

IMPLEMENTASI METODE K-MEANS PADA KELOMPOK BALITA GIZI BURUK

Iqbal Hanan Junaidi¹, Sandy Yustisio Oktaviandra², Rizki Hendra Kusuma^{3*}, Dwi Hartanti⁴

^{1,2,3,4}Informatika

Universitas Duta Bangsa Surakarta

¹202020345@mhs.udb.ac.id, ²202021187@mhs.udb.ac.id, ^{3*}202030381@mhs.udb.ac.id, ⁴dwhartanti@udb.ac.id

Abstrak— Gizi merupakan unsur penting dalam menjaga pertumbuhan, perkembangan, dan kesehatan seseorang. Untuk mengidentifikasi daerah dengan tingkat gizi buruk pada anak balita, dapat menggunakan metode perhitungan. Dalam penelitian ini, digunakan analisis perhitungan data mining sebagai metode perhitungan. Data mining yaitu metode pemrosesan data yang bertujuan menemukan data dengan akurat, yang mudah dipahami. Clustering digunakan untuk mengelompokkan data ke dalam kelompok-kelompok berdasarkan kesamaannya. Proses clustering dilakukan untuk mengelompokkan catatan, observasi, atau kelas yang memiliki objek yang sama. Satu diantara metode data mining yang digunakan adalah K-Means Clustering yang dapat dilakukan tanpa supervisi dan termasuk teknik partisi untuk mengelompokkan data dengan akurasi tinggi.

Kata kunci— Data Mining, Clustering, K-Means, Gizi

Abstract— Nutrition is an important element in maintaining one's growth, development and health. To identify areas with malnutrition levels in children under five, the calculation method can be used. In this study, data mining calculation analysis is used as a calculation method. Data mining is the process of collecting and processing data to extract important information. Clustering is used to group data into groups based on their similarity. The clustering process is carried out to group records, observations, or classes that have the same object. One of the data mining techniques used is K-Means Clustering which can be done without supervision and includes partitioning techniques to group data with high accuracy.

Keywords— Data Mining, Clustering, K-Means, Nutrition

I. PENDAHULUAN

Gizi adalah unsur gizi yang penting untuk pertumbuhan, perkembangan dan pemeliharaan kesehatan seseorang. Gizi seimbang mengacu pada kandungan gizi dari makanan sehari-hari yang dikonsumsi oleh tubuh yang dibutuhkan, termasuk kondisi kesehatan, jenis kelamin, dan usia. Pola makan yang tidak teratur dapat menyebabkan kekurangan nutrisi seperti kekurangan berat badan dan anemia. Di sisi lain, kelebihan nutrisi juga dapat menjadi risiko jika nutrisi tidak seimbang dengan baik[1].

Untuk menghitung jumlah daerah dengan tingkat gizi buruk balita, dapat dianalisis dengan metode perhitungan. Di penelitian ini, digunakan analisis perhitungan data mining sebagai metode perhitungannya. Data mining yaitu metode pemrosesan data yang bertujuan menemukan data dengan akurat, yang mudah dipahami.

Tujuan data mining adalah untuk menemukan pola penting yang tidak diketahui keberadaannya dalam data yang besar dan kompleks. Clustering adalah salah satu teknik pada data mining, yaitu teknik pengelompokan data yang belum didefinisikan label

kelasnya sebelum dijalankan oleh tool data mining [2].

Clustering mengklasifikasikan data ke dalam sejumlah kelompok tertentu berdasarkan kemiripannya. Proses clustering dilakukan untuk mengelompokkan record, observasi, atau kelas yang memiliki objek yang sama[2][3].

Satu diantara metode data mining adalah K-Means Clustering yang dilakukan permodelan tidak harus supervisi serta termasuk satu teknik partisi untuk mengelompokkan data. Tata cara ini pula diketahui selaku tata cara non-hirarki yang mengelompokkan informasi ke dalam satuatau lebih cluster atau kelompok[4][3][5].

Dengan demikian, penelitian yang telah diteliti oleh peneliti dengan judul "Implementasi Metode K-Means Pada Kelompok Balita Gizi Buruk". Tujuan data mining adalah untuk menemukan data dengan akurat, yang mudah dipahami.

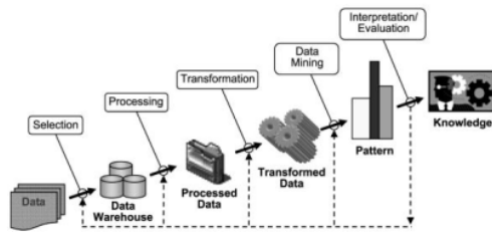
II. TINJAUAN STUDI

A. Data Mining

Data mining yaitu metode pemrosesan data yang bertujuan menemukan data dengan akurat, yang

mudah dipahami. Tahapan dari data mining antara lain :

1. Pembersihan
2. Integrasi
3. Transformasi
4. Data Mining
5. Evaluasi
6. Knowledge



Gambar 23. Tahapan data mining

B. Clustering

Clustering adalah salah satu teknik pada data mining, yaitu teknik pengelompokan atribut yang belum didefinisikan label kelasnya sebelum dijalankan oleh tool data mining [2]. Clustering mengklasifikasikan data ke dalam sejumlah kelompok tertentu berdasarkan kemiripannya. Proses clustering dilakukan untuk mengelompokkan record, observasi, atau kelas yang memiliki objek yang sama[2][3].

C. K-Means Method

K-Means Method adalah tahap non-hierarki untuk bertujuan mengelompokkan data ke cluster. Data yang sesuai karakteristik akan digabungkan menjadi satu kelompok dan informasi dengan karakteristik yang berbeda dikelompokkan dengan kelompok lain agar data dari satu kelompok memiliki variabilitas yang rendah[4][6]. Algoritma K-Means adalah teknik yang bekerja atas dasar pembagian ke dalam kelompok, sehingga prinsip pengelompokan dilakukan secara bertahap[7].

D. RapidMiner

RapidMiner adalah perangkat lunak yang dirancang untuk memudahkan pengguna menggunakan fitur perangkat lunak ini. Hasil yang ditampilkan oleh RapidMiner juga dapat divisualisasikan dengan grafik, menjadikan RapidMiner pilihan untuk mengekstraksi informasi menggunakan metode data mining[10].

III. METODOLOGI PENELITIAN

A. Pengumpulan data

Penelitian ini peneliti menggunakan dataset publik yang diperoleh dari portal resmi nasional yaitu Satu Data Indonesia <https://data.go.id>. Data tentang jumlah status gizi balita berdasarkan indeks balita gizi buruk di Jawa Barat.

B. Data Selection

Seleksi Data adalah proses pemilihan data yang diperlukan dari kumpulan data operasional sebelum dilakukan sesi penggalian data pada KDD. [8]. Jadi, hanya beberapa data relevan dengan penelitian yang kami diambil. Metode ini adalah cara yang dilakukan pemilihan atribut, dimana peneliti akan mengambil atribut mana yang ingin peneliti gunakan pada data mining. Data ini memiliki 8 atribut yaitu kode_provinsi, nama_provinsi, kode_kabupaten_kota, nama_kabupaten_kota, kategori_gizi_buruk, jumlah_balita, satuan dan tahun. Dan yang akan dipilih oleh peneliti ada 3 data atribut, yaitu nama kabupaten kota, jumlah_balita serta tahun.

C. Data Transformation

K- Means Clustering merupakan tata cara yang dapat dicoba apabila informasi yang dipakai merupakan informasi berbentuk angka, hingga proses transformasi data mungkin diperlukan. Proses transformasi sendiri adalah metode dimana untuk mengubah atribut data non-numerik menjadi nilai numerik sehingga algoritma clustering K-Means dapat menangani data ini[9].

D. K-Means Method

K-Means Method adalah tahap non-hierarki untuk bertujuan mengelompokkan data ke cluster. Data yang sesuai karakteristik akan digabungkan menjadi satu kelompok dan informasi dengan karakteristik yang berbeda dikelompokkan dengan kelompok lain agar data dari satu kelompok memiliki variabilitas yang rendah[4][6]. Algoritma K-Means adalah teknik yang bekerja atas dasar pembagian ke dalam kelompok, sehingga prinsip pengelompokan

dilakukan secara bertahap[7]. Metode dari K-Means adalah:

1. Menetapkan kelompok yang akan disesuaikan dengan data.
2. Memindahkan atau menggabungkan data ke cluster.
3. Menghitung centroid dari data yang ada setiap cluster.
4. Tetapkan setiap data ke nilai rata yang paling dekat.
5. Mengulang ke langkah 3, jika terdapat atribut yang bertukar kelompok maka perhitungan nilai rata dilanjutkan, dan atribut kelompok yang konstan perhitungan di berhentikan.

E. Evaluasi

Pada tahap ini merupakan langkah terakhir untuk menggali dan mengevaluasi temuan atau hasil dari proses penelitian, menyajikan data yang mudah dibaca dan dipahami bagi yang membacanya.

IV. HASIL DAN PEMBAHASAN

Dalam penelitian ini, peneliti menggunakan data dari Satu Indonesia Data. Selanjutnya data dapat diolah menggunakan software RapidMiner dengan metode K-Means serta data balita gizi buruk yang cukup banyak.

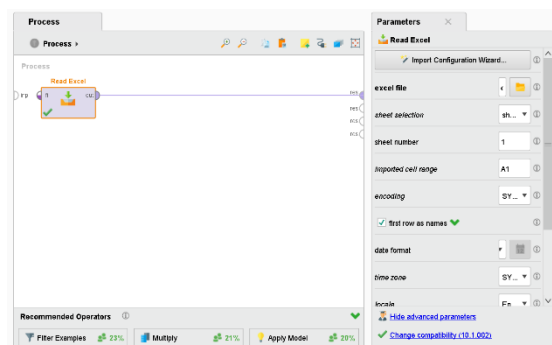
Tabel 1. Dataset Penelitian

Kabupaten Kota	Jumlah Balita	Tahun
Bogor	16821	2021
Sukabumi	6543	2021
Cianjur	5146	2021
Bandung	9596	2021
Garut	5150	2021
Tasikmalaya	4606	2021
Ciamis	1977	2021
Kuningan	2461	2021
Cirebon	7876	2021
Majalengka	2471	2021
Sumedang	2381	2021
Indramayu	5973	2021
Subang	2390	2021
Purwakarta	1228	2021
Karawang	2914	2021
Bekasi	7548	2021

Bandung Barat	5164	2021
Pangandaran	347	2021
Kota Bogor	2824	2021
Kota Sukabumi	621	2021
Kota Bandung	3941	2021
Kota Cirebon	1108	2021
Kota Bekasi	7343	2021
Kota Depok	2397	2021
Kota Cimahi	1221	2021
Kota Tasikmalaya	2477	2021
Kota Banjar	727	2021

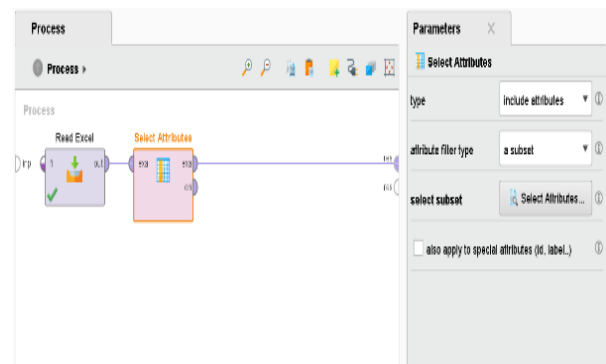
A. Data Selection

Langkat pertama adalah data akan dibaca dengan menggunakan operator *Read Excel*. Selanjutnya kita masukkan datanya.



Gambar 24 Proses Membaca Excel

Selanjutnya memisahkan data antara yang diterapkan serta yang tidak akan diterapkandata menggunakan operator *Select Attributes*. Kemudian atribut yang layak diteliti akan diolah.

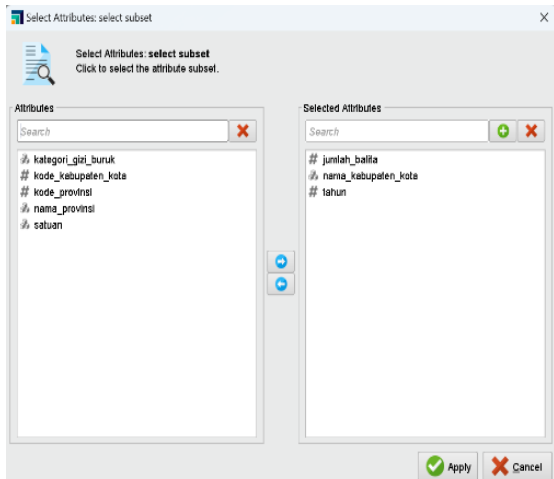


Gambar 25 Proses Seleksi Atribut

Selanjutnya adalah pemilihan data yang akan digunakan untuk penelitian selanjutnya. Disini

penulis memilih data dari jumlah_balita, nama kabupaten kota dan tahun.

digunakan untuk menggunakan angka. Langkah ini penting karena dijelaskan Algoritma K-Means hanya dapat menangani data numerik.



Gambar 26 Memilih atribut dataset

Hasil Select Attributes yang terdiri dari 3 atribut yang telah terpilih yaitu nama kabupaten kota, jumlah_balita dan tahun.

Result History ExampleSet (Select Attributes)

Open in Turbo Prep Auto Model

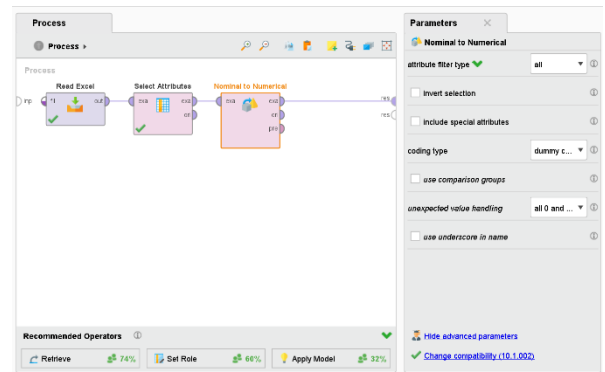
Row No.	nama_kabu...	jumlah_balita	tahun
3	KABUPATEN...	5146	2021
4	KABUPATEN...	9596	2021
5	KABUPATEN...	5150	2021
6	KABUPATEN...	4606	2021
7	KABUPATEN...	1977	2021
8	KABUPATEN...	2461	2021
9	KABUPATEN...	7876	2021
10	KABUPATEN...	2471	2021
11	KABUPATEN...	2381	2021
12	KABUPATEN...	5973	2021
13	KABUPATEN...	2390	2021
14	KABUPATEN...	1228	2021
15	KABUPATEN...	2914	2021

ExampleSet (27 examples, 0 special attributes, 3 regular attributes)

Gambar 27 Hasil data yang dipilih

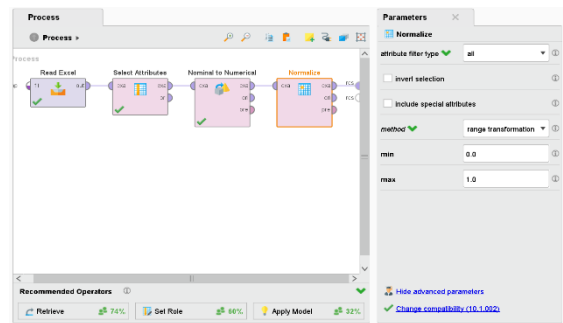
B. Transformasi Data

Data yang telah dipisahkan akan ditransformasikan sesuai dengan algoritma yang digunakan. Menambahkan operator *Nominal to Numerical* untuk tujuan merubah nilai yang



Gambar 28 Operator Nominal to Numerical

Selanjutnya adalah normalisasi data dengan Normalize, yang tujuannya untuk mengurangi gap dari atribut tersebut.



Gambar 29 Operator Normalize

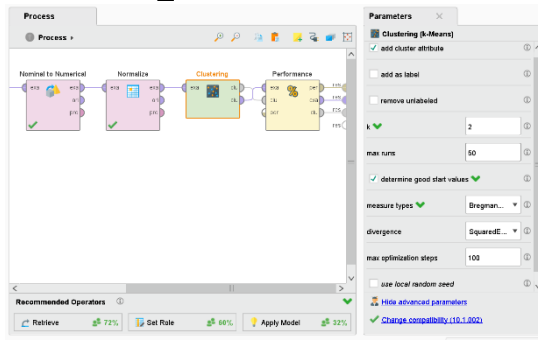
Berikut adalah hasil dari Normalize yang telah diatur dengan metode range transformation dengan parameter min=0.0 dan max=1.0.

nama_kabu...	jumlah_balita	tahun
0	1	1
0	0.376	1
0	0.291	1
0	0.561	1
0	0.292	1
0	0.259	1
0	0.099	1
0	0.128	1
0	0.457	1

Gambar 30 Tabel data hasil normalize

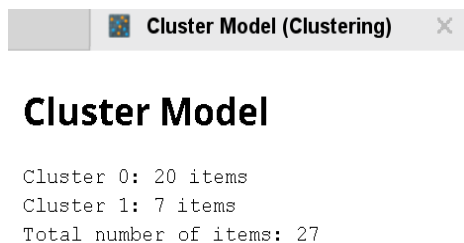
C. K-Means

Kemudian adalah tahap menghitung data mining. Penelitian ini menggunakan clustering atau pengelompokan data dengan metode K-Means dengan nilai $max_runs=50$ dan $k=2$.



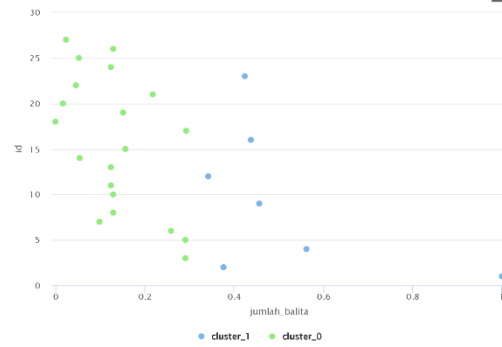
Gambar 31 Operator Clustering K-Means

Dengan menggunakan $k=2$, 2 cluster dari 27 record data dibuat. Cluster 0 dengan total 20 data, sedangkan cluster 1 dengan total 7 data. Dengan demikian, total jumlah data adalah 27 kabupaten di Jawa Barat pada tahun 2021.



Gambar 32 Hasil clustering model

Langkah selanjutnya adalah memvisualisasikan kedalam bentuk diagram scatter sehingga memudahkan untuk membaca data yang telah diolah.



Gambar 33 Diagram scatter 3D

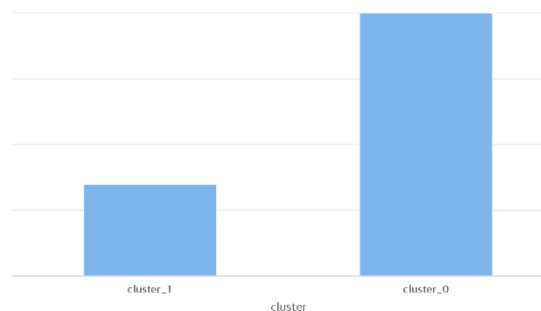
D. Evaluasi

Pada tahap ini, akan menampilkan sekumpulan data yang telah diperoleh dari proses metode data mining K-Means dalam bentuk yang mudah dibaca dan dipahami.

Tabel 2. Tabel Cluster

Cluster	Jumlah Data	Keterangan
0	20	Rendah
1	7	Tinggi

Cluster 0 merupakan tingkat daerah gizi buruk rendah, jumlah data yaitu 20 data kabupaten, sedangkan cluster 1 merupakan tingkat daerah gizi buruk yang tinggi dengan jumlah data yaitu 7 data kabupaten.



Gambar 34 Diagram batang cluster

Data cluster 0 yang merupakan tingkat daerah rendah. 20 Kabupaten yang termasuk kedalam cluster ini adalah dengan data dibawah 5500 balita gizi buruk yaitu Kabupaten Cianjur, Garut, Tasikmalaya, Ciamis, Kuningan, Majalengka, Sumedang, Subang, Purwakarta, Karawang, Bandung, Pangandaran, Bogor, Sukabumi, Kota Kota Bandung, Cirebon, Depok, Cimahi, Tasikmalaya, Banjar pada tahun 2021.

Data cluster 1 yang merupakan cluster tinggi. 7 Kabupaten yang termasuk kedalam cluster ini adalah dengan data diatas 5500 balita gizi buruk yaitu Bogor, Sukabumi, Bandung, Cirebon, Indranayu, Bekasi, Kota Bekasi pada tahun 2021.

V. KESIMPULAN

Hasil dari diskusi penelitian menggunakan metode algoritma K-Means, data Gizi Balita Jawa Barat tahun 2021 dapat dibagi menjadi 2 cluster. Dengan cluster 0 merupakan cluster tingkat daerah gizi buruk rendah dengan jumlah 20 data yaitu Kabupaten yang termasuk kedalam cluster ini adalah dengan data dibawah 5500 balita gizi buruk yaitu Kabupaten Cianjur, Garut, Tasikmalaya, Ciamis, Kuningan, Majalengka, Sumedang, Subang, Purwakarta, Karawang, Bandung, Pangandaran, Bogor, Kabupaten Sukabumi, Kabupaten Bandung, Kabupaten Cirebon, Kabupaten Depok, Kabupaten Cimahi, Kabupaten Tasikmalaya, Kabupaten Banjar pada tahun 2021., sedangkan cluster 1 merupakan cluster tingkat daerah gizi buruk yang tinggi dengan jumlah 7 data yaitu Kabupaten yang termasuk kedalam cluster ini adalah dengan data diatas 5500 balita gizi buruk yaitu Bogor, Sukabumi, Bandung, Cirebon, Indranayu, Bekasi, Kota Bekasi pada tahun 2021..

Peneliti berharap dapat melanjutkan penelitian ini dengan membandingkan hasil dari metode lain. Hasil

studi penelitian dapat menjadi dasar untuk merangkum jumlah balita kurang gizi.

VI. REFERENSI

- [1] M. K-means, O. Purwaningrum, Y. Y. Putra, and A. A. Arifiyanti, "Penentuan Kelompok Status Gizi Balita dengan Menggunakan," vol. 15, no. 2, pp. 129–136, 2021.
- [2] C. A. Rahmat, H. Permatasari, and ..., "Penerapan K-Means Untuk Clustering Kondisi Gizi Balita Pada Posyandu," *J. Media ...*, vol. 7, pp. 207–213, 2023, doi: 10.30865/mib.v7i1.5142.
- [3] A. P. Lestari Br, R. Saragih, and Novriyenni, "Penerapan Metode K-Means Clustering Dalam Pengelompokan Status Kesehatan Ibu Hamil," *J. Inform. Kaputama*, vol. 6, no. 3, 2022.
- [4] W. I. Rahayu, C. Prianto, and E. A. Novia, "Perbandingan Algoritma K-Means Dan Naïve Bayes Untuk Memprediksi Prioritas Pembayaran Tagihan Rumah Sakit Berdasarkan Tingkat Kepentingan Pada Pt. Pertamina (Persero)," *J. Tek. Inform.*, vol. 13, no. 2, pp. 1–8, 2021, [Online]. Available: <https://ejurnal.poltekpos.ac.id/index.php/informatika/article/view/1383>
- [5] M. Y. Matdoan, U. A. Matdoan, and M. Saleh Far-Far, "Algoritma K-Means Untuk Klasifikasi Provinsi di Indonesia Berdasarkan Paket Pelayanan Stunting," *PANRITA J. Sci. Technol. Arts*, vol. 1, no. 2, pp. 41–46, 2022, [Online]. Available: <https://journal.dedikasi.org/pjsta>
- [6] D. Dona and M. Rifqi, "Penerapan Metode K-Means Clustering Untuk Menentukan Status Gizi Baik Dan Gizi Buruk Pada Balita (Studi Kasus Kabupaten Rokan Hulu)," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 7, no. 2, pp. 179–191, 2022, doi: 10.36341/rabit.v7i2.2171.
- [7] V. Syaputri, D. Hartama, F. Anggraini, M. Safii, and R. Dewi, "Penerapan Algoritma K-Means Untuk Menentukan Status Gizi Balita (Studi Kasus: Puskesmas Kecamatan Jawa Maraja Bah Jambi)," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 1, pp. 94–102, 2022, doi: 10.36040/jati.v6i1.4630.
- [8] P. Kepemilikan, A. Kelahiran, and D. I. Jawa, "IMPLEMENTASI ALGORITMA K-MEANS TERHADAP," vol. 18, pp. 1–9, 2023.
- [9] Khairullah, M. H. Rifqo, H. Witriyono, and A. Karolina, "Penerapan Metode K-Means Clustering untuk Pengelompokan Harga Karet di Desa Sengkuang Jaya Provinsi Bengkulu," *J. Kom.*, vol. 2, no. 2, pp. 423–430, 2022.
- [10] 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.