

Clustering Status Gizi Balita menggunakan Metode K-Means pada Posyandu Desa Mekar Wangi

Muhamad Djaelani^{1*}, Martanto², Umi Hayati³

^{1,3} Program Studi Teknik Informatika, STMIK IKMI Cirebon

² Program Studi Manajemen Informatika, STMIK IKMI Cirebon

*Email: muhamaddjaelani047@gmail.com

Abstrak

Kesehatan balita merupakan hal yang sangat penting dalam pembangunan suatu negara. Gizi balita menjadi aspek kunci dalam memastikan pertumbuhan dan perkembangan anak-anak yang sehat. Penelitian ini bertujuan untuk menganalisis Clustering status gizi balita di desa mekar wangi menggunakan algoritma K-Means. Analisis clustering merupakan salah satu metode analisis data mining yang dipengaruhi oleh metode algoritma clustering. Status gizi balita pada posyandu Desa Mekar Wangi dikelompokkan berdasarkan metrik tertentu, seperti berat badan dan tinggi badan, melalui teknik K-Means Clustering. Data mengandung banyak informasi atribut. Setelah data dikumpulkan dan dianalisis, pra-pemrosesan dilakukan untuk menghilangkan data yang tidak valid dan kosong. Hasil analisis clustering menunjukkan bahwa beberapa kelompok balita memiliki status gizi normal, sementara kelompok lain memiliki masalah gizi yang kurang atau lebih. Nilai evaluasi performa Davies Bouldin Index (DBI) yang optimal ditemukan dengan alat RapidMiner dengan $K=2$ dan nilai 0.164 yang mendekati 0 menunjukkan bahwa klaster yang dievaluasi menghasilkan klaster yang baik. Dengan pemahaman yang lebih baik tentang pola gizi balita di Desa Mekar Wangi, petugas Posyandu dapat mengembangkan program yang lebih efisien untuk meningkatkan kualitas gizi anak-anak di Desa Mekar Wangi. petugas Posyandu dapat membantu dalam pengambilan keputusan untuk mengembangkan rekomendasi dan intervensi yang lebih tepat sasaran untuk meningkatkan status gizi balita di desa Mekar Wangi.

Kata kunci: K-Mean Clustering, Gizi Bayi, Posyandu

Abstract

The health of children under five is very important in the development of a country. Toddler nutrition is a key aspect in ensuring the healthy growth and development of children. This study aims to analyze the clustering of nutritional status of toddlers in Mekar Wangi village using the K-Means algorithm. Clustering analysis is a data mining analysis method that is influenced by the clustering algorithm method. The nutritional status of toddlers at the posyandu in Mekar Wangi Village is grouped based on certain metrics, such as body weight and height, using the K-Means Clustering technique. Data contains a lot of attribute information. Once the data is collected and analyzed, pre-processing is performed to remove invalid and empty data. The results of the clustering analysis show that some groups of toddlers have normal nutritional status, while other groups have less or more nutritional problems. The optimal Davies Bouldin Index (DBI) performance evaluation value was found using the RapidMiner tool with $K=2$ and the value of 0.164 which is close to 0 indicates that the evaluated cluster produced a good cluster. With a better understanding of the nutritional patterns of toddlers in Mekar Wangi Village, Posyandu officers can developing a more efficient program to improve the nutritional quality of children in Mekar Wangi Village. Posyandu officers can assist in decision making to develop more targeted recommendations and interventions to improve the nutritional status of toddlers in Mekar Wangi village.

Keywords: K-Means clustering, baby nutrition, posyandu

PENDAHULUAN

Pertumbuhan dan kesehatan balita sangat penting untuk perkembangan masyarakat dan negara. semua bertanggung jawab untuk memastikan bahwa balita adalah generasi masa depan. Balita di Indonesia sering mengalami masalah kekurangan gizi.

Keterlibatan yang kuat dari Puskesmas dan Posyandu sangat penting untuk mengurangi jumlah balita yang menderita gizi buruk. Untuk memastikan perkembangan balita berjalan dengan baik dan membuat mereka terbiasa dengan gaya hidup sehat di masa depan, sangat penting untuk memberi mereka

asupan gizi yang seimbang dan sesuai pada tahap perkembangan yang tepat. Penting untuk diingat bahwa makanan bergizi tidak selalu harus mahal.(Irfiani & Rani, 2018)

Pertumbuhan balita dan kesehatannya sangat penting untuk kemajuan masyarakat dan negara. Semua orang bertanggung jawab untuk memastikan generasi masa depan terdiri dari balita. Di Indonesia, balita sering mengalami kekurangan gizi. Untuk memastikan perkembangan balita berjalan dengan baik dan membuat mereka terbiasa dengan gaya hidup sehat di masa depan, sangat penting untuk memberi mereka asupan gizi yang seimbang dan sesuai pada tahap perkembangan yang tepat. Penting untuk diingat bahwa harga makanan bergizi tidak selalu menentukan kualitasnya. (Zami et al., 2022).

Penelitian terdahulu terkait clustering gizi balita dilakukan oleh (Gustrianda & Mulyana, 2022) yang berjudul Penerapan Data Mining Dalam Pemilihan Produk Unggulan dengan Metode Algoritma *K-Means* Dan *K-Medoids*. Hasil penelitian sebelumnya menunjukkan bahwa nilai Davies Bouldin yang diperoleh dari penerapan algoritma *K-Means* adalah -0,430, sedangkan nilai Davies Bouldin yang dihasilkan dari penggunaan algoritma *K-Medoids* adalah -1,392. Hasil ini mengindikasikan bahwa metode *K-Medoids* memiliki nilai Davies Bouldin yang lebih rendah dibandingkan dengan metode *K-Means*. Oleh karena itu, penggunaan metode *K-Means* lebih disarankan dalam konteks pemilihan produk unggulan.

Sementara penelitian yang dilakukan oleh (Adiputra, 2022) yang berjudul Clustering Penyakit DBD Pada Rumah Sakit Dharma Kerti Menggunakan Algoritma *K-Means*. Hasil penelitian ini menunjukkan bahwa kelompok pasien yang menerima diagnosis Demam Berdarah Dengue (DBD) paling sedikit dalam jumlahnya diidentifikasi dalam kelompok usia antara 25 hingga 45 tahun di Rumah Sakit Dharma Kerti.

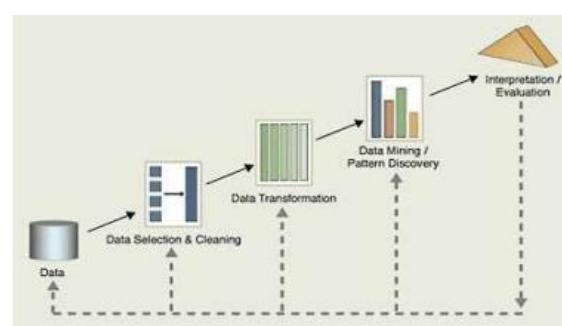
Penelitian selanjutnya dilakukan oleh (Nabila et al., 2021) yang berjudul Analisis Data Mining Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma *K-*

Means. Hasil validasi *Indeks Davies-Bouldin* (DBI) dengan metode perhitungan manual dan menggunakan perangkat RapidMiner menunjukkan perbedaan. Dalam konteks ini, perhitungan manual menghasilkan hasil yang lebih optimal dibandingkan dengan menggunakan alat bantu RapidMiner. Namun, penting untuk dicatat bahwa kedua perhitungan tersebut menghasilkan nilai DBI yang mendekati nol, menunjukkan bahwa klaster yang dievaluasi menghasilkan struktur klaster yang baik.

Pada penelitian ini hanya akan fokus untuk mengolah data pada desa mekar wangi yang sebelumnya belum pernah ada penelitian di desa mekar wangi, pada pengolahan data menggunakan tools rapid miner.

METODE

Pada penelitian ini menggunakan proses tahapan Knowledge Discovery in Databases (KDD) yang terdiri dari beberapa tahapan penting.



Gambar 1. Knowledge Discovery in Databases (KDD)

Tahapan KDD ini dirancang untuk menggali pengetahuan yang berharga dari data.

- Pemilihan Data (Data Selection: Tahapan ini adalah pengumpulan data. Data merupakan bahan mentah yang akan dianalisis. Proses seleksi dilakukan pada data posyandu di Desa Mekar Wangi pada tahun 2023 bulan oktober sebanyak 403 data.
- Pra-pemrosesan Data (Preprocessing): Setelah data terkumpul, tahapan selanjutnya adalah pembersihan data.

Pembersihan data melibatkan identifikasi dan perbaikan masalah dalam data, seperti data yang hilang, data yang duplikat, atau data yang tidak akurat. Tujuan dari tahapan ini adalah memastikan bahwa data yang digunakan dalam analisis adalah berkualitas tinggi dan dapat diandalkan.

- c) Transfomasi Data (Transformation): Tahapan berikutnya adalah transformasi data. Data seringkali perlu diubah atau disesuaikan agar sesuai dengan format dan struktur yang diperlukan untuk analisis. Hal ini mencakup pemilihan variabel yang relevan, normalisasi data, dan pengkodean data jika diperlukan.
- d) Penggalian Data (Data Mining): Mengidentifikasi pola atau informasi menarik dalam suatu kumpulan data dengan menggunakan metode dan algoritma. Pendekatan, metode, dan algoritma dalam data mining memiliki beragam variasi. Dalam konteks penelitian ini, penerapan data mining menggunakan algoritma K-means clustering karena untuk menentukan nilai k setelah melihat nilai centroid.
- e) Evaluasi merupakan penafsiran atau kesimpulan dari hasil data mining. Kesimpulan akhir dibuat dengan melihat hasil analisis data mining K-means clustering di evaluasi dengan cara mencari nilai DBI nya.

HASIL DAN PEMBAHASAN

Hasil penelitian ini menggunakan clustering K-Means dari data posyandu Desa Mekar Wangi, yang dilakukan dengan metode Knowledge Discovery Database (KDD), yang dilakukan dengan tools Rapidminer versi 10.1. Penelitian ini akan menyajikan temuan dan analisis dari penelitian. Penulis menemukan diskusi yang menjelaskan proses yang dilakukan, mulai dari pengambilan data, pengolahan data mentah menjadi data yang telah diproses oleh algoritma, hingga tahap evaluasi. Diharapkan bahwa temuan yang diperoleh akan memberikan dasar untuk penelitian yang akan datang.

1. Data Selection

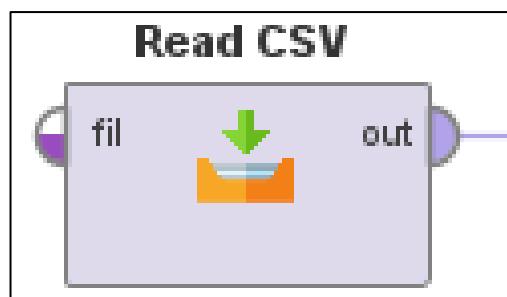
Pada tahap awal Knowledge Discovery in Databases (KDD), penulis memulai dengan seleksi data. Data yang dipilih berasal dari

Posyandu Mekar Wangi. Proses seleksi ini dilakukan di Microsoft Excel untuk memilih dataset yang akan dianalisis dengan menggunakan Tools RapidMiner. Dataset yang digunakan terdiri dari 403 data dan memiliki sembilan atribut, seperti Nama anak, Umur, Umur/Bulan, Tinggi badan, Berat Badan, Lila, dan lila sebagaimana terlihat pada Gambar 2. Tahap seleksi data ini menjadi dasar penelitian selanjutnya, dataset yang terpilih akan diolah untuk pengelompokan data.

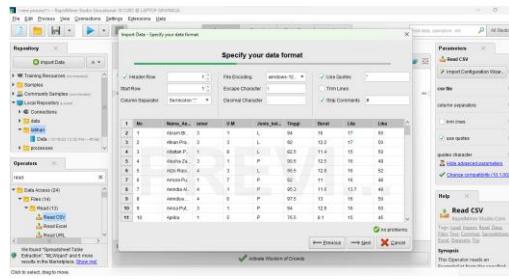
No	Nama_Anak	umur	umur/bulan	Jenis_kelamin	Tinggi	Berat	Lila	Lika
1	Abram Bin Indra	3	1	L	94	16	17	50
2	Afnan Pratama	3	3	L	92	12.	17	50
3	Alfattah Purnama	1	8	L	82.5	11.	15	50
4	Alesha Zahra	3	1	P	90.5	12.	16	49
5	Alzio Rayyan Syahputra	4	3	L	96.5	12.	16	52
6	Amora Putri Cantika	1	7	P	82	11	16	48
7	Anindia Alisha Silvia	4	1	P	95.5	11.	13	48
8	Anindiya Pauziah	4	6	P	97.5	13	16	50
9	Anisa Putri Junaira	3	1	P	94	12.	16	50
10	Aprilia	1	5	P	76.5	8.1	15	45
11	Aqilan Zikri Hamizam	1	5	L	81.6	9.8	16	48
12	Azami Lazuardi H	1	11	L	83	11	16	48
13	Azarine Putri Afifah	4	4	P	100	15.	17	51
14	Azka Malik Ibrahim	1	10	L	81	9.6	15	48
15	Brian	2	10	L	86.5	11.	15	47
16	Ceki Triana kurnia Putra	4	3	L	99.1	15	18	52
17	Daffa Elvano Adhitama Fauzi	3	9	L	100	15.	17	50
...
40	Latif Alfariz	1	1	L	58.1	5.2	11	38
3								

Gambar 2. Dataset Terpilih

Setelah data diselection pada Gambar 2 melalui tahap selection data bahwa data yang diambil adalah No, Nama Anak, umur, dan lainnya. Setelah Data Di Selection tahapan pertama import file dataset Posyandu Desa Mekar Wangi ke Operator Read CSV untuk membaca data xlsx. Pastikan Operator read excel tidak tanda seru dan warna kuning artinya file telah di import dan siap diolah.



Gambar 3. Read CSV



Gambar 4. Hasil Import Data xlsx ke Read CSV

Gambar 4 merupakan hasil dari memasukan data ke dalam operator Read CSV. Selanjutnya klik next dan hasil dari next dapat dilihat pada Gambar 5.

Format your columns.							
Data format		Enter value...					
Replace errors with missing values							
No	integer	Nama_Ak.	umur	umur/bulan	Jenis_kela...	Tinggi	Berat
1	Afram Bin Indra	3	1	L	94	16	
2	Ahan Pratama	3	3	L	92	12	
3	Affattah Purnama	1	8	L	83	11	
4	Alesha Zahra	3	1	P	91	13	
5	Aliso Rayyan Sya...	4	3	L	97	13	
6	Amora Putri Cant...	1	7	P	82	11	
7	Anindya Alisha Sil...	4	1	P	95	30	
8	Anindya Pauziah	4	6	P	97	50	
9	Anisa Putri Junitara	3	1	P	94	13	
10	Aprilia	1	5	P	77	8	
11	Aqilan Zikri Hamiz...	1	5	L	82	10	

Gambar 5. Hasil dari next import data

Pada Gambar 5. menjelaskan bahwa penggantian tipe data dari real menjadi integer adalah untuk menyimpan bilangan bulat atau tanpa ada koma. Maka hasil dari Gambar 5 bisa dilihat pada gambar 6.

Format your columns.							
Data format		Enter value...					
Replace errors with missing values							
No	Nama_Ak.	umur	umur/bulan	Jenis_kela...	Tinggi	Berat	
1	Afram Bin Indra	3	1	L	94	16	
2	Ahan Pratama	3	3	L	92	12	
3	Affattah Purnama	1	8	L	83	11	
4	Alesha Zahra	3	1	P	91	13	
5	Aliso Rayyan Sya...	4	3	L	97	13	
6	Amora Putri Cant...	1	7	P	82	11	
7	Anindya Alisha Sil...	4	1	P	95	12	
8	Anindya Pauziah	4	6	P	98	13	
9	Anisa Putri Junitara	3	1	P	94	13	
10	Aprilia	1	5	P	77	8	
11	Aqilan Zikri Hamiz...	1	5	L	82	10	

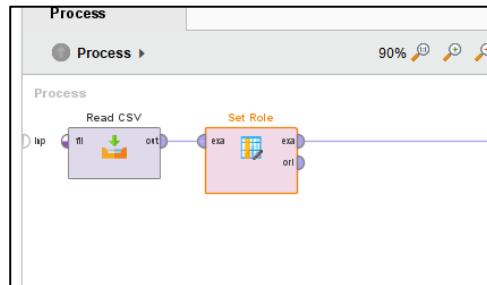
Gambar 6. Hasil dari perubahan tipe data

2. PreProcessing data

Setelah melakukan tahap selection maka tahap selanjutnya adalah melakukan data preprocessing. Pada tahap ini tujuan dari data preprocessing yaitu menambahkan id pada dataset menghapus atribut yang tidak di perlukan, data yang null atau missing, serta mengganti beberapa data yang hilang. Pada tahap preprocessing data penulis memiliki 3 jenis operator yaitu select attributes, operator replace missing value dan set role. Adapun langkah-langkah untuk melakukan preprocessing data dengan 2 operator yaitu sebagai berikut:

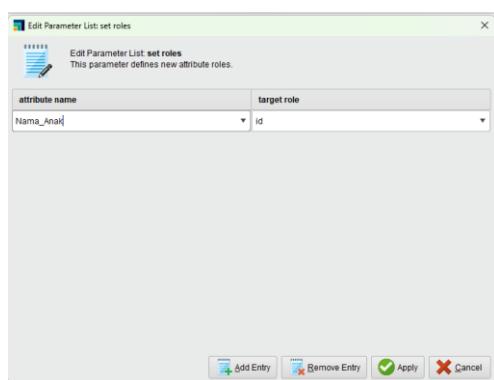
a. Operator Set Role

Langkah pertama dalam preprocessing adalah menggunakan operator set role. Operator set role merupakan operator yang berfungsi untuk menamai id, tujuan dari penamaan id adalah wajib karena dataset wajib memiliki identitas, karena adanya identitas dataset yang digunakan akan mudah untuk dideskripsikan. Penggunaan operator set role dapat dilihat pada Gambar 7.



Gambar 7. Tampilan Operator Set role

Parameter yang dipilih dari set role bisa dilihat pada Gambar 8.

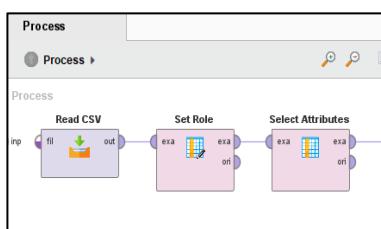


Gambar 8. Parameter set role

Pada Gambar 8 ada 2 parameter yaitu attribut name dan target role, di sini peneliti memilih attribut name memilih nama anak dan target role adalah id, maksud dari gambar 4.6 adalah attribut dengan nama_anak adalah id dari dataset.

b. Operator Select Attributes

Langkah kedua dalam preprocessing data adalah dengan menggunakan operator select attributes. Operator select attribute merupakan operator yang berfungsi untuk memilih attribut yang akan di pakai dan menghapus atribut yang tidak terpakai. Tampilan operator select attributes dapat dilihat pada Gambar 9.



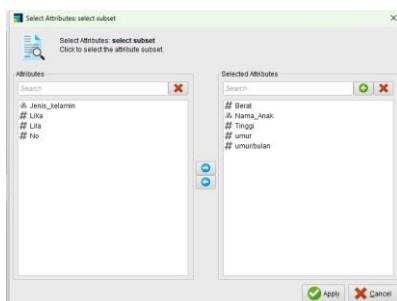
Gambar 9. Operator Select Attributes

Pada operator select attribut terdapat beberapa parameter yang harus di setting atau dan disesuaikan dengan kebutuhan. Pengaturan parameter tersebut dapat di lihat pada Gambar 10.

Parameter	Pilihan	Keterangan
Type attributes	Exclude attributes	Untuk menghilangkan attribut yang tidak dipakai
Attribut filter type	A subset	Pemilihan beberapa attribut
Select subset	No, Lila, Lika	Attribut yang tidak dipilih

Gambar 10. Pengaturan Parameter

Penggunaan select attributes dapat menghilangkan tiga attribut yaitu No, Lila, dan lika, setelah proses pemilihan maka attribut yang dipilih adalah Nama anak, umur, umur/bulan, tinggi dan berat. Hasil dari select attributes dapat dilihat pada Gambar 11.



Gambar 11. Operator Select Attributes

Setelah memilih attribut mana saja yang akan digunakan pada Gambar 11 maka hasil dari operator select attributes dapat dilihat pada Gambar 12.

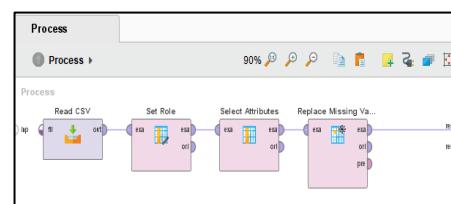
Row No.	Nama_Anak	umur	umurbulan	Tinggi	Berat
130	Abil	0	8	71	8
131	Adina	3	0	92	12
132	Adinata	1	3	?	9
133	Airaa Sania	2	5	89	11
134	Aishwa	3	3	?	9
135	Aisyah Tazkiyah	0	7	?	9
136	Aleatha Zahra	4	0	99	18
137	Altaro Haili Z...	1	11	88	14
138	Ali Syahda	0	7	?	9
139	Alia Ahza	3	8	98	18
140	Alifwan	3	3	?	9
141	Aluma	4	5	?	9
142	Andita	1	5	78	8
143	Andra	4	3	?	9
144	Anthar Pradipt...	3	10	96	13

Gambar 12. Hasil Operator Select Attributes

Pada Gambar 12 pada attribut tinggi dan berat terdapat nilai yang kosong, maka dari itu tahap selanjutnya menggunakan operator replace missing value.

c. Replace Missing Value

Operator replace missing value digunakan untuk mengisi data yang kosong. Selama proses, hanya beberapa sampel yang hilang dan telah diatasi dengan menggantikan nilai pada atribut x. Teknik ini umum digunakan untuk menangani nilai yang hilang. Tampilan operator select attributes dapat dilihat pada Gambar 13.



Gambar 13. Operator Replace Missing Value

Hasil dari penggunaan operator missing value dapat dilihat pada Gambar 14.

Row No.	Nama_Anak	umur	umurbulan	Tinggi	Berat
129	Abinna	2	10	86	13
130	Abil	0	8	71	8
131	Adina	3	0	92	12
132	Adinata	1	3	86	13
133	Airaa Sania	2	5	89	11
134	Aishwa	3	3	86	13
135	Aisyah Tazkiyah	0	7	86	13
136	Aleatha Zahra	4	0	99	18
137	Altaro Haili Z...	1	11	88	14
138	Ali Syahda	0	7	86	13
139	Alia Ahza	3	8	98	18
140	Alifwan	3	3	86	13
141	Aluma	4	5	86	13
142	Andita	1	5	78	8
143	Andra	4	3	86	13

Gambar 14. Hasil Operator Replace Missing Value

Pada Gambar 14 data awal sebelum dilakukan proses replace missing value terdiri dari 403 data, setelah diterapkannya operator ini data tidak ada yang kosong tetapi masih sama seperti data sebelumnya 403 data. Berdasarkan jumlah tersebut, dapat disimpulkan bahwa data yang memiliki nilai kosong akan terisi dan menghasilkan sejumlah data awal sebanyak 403 data.

3. Transformasi Data

Setelah melakukan tahapan data *preprocessing* maka tahapan selanjutnya adalah melakukan data *transformation*, pada tahap ini penulis melakukan perubahan data pada umur dan umur/bulan. Tampilan dapat di lihat pada Gambar 15.

No	Nama Anak	umur	umur/bulan	Tinggi	Berat
1	Abram Bin Indra	3	1	91	13
2	Afnan Pratama	3	3	97	13
3	Alfattah Purnama	1	8	82	11
4	Alesha Zahra	3	1	95	12
5	Alzio Rayyan Syahputra	4	3	98	13
6	Amora Putri Cantika	1	7	94	13
7	Anindya Alisha Silvia	4	1	77	8
8	Anindiya Pauziah	4	6	82	10
9	Anisa Putri Junaira	3	1	83	11
10	Aprilia	1	5	100	16
11	Aqlan Zikri Hanizam	1	5	81	10
12	Azam Lazuardi H	1	11	87	12
13	Azarine Putri Afifah	4	4	99	15
14	Azka Malik Ibrahim	1	10	100	16
15	Brian	2	10	100	14
16	Ceki Triana Kurnia Putra	4	3	76	9
17	Daffa Elvano Adhitama Fauzi	3	9	95	14
...
403	Latif Alfariz	1	1	58	5

Gambar 15. Tampilan yang akan di transformasi

Pada Gambar 16 merupakan tampilan awal dataset dari excel, maka penulisan merubah nama bayi menjadi bayi 1 sampai bayi 403, dan umur dan umur/bulan. Tampilan dapat dilihat pada Gambar 16.

No	Nama Anak	Umur	Tinggi	Berat
1	Balita ke 1	37	94	16
2	Balita ke 2	39	92	12
3	Balita ke 3	20	83	11
4	Balita ke 4	37	91	13
5	Balita ke 5	51	97	13
6	Balita ke 6	19	82	11
7	Balita ke 7	49	95	12
8	Balita ke 8	54	98	13
9	Balita ke 9	37	94	13
10	Balita ke 10	17	77	8
11	Balita ke 11	17	82	10
12	Balita ke 12	23	83	11
13	Balita ke 13	52	100	16
14	Balita ke 14	22	81	10
15	Balita ke 15	34	86.5	11.7
16	Balita ke 16	51	99.1	15
17	Balita ke 17	45	100	15.8
18	Balita ke 18	57	100.1	14.1
...
403	Balita ke 403	13	58.1	5.2

Gambar 16. Hasil Transformasi excel ke excel

Gambar 17 menunjukan hasil dari excel, yang awal terdapat attribut umur dan umur/bulan menjadi 1 attribut, yaitu umur dalam bentuk bulan dengan nama attribut umur dan nama bayi menjadi balita 1 sampai balita 403.

No	Nama_Anak	Umur	Tinggi	Berat
1	Balita ke 1	37	94	16
2	Balita ke 2	39	92	12
3	Balita ke 3	20	82.5	11.4
4	Balita ke 4	37	90.5	12.5
5	Balita ke 5	51	96.5	12.8
6	Balita ke 6	19	82	11
7	Balita ke 7	49	95.3	11.6
8	Balita ke 8	54	97.5	13
9	Balita ke 9	37	94	12.9
10	Balita ke 10	17	76.5	8.1
11	Balita ke 11	17	81.6	9.8
12	Balita ke 12	23	83	11
13	Balita ke 13	52	100	15.7
14	Balita ke 14	22	81	9.6
15	Balita ke 15	34	86.5	11.7
16	Balita ke 16	51	99.1	15
17	Balita ke 17	45	100	15.8
18	Balita ke 18	57	100.1	14.1
...
403	Balita ke 403	13	58.1	5.2

Gambar 17. Hasil Perubahan Transformasi

Dari Gambar 17 merupakan tampilan dalam excel, maka jika dataset yang dimasukan kedalam rapid miner akan tampil seperti pada Gambar 18.

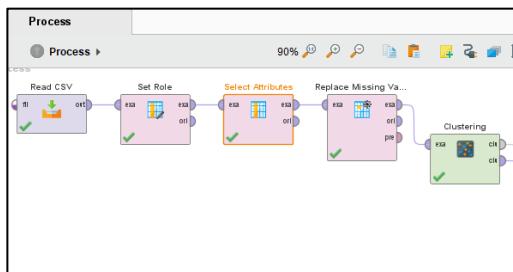
Row No.	Nama_Anak	Tinggi	Berat
1	Balita ke 1	94	16
2	Balita ke 2	92	12
3	Balita ke 3	83	11
4	Balita ke 4	91	13
5	Balita ke 5	97	13
6	Balita ke 6	82	11
7	Balita ke 7	95	12
8	Balita ke 8	98	13
9	Balita ke 9	94	13
10	Balita ke 10	77	8
11	Balita ke 11	82	10
12	Balita ke 12	83	11
13	Balita ke 13	100	16
14	Balita ke 14	81	10
15	Balita ke 15	87	12

Gambar 18. Hasil Tampilan Transformasi data

4. Data Mining

Proses penelitian ini menggunakan algoritma k-means. Dalam penelitian ini

dibantu rapidminer untuk algoritma k-means clustering. Model data mining yang digunakan dalam penelitian ini adalah model pengelompokan atau clustering. Data mining diimplementasikan dengan menggunakan RapidMiner. Proses pemodelan RapidMiner dapat dilihat pada Gambar 19. Operator yang digunakan adalah clustering. Operator clustering digunakan untuk mengelompokkan data.



Gambar 19. Proses Pemodelan

Pada proses clustering, terdapat beberapa parameter yang harus di setting seperti pada Gambar 20.

Parameter	Pilihan	Isi
K min	2-10 cluster	2
K max	10	20
Measure type	Numerical Measures	numericalMeasure
Numerical measure	Euclidean Distance	EuclideanDistance
Max optimization	100	100

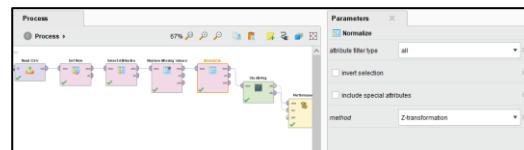
Gambar 20. Parameter Clustering

Langkah yang kedua yaitu melakukan evaluasi atau menguji hasil clustering dengan menerapkan index validitas DBI pada parameter Cluster Distance Performance. Dalam pengujian DBI, cluster yang memiliki nilai DBI terkecil atau mendekati 0 dianggap sebagai cluster yang terbaik. Pada tahapan ini, peneliti melakukan eksperimen pengujian DBI dengan membandingkan nilai dari 3 model proses pengujian. Pengujian yang pertama yaitu menerapkan operator Normalize dengan metode Z-Transformation. Pengujian yang kedua yaitu menerapkan operator Normalize dengan metode Range Transformation. Pengujian yang ketiga yaitu tidak menerapkan operator Normalize.

(a) Menerapkan Operator Normalize (Z- Transformation)

Pada Rapidminer operator Normalize digunakan untuk menskalakan nilai sesuai dengan rentang tertentu, tujuannya agar nilai pada setiap atribut memiliki

rentang nilai yang sama. Normalize dengan metode Z-transformation merupakan normalisasi statistik, dimana penentuan rentang nilai ditentukan secara otomatis. Model proses yang menerapkan operator Normalize dengan metode Z-Transformation seperti Gambar 21.



Gambar 21. Normalize Z-Transformation

Pengujian DBI pada proses yang menerapkan operator Normalize dengan metode Z Transformation, terdapat beberapa parameter yang harus diatur. Pengaturan pada parameter ini seperti pada Gambar 22.

Operator	Parameter	Pilihan	Keterangan
Normalize (Z-transformation)	attribute filter type	Subet	Tipe yang digunakan
	Attributes	Time,Source,Destinatio n,_Protocol,Length	Atribut yang dinormalisasi
	Method	Z-Transformation	Metode normalisasi yang dipilih
Cluster distance Performance	Main Criterion		kriteria yang dipilih untuk evaluasi clustering
	Normalize	Ceklis	Hasil dinormalisasi
	Maximize	Ceklis	Hasil dimaksimalkan

Gambar 22. Parameter Normalize (Z- Transformation)

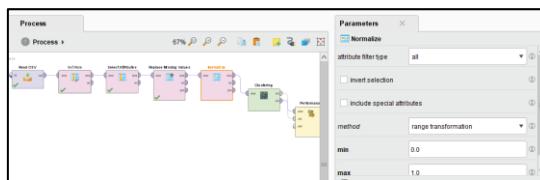
Hasil pengujian DBI pada model proses yang menggunakan operator *Normalize* dengan metode *Z-Transformation* dengan percobaan K = 2 hingga K = 10 seperti pada Gambar 23.

NORMALIZE (Z-TRANSFORMATION)		
Cluster (K)	DBI	Jumlah Anggota Cluster
2	0.164	Cluster 0: 398 items, Cluster 1: 5 items Total number of items: 403
3	0.251	Cluster 0: 274 items, Cluster 1: 124 items Cluster 2: 5 items, Total number of items: 403
4	0.249	Cluster 0: 128 items, Cluster 1: 95 items, Cluster 2: 175 items, Cluster 3: 5 items Total number of items: 403
5	0.234	Cluster 0: 94 items, Cluster 1: 128 items Cluster 2: 175 items, Cluster 3: 4 items Cluster 4: 2 items Total number of items: 403
6	0.246	Cluster 0: 165 items Cluster 1: 27 items Cluster 2: 4 items, Cluster 3: 80 items, Cluster 4: 125 items, Cluster 5: 2 items Total number of items: 403
7	0.276	Cluster 0: 137 items, Cluster 1: 2 items Cluster 2: 89 items, Cluster 3: 69 items Cluster 4: 27 items, Cluster 5: 4 items Cluster 6: 75 items, Total number of items: 403
8	0.285	Cluster 0: 98 items, Cluster 1: 47 items, Cluster 2: 42 items, Cluster 3: 3 items, Cluster 4: 17 items, Cluster 5: 134 items Cluster 6: 4 items, Cluster 7: 59 items Cluster 8: 5 items, Total number of items: 403
9	0.272	Cluster 0: 45 items, Cluster 1: 4 items, Cluster 2: 53 items, Cluster 3: 58 items Cluster 4: 65 items, Cluster 5: 17 items Cluster 6: 124 items, Cluster 7: 2 items Cluster 8: 5 items, Total number of items: 403
10	0.257	Cluster 0: 119 items, Cluster 1: 2 items Cluster 2: 10 items, Cluster 3: 4 items Cluster 4: 54 items, Cluster 5: 40 items Cluster 6: 64 items, Cluster 7: 62 items Cluster 8: 2 items, Cluster 9: 79 items Total number of items: 403

Gambar 23. Hasil Pengujian DBI

(b) Menerapkan Operator Normalize (Range Transformation)

Pada Rapidminer, Normalize dengan metode range transformation merupakan normalisasi yang rentang nilainya dapat ditentukan oleh peneliti. Dalam pengujian ini, rentang nilai yang diterapkan yaitu dari skala minimal 0 dan maksimal 10. Model proses yang menerapkan operator Normalize dengan metode Range Transformation seperti yang digambarkan pada Gambar 24.



Gambar 24. Normalize Range Transformation

Pengujian DBI pada proses yang menerapkan operator Normalize dengan metode Range Transformation, terdapat beberapa parameter yang harus diatur. Pengaturan pada parameter ini seperti pada Gambar 25.

Operator	Parameter	Pilihan	Keterangan
Normalize (Z-transformation)	attribute filter type	Subset	Tipe yang digunakan
	Attributes	Time, Source, Destination, Protocol, Length	Attribut yang dinormalisasi
	Method	Range transformation	Metode normalisasi yang dipilih
	Min	0.0	Nilai minimum rentang skala
	Max	10.0	Nilai maksimum rentang skala
Cluster distance Performance	Main Criterion		kriteria yang dipilih untuk evaluasi clustering
	Normalize	Ceklis	Hasil dinormalisasi
	Maximize	Ceklis	Hasil dimaksimalkan

Gambar 25. Parameter Normalize

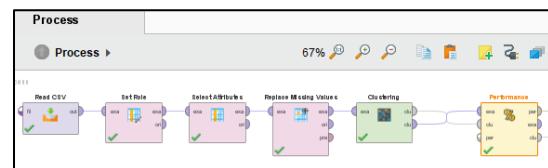
Hasil pengujian DBI pada model proses yang menggunakan operator Normalize dengan metode Range Transformation dengan percobaan $K = 2$ hingga $K = 10$ dapat dilihat pada Gambar 26.

NORMALIZE (RANGE TRANSFORMATION)		
Cluster (K)	DBI	Jumlah Anggota Cluster
2	0.319	Cluster 0: 384 items, Cluster 1: 119 items, Total number of items: 403
3	0.298	Cluster 0: 183 items, Cluster 1: 94 items, Cluster 2: 126 items, Total number of items: 403
4	0.234	Cluster 0: 93 items, Cluster 1: 5 items, Cluster 2: 126 items, Cluster 3: 179 items, Total number of items: 403
5	0.253	Cluster 0: 160 items, Cluster 1: 7 items, Cluster 2: 126 items, Cluster 3: 84 items, Cluster 4: 84 items, Total number of items: 403
6	0.279	Cluster 0: 91 items, Cluster 1: 136 items, Cluster 2: 28 items, Cluster 3: 67 items, Cluster 4: 77 items, Cluster 5: 4 items, Total number of items: 403
7	0.250	Cluster 0: 56 items, Cluster 1: 71 items, Cluster 2: 145 items, Cluster 3: 77 items, Cluster 4: 101 items, Cluster 5: 2 items, Cluster 6: 4 items, Total number of items: 403
8	0.251	Cluster 0: 124 items, Cluster 1: 2 items, Cluster 2: 55 items, Cluster 3: 10 items, Cluster 4: 16 items, Cluster 5: 67 items, Cluster 6: 43 items, Cluster 7: 91 items Total number of items: 403
9	0.258	Cluster 0: 109 items, Cluster 1: 30 items, Cluster 2: 4 items, Cluster 3: 59 items, Cluster 4: 2 items, Cluster 5: 79 items, Cluster 6: 16 items, Cluster 7: 45 items, Cluster 8: 51 items, Cluster 9: 1 item, Total number of items: 403
10	0.263	Cluster 0: 109 items, Cluster 1: 2 items, Cluster 2: 16 items, Cluster 3: 16 items, Cluster 4: 4 items, Cluster 5: 16 items, Cluster 6: 28 items, Cluster 7: 79 items, Cluster 8: 39 items, Cluster 9: 38 items Total number of items: 403

Gambar 26. Hasil Pengujian DBI

(c) Tidak Menerapkan Operator Normalize

Pengujian hasil clustering dengan model proses yang tidak menerapkan Normalize tekniknya langsung menghubungkan operator clustering pada operator Cluster Distance Performance. Pada parameter operator Performance yang diatur hanya pada main centroid yaitu memilih (Davies Bouldin) dan beri tanda ceklis pada opsi normalize dan maximize. Model proses yang tidak menerapkan operator Normalize dapat dilihat pada Gambar 27.



Gambar 27. Tanpa Pengujian Operator Normalize

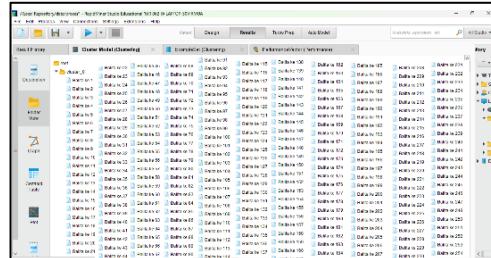
Hasil pengujian DBI pada model proses yang tidak menggunakan operator Normalize dengan percobaan $K = 2$ hingga $K = 10$ dapat dilihat pada Gambar 28.

NORMALIZE (RANGE TRANSFORMATION)		
Cluster (K)	DBI	Jumlah Anggota Cluster
2	0.338	Cluster 0: 125 items, Cluster 1: 278 items, Total number of items: 403
3	0.251	Cluster 0: 124 items, Cluster 1: 274 items, Cluster 2: 5 items, Total number of items: 403
4	0.241	Cluster 0: 128 items, Cluster 1: 5 items, Cluster 2: 83 items, Cluster 3: 177 items, Total number of items: 403
5	0.272	Cluster 0: 28 items, Cluster 1: 165 items, Cluster 2: 125 items, Cluster 3: 5 items, Cluster 4: 80 items Total number of items: 403
6	0.249	Cluster 0: 125 items, Cluster 1: 4 items, Cluster 2: 2 items, Cluster 3: 35 items, Cluster 4: 75 items, Cluster 5: 162 items, Total number of items: 403
7	0.267	Cluster 0: 136 items, Cluster 1: 68 items, Cluster 2: 76 items, Cluster 3: 4 items, Cluster 4: 2 items, Cluster 5: 90 items, Cluster 6: 27 items, Total number of items: 403
8	0.275	Cluster 0: 129 items, Cluster 1: 19 items, Cluster 2: 55 items, Cluster 3: 2 items, Cluster 4: 100 items, Cluster 5: 4 items, Cluster 6: 38 items, Cluster 7: 56 items, Total number of items: 403
9	0.290	Cluster 0: 126 items, Cluster 1: 20 items, Cluster 2: 4 items, Cluster 3: 43 items, Cluster 4: 53 items, Cluster 5: 2 items, Cluster 6: 77 items, Cluster 7: 61 items, Cluster 8: 17 items, Total number of items: 403
10	0.266	Cluster 0: 53 items, Cluster 1: 2 items, Cluster 2: 64 items, Cluster 3: 7 items, Cluster 4: 2 items, Cluster 5: 126 items, Cluster 6: 17 items, Cluster 7: 43 items, Cluster 8: 87 items, Cluster 9: 2 items, Total number of items: 403

Cluster Model

Cluster 0: 398 items
 Cluster 1: 5 items
 Total number of items: 403

Gambar 30. Cluster Model



Gambar 28. Hasil Pengujian DBI

Berdasarkan ketiga pengujian yang telah dilakukan, maka hasil rekapitulasi perbandingan nilai DBI dari ketiga proses ini dapat disimpulkan seperti pada Gambar 29.

Cluster (K)	Nilai DBI		
	Normalize (Z-Transformation)	Normalize (Range Transformation)	Tidak menerapkan Normalize
2	0.164	0.319	0.338
3	0.251	0.298	0.251
4	0.259	0.234	0.241
5	0.234	0.253	0.272
6	0.246	0.279	0.249
7	0.276	0.250	0.267
8	0.285	0.251	0.275
9	0.272	0.258	0.290
10	0.257	0.263	0.266

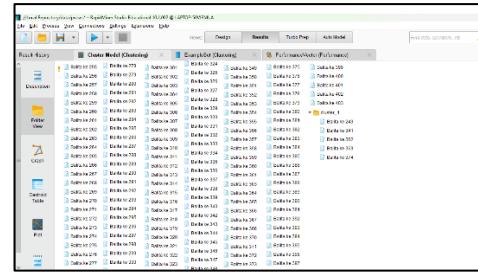
Gambar 29. Komparasi Nilai DBI

5. Evaluation

Berdasarkan Gambar 29 menunjukkan bahwa setiap cluster memiliki nilai DBI yang berbeda. Pengujian DBI dari jumlah K = 2 sampai K = 10 berdasarkan model proses yang digunakan memperoleh nilai DBI yang bervariasi. Sesuai dengan fungsinya yaitu DBI dapat mengukur validitas suatu cluster, maka hasil cluster lebih optimal jika nilai DBI semakin kecil atau mendekati 0. Dalam penelitian ini nilai DBI terkecil pada pengujian Performance Davies-Bouldin adalah 0,164. Nilai ini diperoleh dari model proses yang tidak menerapkan operator Normalize dengan K = 2.

Pada K = 2, maka diperoleh anggota cluster 0 adalah 398 dan anggota cluster 1 berjumlah 5 anggota.

Gambar 31. Hasil Folder View Cluster 0



Gambar 32. Hasil Folder view Cluster 0 dan Cluster1

Dalam menentukan cluster yang terbaik antara cluster 0 dan cluster 1 menggunakan Performance Centroid Distance dari masing-masing cluster serta menerapkan evaluasi DBI.

PerformanceVector

PerformanceVector:
 Avg. within centroid distance: 0.599
 Avg. within centroid distance_cluster_0: 0.577
 Avg. within centroid distance_cluster_1: 2.397
 Davies Bouldin: 0.164

Gambar 33. Performance Vector

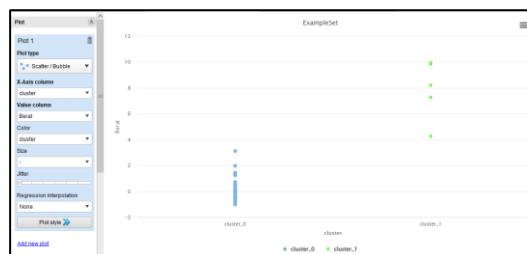
Pada Gambar 33 dapat diketahui informasi mengenai Performance dari hasil clustering. Informasi yang diperoleh yaitu nilai DBI (Davies-Bouldin Index) dan nilai rata-rata jarak dari setiap cluster ke titik pusat (Avg. within centroid distance). Nilai dari Avg. within centroid distance dapat digunakan untuk menentukan cluster yang terbaik. Pada Gambar 34 menjelaskan proses dalam

menentukan cluster yang memiliki nilai rata-rata jarak terdekat.

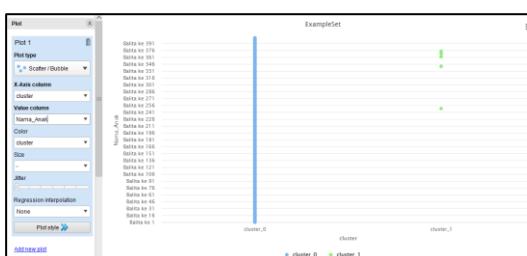
Cluster	Avg. within centroid distance	Avg. within centroid distance cluster	Selisih centroid (Avg. within centroid distance – Avg. within centroid distance cluster)
Cluster 0	0.599	0.577	0.022
Cluster 1	0.599	2.397	-1.798

Gambar 34. Perhitungan selisih jarak rata-rata centroid

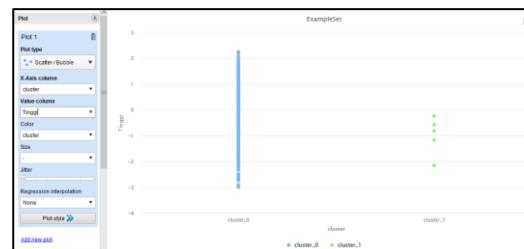
Sesuai dengan aturan DBI yakni cluster yang memiliki nilai terkecil atau mendekati 0 merupakan cluster yang terbaik. Tabel 4.110 memperlihatkan bahwa cluster terbaik adalah cluster 0 karena memperoleh nilai rata-rata selisih jarak yang lebih kecil dari cluster 1 yakni 0.022. Pada output example set view visualization plot dengan tipe Scatter/Bubble dari masing-masing atribut menunjukkan bahwa cluster 0 merupakan cluster yang memiliki nilai DBI terbaik atau mendekati 0. Plot Scatter/Bubble dari setiap atribut dapat dilihat pada Gambar 35.



Gambar 35. Plot Scatter Attribut Berat



Gambar 36. Plot Scater Attribut Nama_Anak



Gambar 37. Plot Scatter Attribut Tinggi

Berdasarkan Gambar (35, 36, 37) menunjukkan bahwa pada *cluster* 0 dan *cluster* 1 memiliki grafik *Plot Scatter/Bubble* yang berbeda. Pada *cluster* 0 jarak setiap anggotanya berdekatan dan mendekati nilai 0, sedangkan pada *cluster* 1 jarak antara anggota tidak berdekatan. Sesuai dengan nilai DBI dan tampilan *visualization* dari *plot scatter/bubble* dapat dibuktikan bahwa *cluster* 0 merupakan *cluster* yang terbaik.

Rata-rata nilai dari masing-masing atribut seperti pada Gambar 38.

Attribute	cluster_0	cluster_1
Tinggi	0.012	-4.988
Berat	-0.099	7.902

Gambar 38. Centroid Tabel

No	Standar Anthropometri (Tinggi Badan Ideal Berdasarkan Usia)
1	Bayi Usia 0 - 3 Bulan: Tinggi Badan 40,4 – 60 cm
2	Bayi Usia 4 - 6 Bulan: Tinggi Badan 60,5 – 66,0 cm
3	Bayi Usia 7 – 9 Bulan: Tinggi Badan: 67,5 – 70,5 cm
4	Bayi Usia 10 – 12 Bulan: Tinggi Badan: 72 – 74,5 cm
5	Balita Usia 13 – 24 Bulan: Tinggi Badan: 82 – 92 cm
6	Balita Usia 25 – 32 Bulan: Tinggi Badan: 83 – 95 cm
7	Balita Usia 33 – 44 Bulan: Tinggi Badan: 84 – 97 cm
8	Balita Usia 44 – 58 Bulan: Tinggi Badan: 85 – 98 cm

Gambar 39. Standar Anthropometri tinggi badan ideal berdasarkan usia
(Sumber: (Apriyani et al., 2023))

Berdasaran Standar Anthropometri dapat dikategorikan status gizi balita seperti Gambar 40.

cluster	anggota cluster	total
cluster 0	398	Jumlah balita gizi kurang : 67 Jumlah balita gizi normal : 224 Jumlah balita gizi lebih : 107
cluster 1	5	Jumlah balita gizi kurang : 2 Jumlah balita gizi normal : 3 Jumlah balita gizi lebih : -

Gambar 40. Informasi Jumlah Balita

SIMPULAN

Penelitian ini melakukan 9 kali percobaan dengan k yang berbeda. Dan diketahui dari 10 kali percobaan, nilai k yang paling optimal terdapat pada k = 2 dengan hasil DBI nya adalah 0.164. Berdasarkan hasil pengujian data menggunakan tools RapidMiner pada pengelompokan kasus Stunting balita di Desa Tegalwangi per akhir bulan Januari 2023 menggunakan algoritma K-Means Clustering melalui 3 parameter yaitu umur, berat badan dan tinggi badan dimana dibagi menjadi 2 cluster yaitu cluster 0 terdapat 392 balita yaitu Shanum, Rizka, Nurjanah, dan lainnya, cluster 1 terdapat 3 balita yaitu Ezra, M Abidza dan Abd Mahmud. Jumlah balita status Gizi Normal sebanyak 227 balita jumlah balita status Gizi kurang sebanyak 69 balita, dan jumlah status gizi lebih sebanyak 107 balita berdasarkan Standar Antropometri standar ideal balita. Berdasarkan hasil pengujian nilai Davies Bouldin Index (DBI) perhitungan menggunakan bantuan tools RapidMiner menghasilkan nilai evaluasi performa yang optimal berada di K=2 memiliki nilai 0.164 dimana nilai tersebut mendekati 0 yang berarti klaster yang di evaluasi menghasilkan Cluster yang baik.

Untuk penelitian selanjutnya diharapkan menggunakan perhitungan manual atau bahkan menggabungkan 2 hasil manual dan tools Rapid miner. Selain itu jumlah data bisa ditambah, peneliti memilih data di desa mekar wangi diharap bagi penelitian selanjutnya data perkecamatan yang ada di wilayah kabupaten/kota Bandung.

DAFTAR PUSTAKA

- Adiputra, I. N. M. (2022). Clustering Penyakit Dbd Pada Rumah Sakit Dharma Kerti Menggunakan Algoritma K-Means. *INSERT: Information System and Emerging Technology Journal*, 2(2), 99. <https://doi.org/10.23887/insert.v2i2.41673>
- Apriyani, P., Dikananda, A. R., & Ali, I. (2023). Penerapan Algoritma K-Means dalam Klasterisasi Kasus Stunting Balita Desa Tegalwangi. *Hello World Jurnal Ilmu Komputer*, 2(1), 20–33. <https://doi.org/10.56211/helloworld.v2i1.230>
- Gustrianda, R., & Mulyana, D. I. (2022).

- Penerapan Data Mining Dalam Pemilihan Produk Unggulan dengan Metode Algoritma K-Means Dan K-Medoids. *Jurnal Media Informatika Budidarma*, 6(1), 27. <https://doi.org/10.30865/mib.v6i1.3294>
- Irfiani, E., & Rani, S. S. (2018). Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita. *Jurnal Sistem Dan Teknologi Informasi (JUSTIN)*, 6(4), 161. <https://doi.org/10.26418/justin.v6i4.29024>
- Nabila, Z., Rahman Isnain, A., & Abidin, Z. (2021). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means. *Jurnal Teknologi Dan Sistem Informasi (JTSI)*, 2(2), 100. <http://jim.teknokrat.ac.id/index.php/JTSI>
- Zami, A. Z., Nurdian, O., & Dwilestari, G. (2022). Klasifikasi Kondisi Gizi Bayi Bawah Lima Tahun Pada Posyandu Melati Dengan Menggunakan Algoritma Decision Tree. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 3(3), 305. <https://doi.org/10.30865/json.v3i3.3892>