K-medoids berdasarkan indikator kemiskinan di daerah jawa timr. Diperoleh hasil untuk cluster 1 sebanyak 30 kota/kabupaten termasuk kategori rumah dengan sanitasi layak sedangkan cluster 2 sebanyak 8 kota/kabuoaten tergolong rumah dengan sanitasi tidak layak [6].

Penelitian yang dilakukan oleh[8] memanfaatkan teknik k-medoids dalam menagani informasi promosi yang tidak merata dengan cara melakukan pengelompokan kepada sipenerima informasi dalam hal ini mahasiswa serta asal daerahnya. Penelitian selanjtnya dengan menggunakan metode Naïve Bayes Classifier dengan membagi sampel data secara acak dan mengelompokkan data tersebut. Klasifikasi status gizi balita stunting dengan Naïve Bayes Classifier dengan pengujian K-Fold Cross Validation [2]. Berdasarkan permasalahan diatas penulis ingin melakukan penelitian pengelompokkan balita stunting dengan teknik k-medoids .

## 2. Metodologi Penelitian

Dibawah ini adalah tahapan yang penulis gunakan dalam penelitian mulai dari identifikasi masalah sampai memperoleh hasil akhirnya:



**Gambar 1.** Kaidah Penelitian

Tahap pertama dalam penelitian ini adalah melakukan identifikasi masalah yaitu menentukan variable. Adapun variable nya adalah usia, status ekonomi, kondisi jamban dan imunisasi. Kemudian mengumpulkan buku serta jurnal-jurnal referensi yang dapat dijadikan rujukan dalam penelitian ini. Kemudian mengumpulkan data dengan cara meminta langsung data stunting ke puskesmas sungai tarab II serta wawancara dengan pihak terkait. Setelah data

diperoleh maka langkah selanjutnya mengolah data dengan teknik k-medoids. Aplikasi rapidminer digunakan sebagai pembuktian sekaligus perbandingan dengan perhitungan manual menggunakan teknik k-medoids. Tahap terakhir dalam penelitian ini adalah menarik kesimpulan dari hasil penelitian yang telah dilakukan.

Algoritma K-Medoids merupakan suatu metode partisi clustering untuk mengelompokkan sekumpulan (n) objek menjadi sejumlah (k) cluster. Algoritma K-Medoids menggunakan objek pada kumpulan objek untuk mewakili sebuah cluster.[9] Objek yang terpilih untuk mewakili sebuah cluster disebut medoid. Dibawah ini tahapan dalam perhitungan algoritma K-Medoids diantaranya [10]:

- 1) Inisialisasikan pusat cluster sebanyak jumlah cluster (k).
- 2) Alokasikan setiap data (objek) ke cluster terdekat menggunakan persamaan ukuran jarak Euclidian Distance dengan persamaan:

$$d(x_{ij}, c_{kj}) = \sqrt{\sum_{j=i}^p \sum_{i=1}^n (x_{ij} - c_{kj})^2} \tag{1}$$

- 3) Pilih secara acak objek pada masing-masing cluster sebagai kandidat medoid baru.
- 4) Hitung jarak setiap objek yang berada pada masing-masing cluster dengan kandidat medoids baru.
- 5) Hitung total simpangan (S) dengan menghitung nilai total distance baru – total distance lama. Jika  $S < 0$ , maka tukar objek dengan data cluster untuk membentuk sekumpulan k objek baru sebagai medoid.
- 6) Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan medoids, sehingga didapatkan cluster beserta anggota masing-masing cluster.

### 3. Hasil Dan Pembahasan

Data yang digunakan dalam penelitian ini adalah data pasien anak tahun 2022 dengan sampel data yang digunakan sebanyak 20 sampel. Adapun data mentah sebelum dinormalisasi seperti Tabel 1:

**Tabel 1.** Sampel Data Pasien Stunting Tahun 2022

No	Inisial Nama	Usia	Status Ekonom	Jamban	Imunisasi
1	VR	57	Mampu	Sehat	Lengkap
2	M5	37	Mampu	Tidak Sehat	Lengkap
3	BQ	40	Tidak Mampu (bekerja)	Sehat	Tidak Lengkap
4	AN	49	Mampu	Sehat	Lengkap
5	AG	58	Tidak Mampu (bekerja)	Tidak Sehat	Lengkap
6	AV	19	Tidak Mampu (tidak bekerja)	Sehat	Lengkap
7	GA	35	Mampu	Sehat	Lengkap
8	AG	57	Tidak Mampu (bekerja)	Tidak Sehat	Tidak Lengkap
9	FD	16	Tidak Mampu (bekerja)	Sehat	Lengkap
10	AG	54	Tidak Mampu (bekerja)	Tidak Sehat	Lengkap
11	HZ	34	Tidak Mampu (tidak bekerja)	Sehat	Lengkap
12	ANG	15	Tidak Mampu (bekerja)	Sehat	Lengkap
13	FA	56	Mampu	Sehat	Lengkap
14	AF	58	Mampu	Sehat	Lengkap
15	AM	30	Tidak Mampu (bekerja)	Tidak Sehat	Lengkap
16	MFR	18	Tidak Mampu (bekerja)	Sehat	Tidak Lengkap
17	BN	35	Tidak Mampu (tidak bekerja)	Tidak Sehat	Tidak Lengkap
18	AR	36	Tidak Mampu (bekerja)	Sehat	Lengkap
19	AA	22	Tidak Mampu (bekerja)	Sehat	Lengkap
20	NJ	23	Tidak Mampu (bekerja)	Sehat	Lengkap

Dari data yang terdapat pada table diatas langkah pertama dalam melakukan perhitungan dengan teknik k-medoids adalah melakukan cleaning data tujuannya adalah menghilangkan data yang tidak relevan, jika semua data telah relevan maka lakukan transformasi data, karena dalam teknik k-medoids data yang di proses adalah berupa data numeric atau angka seperti pada Tabel 2:

**Tabel 2.** Hasil dari Transformasi Data Pasien Stunting

No	Inisial Nama	Usia	Status Ekonomi	Jamban	Imunisasi
1	VR	57	1	1	1
2	MS	37	1	2	1
3	BQ	40	2	1	2
4	AN	49	1	1	1
5	AG	58	2	2	1
6	AV	19	3	1	1
7	GA	35	1	1	1
8	AG	57	2	2	2
9	FO	16	2	1	1
10	AG	54	2	2	1
11	HZ	34	3	1	1
12	ANG	15	2	1	1
13	FA	56	1	1	1
14	AF	58	1	1	1
15	AM	30	2	2	1
16	MR	18	2	1	2
17	BN	35	3	2	2
18	AR	36	2	1	1
19	AA	22	2	1	1
20	NJ	23	2	1	1

Tahap selanjutnya adalah melakukan normalisasi data. Pada penelitian ini varibael yang digunakan dalam pengelompokkan kasus stunting adalah usia, status ekonomi, jamban dan imunisasi maka data yang dinormmalisasikan adalah varibael yang digunakan, seperti pada tabel di bawah ini:

**Tabel 3.** Hasil Normalisasi Data

Usia	Status Ekonomi	Jamban	Imunisasi
0.977	0.000	0.000	0.000
0.512	0.000	1.000	0.000
0.581	0.500	0.000	1.000
0.791	0.000	0.000	0.000
1.000	0.500	1.000	0.000
0.093	1.000	0.000	0.000
0.465	0.000	0.000	0.000
0.977	0.500	1.000	1.000
0.023	0.500	0.000	0.000
0.907	0.500	1.000	0.000
0.442	1.000	0.000	0.000
0.000	0.500	0.000	0.000
0.953	0.000	0.000	0.000
1.000	0.000	0.000	0.000
0.349	0.500	1.000	0.000
0.070	0.500	0.000	1.000
0.465	1.000	1.000	1.000
0.488	0.500	0.000	0.000
0.163	0.500	0.000	0.000
0.186	0.500	0.000	0.000

Dalam menganalisis kebutuhan menggunakan teknik k-medoids yaitu mengelompokkan data berdasarkan nilai rata-rata yang diperoleh dari penelitian maka dibutuhkan cluster serta menentukan titik awal cluster. Pada penelitian ini penulis menggunakan 2 cluster dalam pengelompokan data. Adapun untuk titik awal cluster dipilih secara acak, dan penulis memilih data 7 dan 15 sebagai pusat awal cluster.

**Tabel 4.** Penentuan pusat awal cluster

	Usia	Status Ekc	Jamban	Imunisasi
C1	0.465	0.000	0.000	0.000
C2	0.349	0.500	1.000	0.000

Aplikasikan setiap data (objek) ke cluster terdekat menggunakan persamaan ukuran jarak Euclidean Distance. Berikut adalah perhitungan persamaan jarak Euclidean Distance pada data kasus balita yang mengalami stunting :

$$D_{vr1} = \sqrt{(0,465 - 0,977)^2 + (0,000 - 0,000)^2 + (0,000 - 0,000)^2 + (0,000 - 0,000)^2} = 0.511627907$$

$$D_{vr2} = \sqrt{(0,349 - 0,977)^2 + (0,500 - 0,000)^2 + (1,000 - 0,000)^2 + (0,000 - 0,000)^2} = 1.282289816$$

$$D_{ms1} = \sqrt{(0,465 - 0,512)^2 + (0,000 - 0,000)^2 + (0,000 - 1,000)^2 + (0,000 - 0,000)^2} = 1.001081081$$

$$D_{ms2} = \sqrt{(0,349 - 0,512)^2 + (0,500 - 0,000)^2 + (1,000 - 1,000)^2 + (0,000 - 0,000)^2} = 1.282289816$$

⋮  
⋮  
⋮

$$D_{nj1} = \sqrt{(0,465 - 0,186)^2 + (0,000 - 0,500)^2 + (0,000 - 1,000)^2 + (0,000 - 0,000)^2} = 0.572608012$$

$$D_{nj2} = \sqrt{(0,349 - 0,186)^2 + (0,500 - 0,500)^2 + (1,000 - 1,000)^2 + (0,000 - 0,000)^2} = 1.013163763$$

Hasil perhitungan jarak ke setiap k-medoids awal dapat dilihat pada tabel di bawah ini:

**Tabel 5.** Hasil perhitungan iterasi Ke-1

Nama	Jarak k- medoids		Kedekatan	Cluster
	C1	C2		
VR	0.511628	1.28229	0.511627907	C1
MS	1.001081	0.525833	0.525833444	C2
BQ	1.124064	1.433207	1.124064421	C1
AN	0.325581	1.202182	0.325581395	C1
AG	1.239395	0.651163	0.651162791	C2
AV	1.066983	1.146927	1.066983232	C1
GA	0	1.124064	0	C1
AG	1.584854	1.180791	1.180790909	C2
FO	0.667264	1.051667	0.667263569	C1
AG	1.202182	0.55814	0.558139535	C2
HZ	1.00027	1.121897	1.00027038	C1
ANG	0.682886	1.059097	0.6828859	C1
FA	0.488372	1.271064	0.488372093	C1

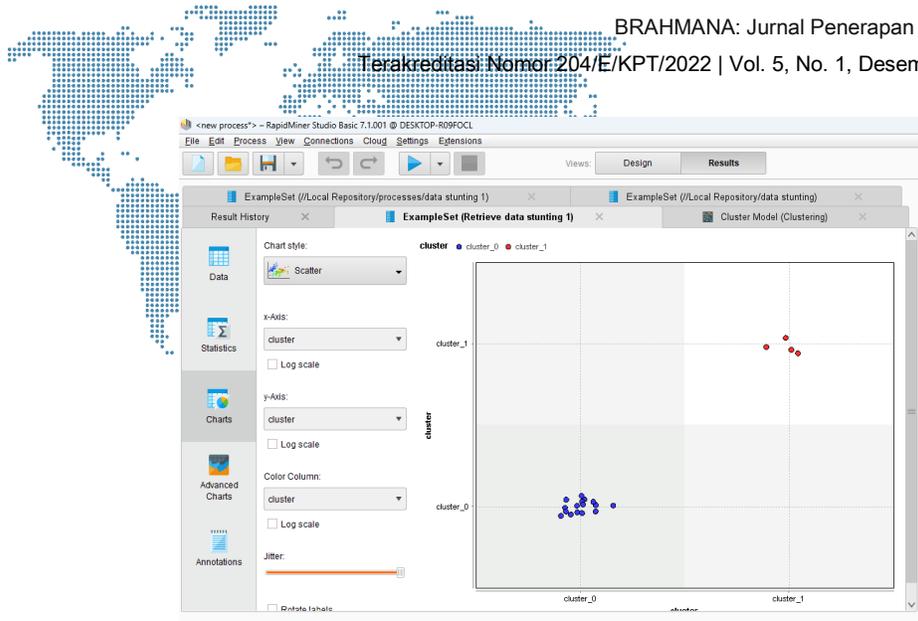
Nama	Jarak k-medoids		Kedekatan	Cluster
AF	0.534884	1.293837	0.534883721	C1
AM	1.124064	0	0	C2
MR	1.185876	1.441485	1.185875501	C1
BN	1.732051	1.124064	1.124064421	C2
AR	0.500541	1.009688	0.500540541	C1
AA	0.584295	1.017159	0.584295094	C1
NJ	0.572608	1.013164	0.572608012	C1
<b>Total cos</b>			<b>13.28524287</b>	

Setelah diperoleh jarak dari setiap objek pada iterasi 1 maka dilanjutkan melakukan perhitungan iterasi 2,3...n hingga memperoleh hasil selisih antar total cost iterasi lebih dari 0. Jika hasil selisih besar dari 0 maka proses perhitungan dihentikan. Hasil cluster terakhir dapat dilihat pada tabel dibawah ini:

**Tabel 6.** Hasil Cluster

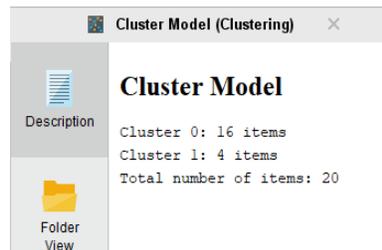
Nama	Cluster Terdekat
VR	C1
MS	C2
BQ	C1
AN	C1
AG	C2
AV	C1
GA	C1
AG	C2
FO	C1
AG	C2
HZ	C1
ANG	C1
FA	C1
AF	C1
AM	C2
MR	C1
BN	C2
AR	C1
AA	C1
NJ	C1

Pengaplikasikan Rapidminer dalam pengelompokan data pasien stunting menggunakan teknik k-medoids. Data yang diinputkan dalam aplikasi rapidminer sebelumnya telah disimpan di ms excel, terdapat lima kolom dari tabel data pasien yaitu nama, usia, status ekonomi, jamban dan imunisasi. Hasil akhir dapat diperoleh seperti pada gambar grafik di bawah ini, yang mana menunjukkan ada 4 cluster yang termasuk dalam cluster 2 ditandai dengan warna merah dan 16 termasuk dalam cluster 1 ditandai dengan warna biru sesuai dengan perhitungan yang dilakukan secara manual menggunakan teknik k-medoids.



**Gambar 2.** Hasil pengolahan data berupa grafik

Hasil pengujian lain dapat dilihat pada gambar 4 di bawah ini, yang mana diperoleh hasil pengolahan datanya untuk cluster 1 sebanyak 16 data yang mana termasuk kategori balita yang tidak stunting sedangkan clster 2 sebanyak 4 data termasuk pada kasus balita stunting. Total data yang digunakan sebagai sampel perhitungan dalam penelitian ini sebanyak 20 data.



**Gambar 3.** Cluster Model

ExampleSet (20 examples, 0 special attributes, 5 regular attributes)

Row No.	Nama	Usia	Status Ekon...	Jamban	Imunisasi
1	VR	0.977	0	0	0
2	MS	0.512	0	1	0
3	BQ	0.581	0.500	0	1
4	AN	0.791	0	0	0
5	AG	1	0.500	1	0
6	AV	0.093	1	0	0
7	GA	0.465	0	0	0
8	AG	0.977	0.500	1	1
9	FO	0.023	0.500	0	0
10	AG	0.907	0.500	1	0
11	HZ	0.442	1	0	0
12	ANG	0	0.500	0	0
13	FA	0.953	0	0	0
14	AF	1	0	0	0
15	AM	0.349	0.500	1	0

**Gambar 4.** Tabel hasil pengelompokkan data stunting

#### 4. Kesimpulan

Data mining dengan teknik k-medoids dinilai mampu melakukan pengelompokan data pasien stunting dengan menggunakan dua cluster yaitu cluster 1 untuk kasus tidak stunting dan cluster 2 untuk kasus stunting. Adapun variabel yang digunakan dalam perhitungan dalam menentukan kasus stunting pada penelitian ini adalah usia, status ekonomi, kondisi jamban dan imunisasi lengkap atau tidak. Agar hasil perhitungan yang dilakukan dengan teknik k-medoids valid maka dilakukan pengujian dengan aplikasi RapidMiner dan diperoleh hasil yang sama antara proses perhitungan menggunakan software RapidMiner dengan perhitungan manual menggunakan teknik k-medoids. Untuk hasil akhir diperoleh cluster 1 sebanyak 16 data dengan kategori tidak stunting dan cluster 2 sebanyak 4 data dan termasuk pada kasus stunting dengan total sampel data yang digunakan dalam penelitian ini berjumlah 20 data sampel.

#### Daftar Pustaka

- [1] Y. Yuwanti, F. M. Mulyaningrum, and M. M. Susanti, "Faktor – Faktor Yang Mempengaruhi Stunting Pada Balita Di Kabupaten Grobogan," *J. Keperawatan dan Kesehatan Masy. Cendekia Utama*, vol. 10, no. 1, p. 74, 2021, doi: 10.31596/jcu.v10i1.704.
- [2] R. R. R. Arisandi, B. Warsito, and A. R. Hakim, "Aplikasi Naïve Bayes Classifier (Nbc) Pada Klasifikasi Status Gizi Balita Stunting Dengan Pengujian K-Fold Cross Validation," *J. Gaussian*, vol. 11, no. 1, pp. 130–139, 2022, doi: 10.14710/j.gauss.v11i1.33991.
- [3] S. Narulita, P. Prihati, A. T. Oktaga, and ..., "Performansi Algoritma Clustering K-Means untuk Penentuan Status Malnutrisi pada Balita," *J. Informasi, Sains ...*, vol. 06, no. 1, 2023, [Online]. Available: <https://isaintek.polinef.ac.id/index.php/isaintek/article/view/128%0Ahttps://isaintek.polinef.ac.id/index.php/isaintek/article/download/128/124>
- [4] A. Fadilah, M. N. Pangestu, S. Lumbanbatu, and S. Defiyanti, "Pengelompokan Kabupaten/Kota Di Indonesia Berdasarkan Faktor Penyebab Stunting Pada Balita Menggunakan Algoritma K-Means," *JIKO (Jurnal Inform. dan Komputer)*, vol. 6, no. 2, p. 223, 2022, doi: 10.26798/jiko.v6i2.581.
- [5] I. Puspita, M. N. Hayati, A. Nohe, L. S. Terapan, and U. Mulawarman, "Pengelompokan Puskesmas Berdasarkan Kasus Balita Stunting di Kabupaten Paser Menggunakan Metode K-medoids Grouping Community Health Centers Based on Stunting Toddler Cases in Paser Regency Using the K-medoids Method," vol. 14, pp. 1–10, 2023.
- [6] F. Alfiah, A. Almadayani, D. Al Farizi, and E. Widodo, "Analisis Clustering K-Medoids Berdasarkan Indikator Kemiskinan di Jawa Timur Tahun 2020," *J. Ilm. Sains*, vol. 22, no. 1, p. 1, 2021, doi: 10.35799/jis.v22i1.35911.
- [7] H. Pohan, M. Zarlis, E. Irawan, H. Okprana, and Y. Pranayama, "Penerapan Algoritma K-Medoids dalam Pengelompokan Balita Stunting di Indonesia," *JUKI J. Komput. dan Inform.*, vol. 3, no. 2, pp. 97–104, 2021, doi: 10.53842/juki.v3i2.69.
- [8] Novita Lestari Anggreini, "Teknik Clustering Dengan Algoritma K-Medoids Untuk Menangani Strategi Promosi Di Politeknik Tede Bandung," *J. Teknol. Inf. dan Pendidik.*, vol. 12, no. 2, pp. 1–7, 2019, [Online]. Available: <http://tip.ppj.unp.ac.id>
- [9] D. A. Alodia, A. P. Fialine, D. Endriani, and E. Widodo, "Implementasi Metode K-Medoids Clustering untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan," *Sepren*, vol. 2, no. 2, pp. 1–13, 2021, doi: 10.36655/sepren.v2i2.606.
- [10] S. Nurlaela, A. Primajaya, and T. N. Padilah, "Algoritma K-Medoids Untuk Clustering Penyakit Maag Di Kabupaten Karawang," *INFORMATIKA*, vol. 12, no. 2, p. 56, 2020, doi: 10.36723/juri.v12i2.234.