

Implementasi Algoritma K- Means Clustering untuk Mengelompokkan Data Gizi Balita pada Kecamatan Garoga Tapanuli Utara

Dody W Sitohang¹, Alex Rikki²

¹STIKOM MEDAN, Jl. Jamin Ginting No.285, Padang Bulan, Medan Baru, Kota Medan, Sumatera Utara, Indonesia

²Universitas Imelda Medan, Jl. Bilal Ujung No.24,52, Pulo Brayon Darat I, Kec. Medan, Sumatera Utara, Indonesia

ARTICLE INFORMATION

Received: Agustus 30,2019
Revised: September 26,2019
Available online: Oktober 05,2019

KEYWORDS

Data mining, Clustering, K-Means,

CORRESPONDENCE

Phone: +6282275847123
E-mail:dodywsitohang@gmail.com¹,
alexrikisinaga@gmail.com²

A B S T R A K

Kesenjangan taraf kehidupan masyarakat di wilayah Kecamatan Garoga sampai saat ini belum merata, masih banyak masyarakat mengkonsumsi gizi yang kurang baik, sehingga dapat menimbulkan gizi yang kurang baik terhadap bayi yang ada di Kecamatan Garoga. Algoritma data mining yang cukup populer digunakan baik dalam dunia bisnis, akademik, ataupun industri adalah algoritma K-Means. AlgoritmaK-Means clustering merupakan salah satu metode clustering non hirarki yang mempartisi data menjadi beberapa cluster/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Mengelompokkan pola gizi balita dengan Algoritma K-Means memang sangat rumit dan dibutuhkan ketelitian yang sangat tinggi dalam penyesuaiannya dengan menggunakan Software Rapidminer..

1. PENDAHULUAN

Kecamatan Garoga merupakan bagian dari salah satu kecamatan yang berada di Kabupaten Tapanuli Utara, Sumatra Utara. Ibukota kecamatan ini berada di desa Garoga Sibargot. Wilayah kecamatan Garoga memiliki luas wilayah 567,58km², kepadatan penduduk mencapai 28 jiwa/km². Dan memiliki jumlah Desa sebanyak 13 Desa. Kesenjangan taraf kehidupan masyarakat di wilayah Kecamatan Garoga sampai saat ini belum merata, masih banyak masyarakat yang kurang mengkonsumsi gizi yang baik, sehingga menimbulkan gizi yang kurang baik terhadap bayi yang ada di Kecamatan Garoga[1].

Algoritma data mining yang cukup populer digunakan baik dalam dunia bisnis, akademik, ataupun industri adalah algoritma K-Means. AlgoritmaK-Means clustering merupakan salah satu metode clustering non hirarki yang mempartisi data menjadi beberapa cluster/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain[2].

Oleh karena itu, guna mendukung kelancaran untuk mengelompokkan data gizi balita di kecamatan Garoga perlu menggunakan suatu sistem yang dapat memberi solusi dan mengefektifkan kinerja pada bagian kesehatan gizi balita. Sistem yang akan dibuat berfungsi untuk mengetahui kesehatan gizi baik, buruk dan sedang. Karena data kesehatan gizi balita yang kurang efektif akan berdampak pada terhambatnya proses mengelompokkan data gizi balita[3].

Salah satu teknik yang dapat digunakan untuk menentukan tingkat kesehatan bayi dan balita ini dapat dilakukan dengan teknik data mining. Dengan menggunakan teknik data mining, setiap kumpulan atau gudang data dapat memberikan pengetahuan penting menjadi informasi yang sangat berharga bagi suatu pemerintahan. Pada sebuah pemerintahan, suatu sistem dapat digunakan untuk memperoleh informasi yang menunjang setiap kegiatan pada pengambilan suatu keputusan. Data mining dilihat dari sisi teknik pengolahan data menyediakan sejumlah algoritma yang dapat digunakan untuk menggali informasi tersembunyi dari kumpulan data yang multidimensi.

Penelitian ini menggunakan algoritma K-Means untuk mengelompokkan atau mengklaster desa-desa yang ada di Kecamatan Garoga kabupaten Tapanuli Utara berdasarkan kemiripan karakteristik daerah yang ditinjau dari nilai indikator kesehatan yaitu angka kematian bayi dan balita, angka kesakitan bayi dan balita, dan status gizi bayi dan balita. Clustering merupakan pengelompokan data, observasi, atau kasus ke dalam kelas dari objek serupa. Clustering berbeda dengan *classification*, bahwa tidak ada variabel target clustering. Pengelompokan tidak mencoba untuk mengklasifikasikan memperkirakan, atau memprediksi nilai variabel sasaran. Sebaliknya, algoritma klasterisasi berusaha untuk mencari bagian dari seluruh dataset menjadi sub kelompok yang relatif homogen atau kelompok, dimana kesamaan data dalam cluster dimaksimalkan, dan kesamaan data yang di luar klaster diminimalkan[4].

2. LANDASAN TEORI

a. Data Mining

Data mining merupakan kiasan dari bahasa inggris, mine.jikamine berarti menambang sumber daya yang tersembunyi di dalam tanah, maka Data Mining merupakan penggalian makna yang tersembunyi dari kumpulan data yang sangat besar. Karena itu

Data Mining sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (artificial intelligent), machine learning, statistik dan basis Data[5].

Data mining adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya. Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data. Data mining adalah langkah analisis terhadap proses penemuan pengetahuan didalam basis data atau knowledge discovery in databases yang disingkat KDD[6]. Data mining adalah campuran dari statistik, kecerdasan buatan, dan riset basis data yang masih berkembang. Data Mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu:

1. Classification

Suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Salah satu contoh yang mudah dan populer adalah dengan Decision tree yaitu salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi. Decision tree adalah model prediksi menggunakan struktur pohon atau struktur berhirarki.

2. Association

Digunakan untuk mengenali kelakuan dari kejadiankejadian khusus atau proses dimana hubungan asosiasi muncul pada setiap kejadian. Salah satu contohnya adalah Market Basket Analysis, yaitu salah satu metode asosiasi yang menganalisa kemungkinan pelanggan untuk membeli beberapa item secara bersamaan.

3. Clustering

Digunakan untuk menganalisis pengelompokan berbeda terhadap data, mirip dengan klasifikasi, namun pengelompokan belum didefinisikan sebelum dijalankannya tool data mining. Biasanya menggunakan metode neural network atau statistik. Clustering membagi item menjadi kelompok-kelompok berdasarkan yang ditemukan tool data mining.

b. K-Means

Beberapa distance space dapat diimplementasikan untuk menghitung jarak (distance) antara data dan centroid termasuk di antaranya Manhattan/City Block Distance, Euclidean Distance dan Minkowski Distance Tetapi secara umum distance space yang sering digunakan adalah Manhattan dan Euclidean. Euclidean sering digunakan karena penghitungan jarak dalam distance space ini merupakan jarak terpendek yang bisa didapatkan antara dua titik yang diperhitungkan, sedangkan Manhattan sering digunakan karena kemampuannya dalam mendeteksi keadaan khusus seperti keberadaan outliers dengan lebih baik.

Objective Function adalah pernyataan kuantitatif dari kasus optimasi, sebagai contoh: memaksimalkan benefit, dengan menentukan biaya operasi minimum. Objective Function yang digunakan untuk metode Hard K-Means, adalah sebagai berikut:

$$J(U, V) = \sum_{k=1}^N \sum_{i=1}^c a_{ik} D(x_k, v_i)^2$$

Gambar 2.1. Objective function

Objective Function Hard K-Means dimana:

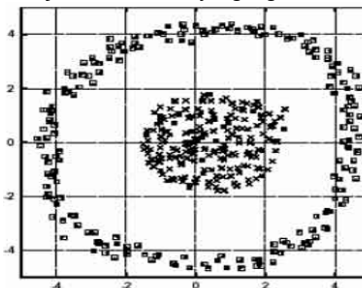
N : Jumlah data

c : Jumlah cluster

a_{ik} : Keanggotaan data ke-k ke cluster ke-i

v_i : Nilai centroid cluster ke-i

Nilai a_{ik} mempunyai nilai 0 atau 1. Apabila suatu data merupakan anggota suatu kelompok maka nilai $a_{ik}=1$ dan sebaliknya. Untuk metode Fuzzy K-Means Objective Function yang digunakan adalah sebagai berikut:



Gambar 1: Objective Function Fuzzy Means[6]

dimana:

N : Jumlah data

c : Jumlah cluster

m : Weighting exponent

u_{ik} : Membership function data ke-k ke cluster ke-i

v_i : Nilai centroid cluster ke-i

Di sini u_{ik} bisa mengambil nilai mulai dari 0 sampai 1.

Beberapa permasalahan yang sering muncul pada saat menggunakan metode K-Means untuk melakukan pengelompokan data adalah:

1. Ditemukannya beberapa model clustering yang berbeda
2. Pemilihan jumlah cluster yang paling tepat
3. Kegagalan untuk converge
4. Pendeteksian outliers
5. Bentuk masing-masing cluster
6. Masalah overlapping

Keenam permasalahan ini adalah beberapa hal yang perlu diperhatikan pada saat menggunakan K-Means dalam mengelompokkan data[2].:

Permasalahan 1, umumnya disebabkan oleh perbedaan proses inisialisasi anggota masing-masing cluster. Proses inisialisasi yang sering digunakan adalah proses inisialisasi secara random. Dalam suatu studi perbandingan, proses inisialisasi secara random mempunyai kecenderungan untuk memberikan hasil yang lebih baik dan independent, walaupun dari segi kecepatan untuk convergen lebih lambat.

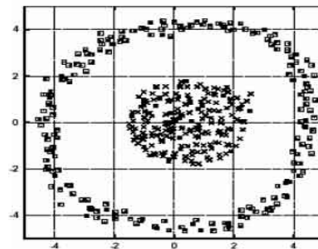
Permasalahan 2, merupakan masalah laten dalam metode K-Means. Beberapa pendekatan telah digunakan dalam menentukan jumlah cluster yang paling tepat untuk suatu dataset yang dianalisa termasuk di antaranya Partition Entropy (PE) dan GAP Statistics.

Permasalahan 3, kegagalan untuk converge, secara teori memungkinkan untuk terjadi dalam metode Hard K-Means maupun Fuzzy K-Means. Kemungkinan ini akan semakin besar terjadi untuk metode Hard K-Means, karena setiap data di dalam dataset dialokasikan secara tegas (hard) untuk menjadi bagian dari suatu cluster tertentu. Perpindahan suatu data ke suatu cluster tertentu dapat mengubah karakteristik model clustering yang dapat menyebabkan data yang telah dipindahkan tersebut lebih sesuai untuk berada di cluster semula sebelum data tersebut dipindahkan. Demikian juga dengan keadaan sebaliknya.

Kejadian seperti ini tentu akan mengakibatkan pemodelan tidak akan berhenti dan kegagalan untuk converge akan terjadi. Untuk Fuzzy K-Means, walaupun ada, kemungkinan permasalahan ini untuk terjadi sangatlah kecil, karena setiap data dilengkapi dengan membership function (Fuzzy K-Means) untuk menjadi anggota cluster yang ditemukan.

Permasalahan 4, merupakan permasalahan umum yang terjadi hampir di setiap metode yang melakukan pemodelan terhadap data. Khusus untuk metode K-Means hal ini memang menjadi permasalahan yang cukup menentukan. Beberapa hal yang perlu diperhatikan dalam melakukan pendeteksian outliers dalam proses pengelompokan data termasuk bagaimana menentukan apakah suatu data item merupakan outliers dari suatu cluster tertentu dan apakah data dalam jumlah kecil yang membentuk suatu cluster tersendiri dapat dianggap sebagai outliers. Proses ini memerlukan suatu pendekatan khusus yang berbeda dengan proses pendeteksian outliers di dalam suatu dataset yang hanya terdiri dari satu populasi yang homogen.

Permasalahan 5, menyangkut bentuk cluster yang ditemukan. Tidak seperti metode data clustering lainnya, K-Means umumnya tidak menginginkan bentuk dari masing-masing cluster yang mendasari model yang terbentuk, walaupun secara natural masing-masing cluster umumnya berbentuk bundar. Untuk dataset yang diperkirakan mempunyai bentuk yang tidak biasa, beberapa pendekatan perlu untuk diterapkan.



Gambar 2: Salah Satu Dataset yang Mempunyai Bentuk Khusus[4]

Permasalahan 6, masalah overlapping sebagai permasalahan terakhir sering sekali diabaikan karena umumnya masalah ini sulit terdeteksi. Hal ini terjadi untuk metode Hard K-Means dan Fuzzy K-Means, karena secara teori, metode ini tidak dilengkapi feature untuk mendeteksi apakah di dalam suatu cluster ada cluster lain yang kemungkinan tersembunyi.

c. Pengertian data gizi balita

Gizi adalah suatu proses menggunakan makanan yang dikonsumsi secara normal melalui proses digesti, absorpsi, transportasi, penyimpanan, metabolisme dan pengeluaran zat-zat yang tidak digunakan untuk mempertahankan kehidupan, pertumbuhan dan fungsi normal dari organ-organ serta menghasilkan energi[7].

Balita adalah kelompok anak yang berumur dibawah lima tahun. Kelompok anak ini menjadi istimewa karena menuntut perhatian yang intensif untuk mengoptimalkan pertumbuhan dan perkembangannya. Lima tahun pertama dari kehidupan seorang manusia adalah fondasi bagi seluruh kehidupan di dunia. Sumber daya manusia yang berkualitas baik fisik, psikis, maupun intelegensinya berawal dari balita yang sehat. Balita adalah anak usia dibawah lima tahun yang berumur 0-4 tahun 11 bulan[8].

Penilaian Status Gizi (PSG) adalah sebuah metode mendeskripsikan kondisi tubuh sebagai akibat keseimbangan makanan yang dikonsumsi dengan penggunaannya oleh tubuh, yang biasanya dibandingkan dengan suatu nilai normatif yang ditetapkan (WHO, 2005). Penilaian status gizi dengan menggunakan data antropometri antara lain berat badan menurut umur (BB/U), tinggi badan menurut umur (TB/U), berat badan menurut tinggi badan (BB/TB), dan indeks masa tubuh menurut umur (IMT/U) World Health Organization (WHO) merekomendasikan pengukuran antropometri pada bayi dan balita menggunakan grafik yang dikembangkan oleh WHO dan center for disease Control and Prevention (CDC). Grafik tersebut menggunakan indikator Z-score sebagai standar deviasi rata-rata dan persentil median. Indikator pertumbuhan digunakan untuk menilai pertumbuhan anak dengan mempertimbangkan faktor umum dan hasil pengukuran tinggi badan dan berat badan, lingkaran kepala dan lingkaran lengan atas. Indeks yang umum digunakan untuk menemukan status gizi bayi dan balita adalah sebagai berikut:

1. Berat badan menurut umur (BB/U)
2. Panjang atau tinggi badan menurut umur (PB/U atau TB/U)
3. Berat badan menurut panjang badan atau tinggi badan (BB/PB atau BB/TB)
4. Indeks massa tubuh menurut umur (IMT/U)

Cara penghitungan rumus Z skor / Standar Deviasi sebagai berikut.

Rumus perhitungan **Z-skor** = **Nilai individu subyek** – **Nilai Median Baku Rujukan**

Keterangan: Nilai simpangan baku rujukan

Nilai simpangan baku rujukan untuk BB/U adalah 11% dari nilai median baku rujukan, simpangan baku rujukan untuk BB/TB adalah 10% dari nilai median baku rujukan, serta simpangan baku rujukan untuk TB/U adalah 5% dari nilai median baku rujukan. Berikut tabel baku rujukan BB/U.

Tabel 1. Tabel Baku Rujukan Penilaian Status Gizi Anak Perempuan dan Laki-laki usia 48-59 Bulan Menurut Berat Badan dan Umur (BB/U)

Umur (bulan)	Median Berat Badan Anak Laki-Laki (kg)	Median Berat Badan Anak Perempuan (kg)
48	16,7	16,0
49	16,9	16,1
50	17,0	16,2
51	17,2	16,4
52	17,4	16,5
53	17,5	16,7
54	17,7	16,8
55	17,9	17,0
56	18,0	17,1
57	18,2	17,2
58	18,3	17,4
59	18,5	17,5

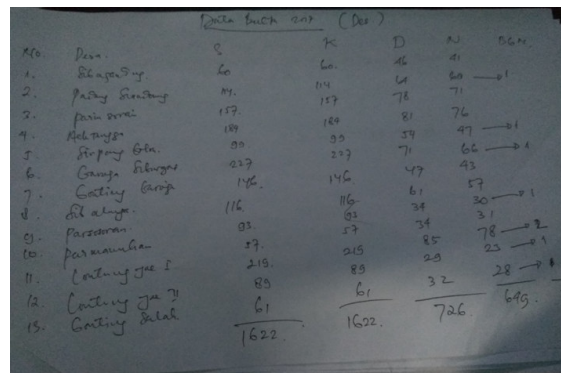
Sumber: Supariasa (2002).

3. METODOLOGI PENELITIAN

Dalam melaksanakan penelitian ini, objek penelitian ini dilakukan di UPT Puskesmas Garoga Kecamatan Garoga Kabupaten Tapanuli Utara. Data – data yang diperlukan dalam penelitian ini adalah data yang berhubungan dengan penelitian data yang dibutuhkan antara lain :

1. Data Primer

Data primer yaitu data yang digunakan untuk proses pengelompokan gizi balita yang diperoleh dari bagian yang menangani perkembangan gizi balita di UPT Puskesmas Garoga yang bernama Rosita Limbong. Dari hasil observasi dan wawancara data yang dikumpulkan sebagai berikut



No	Nama	Umur (Bulan)	BB (kg)	TB (cm)	IMT (kg/m ²)	Stad
1.	Bela	60	60	46	41	0,64
2.	Paity Siantang	44	114	64	60	→ 1
3.	Paity Siantang	44	114	78	71	
4.	Paity Siantang	157	184	81	76	
5.	Paity Siantang	184	227	71	47	→ 1
6.	Paity Siantang	227	277	71	66	→ 1
7.	Paity Siantang	146	146	47	43	
8.	Paity Siantang	116	116	61	57	
9.	Paity Siantang	83	83	34	30	→ 1
10.	Paity Siantang	77	77	34	31	
11.	Paity Siantang	215	215	85	78	→ 2
12.	Paity Siantang	85	85	23	23	→ 1
13.	Paity Siantang	61	61	32	28	→ 1
14.	Paity Siantang	61	61	32	28	→ 1
15.	Paity Siantang	61	61	32	28	→ 1
		1622	1622	726	685	

Gambar 3: Data Primer

2. Data Sekunder

Data sekunder yaitu data yang diperoleh dari buku – buku literatur yang berkaitan dengan objek yang diteliti

a. Metode Pungumpulan Data

1. Metode Observasi

Metode Observasi yaitu cara pengumpulan data dengan melakukan pengamatan dan pencatatan secara sistematis mengenai data – data gizi balita di UPT Puskesmas Garoga

2. Metode Wawancara

Metode wawancara yaitu cara pengumpulan data dengan melakukan tanya jawab langsung kepada bagian penanganan perkembangan gizi balita di UPT Puskesmas. Adapun beberapa Pertanyaan yang saya berikan sebagai berikut :

- Apa yang dilakukan petugas, jika menemukan anak yang mengalami kekurangan gizi ?
- Menurut bapak, apa factor penyebab kekurangan gizi pada anak ?
- Apa ciri- ciri kekurangan gizi balita ?
- Apa yang dilakukan petugas, jika menemukan anak yang mengalami kurang gizi ?

3. Metode Kepustakaan

Metode ini dilakukan dengan cara mengambil bahan dan literature serta sumber lain yang mendukung dan berhubungan dengan topic penelitian yaitu jurnal yang membahas tentang Algoritma K - Means Clustering.

b. Metode Analisa Data

Metode analisa data yang digunakan untuk mengelompokkan data gizi balita di kecamatan Garoga menggunakan Algoritma K – Means clustering.

d. Diagram Alir Langkah Penelitian

Penelitian merupakan proses serangkaian langkah-langkah yang dilakukan secara terencana dan sistematis untuk mendapatkan pemecahan masalah atau mendapatkan jawaban. Dalam melakukan penelitian ini terdapat langkah-langkah penelitian yaitu penulis menentukan objek dan lokasi, mengidentifikasi masalah, membuat rumusan, dan batasan masalah, menentukan tujuan dan kegunaan penelitian. Penulis mengumpulkan data-data yang diperlukan mengolah data, menganalisa serta merancang dan menerapkan hasil penelitian.



Gambar 4 : Diagram alir langkah penelitian

4. HASIL DAN PEMBAHASAN

Proses pengujian dilakukan untuk menentukan pengelompokan UPT Puskesmas Garoga Kecamatan Garoga Kabupaten Tapanuli Utara menggunakan algoritma *K-Means Clustering*. Algoritma *K-Means Clustering* digunakan untuk dapat menentukan jumlah *cluster* yang tepat untuk masing-masing kelompok gizi baik, buruk dan sedang.

Implementasi proses Algoritma *K-Means Clustering* pada aplikasi *Rapid Miner 7.3* mencakup spesifikasi *hardware* dan *software* yang dibutuhkan. Berhubungan dengan aplikasi yang digunakan dalam pengujian adalah aplikasi yang menggunakan salah satu *database Microsoft Excel*. Adapun tahapan evaluasi pengelompokan UPT Puskesmas Garoga Kecamatan Garoga Kabupaten Tapanuli Utara.

a. Analisa Data

Pada UPT Puskesmas Garoga Kecamatan Garoga Kabupaten Tapanuli Utara terdapat 13 desa dan semua data balita di tiap desa akan dijadikan sampel untuk penerapan Data Mining. Data-data tersebut tersedia dalam file *Microsoft Excel* yang telah

diinput oleh penulis. Hal ini di karenakan pada proses pengujian, aplikasi *Rapid Miner* dapat mengolah file tipe *.xls*. Adapun data yang akan diolah pada penelitian ini dapat dilihat pada table 2 berikut:

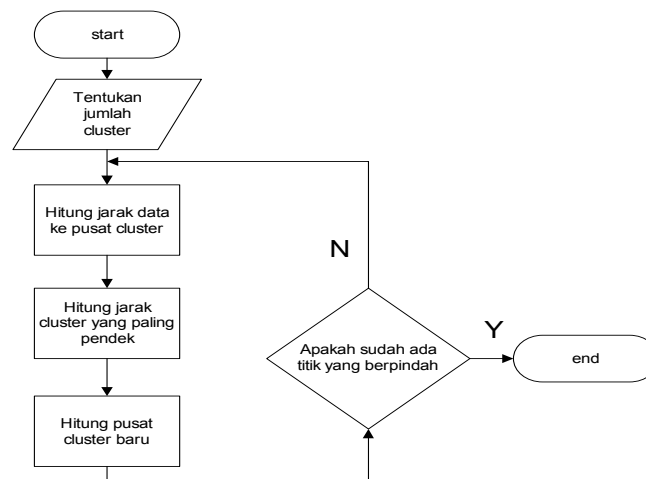
Tabel 2. UPT Puskesmas Garoga Kecamatan Garoga Kabupaten Tapanuli Utara

NO	NAMA BALITA	BERAT BADAN (KG)	TINGGI BADAN (cm)
1	Natalia Pasaribu	2.5	30.2
2	Mira Marpaung	2.8	48.8
3	Andreas Sormin	3.1	50.8
4	Alfredo sormin	3.2	50.8
5	Anroi Sipahutar	3.4	50.3
6	Enjelika Sipahutar	2.7	4.1
7	Muhammad Pirmansha Pasaribu	3.4	49.8
8	Sartika Goltom	3.2	41.8
9	Kristop Lubis	3.7	50.1
10	Ayu Pane	3.6	45.6
11	Nurhidayah	4.2	53.7
12	Dewi Pakpahan	4.3	50.5
13	Esti Sibarani	4.1	48.7
14	Exael Situmorang	4.2	52.7
15	Uba Wahyu Pakpahan	4.5	54.7
16	Konida Munthe	4.1	52.8
17	Enjel sidabutar	4.6	54.5
18	Wanris Ritonga	4.3	51.6
19	Gomgom Sabar Hasibuan	4.8	56.7
20	Halasson	5.1	56.7
21	Wiwin Pasaribu	4.8	54.2
22	Laura Lara Hasibuan	4.8	55.7
23	Joel Hasibuan	4.9	50.4
24	Rodami Parsaoran Silitonga	4.2	53.7
25	Putra Sudiantono	5.4	55.5

Variabel yang digunakan pada penelitian ini adalah variable Berat badan dan Tinggi badan dan selanjutnya akan dilakukan penentuan jumlah *cluster* dan titik pusat *cluster* awal yang dilakukan secara random.

b. Analisa *K-Means Clustering*

Metode yang digunakan penulis pada penelitian ini adalah metode *Data Mining* algoritma *K-Means Clustering* dimana proses utama adalah segmentasi atau pengelompokan data balita yang terdapat pada UPT Puskesmas Garoga Kecamatan Garoga Kabupaten Tapanuli Utara. Dan pada tahap ini penulis mengelompokkan data gizi balita sebanyak 3 kelompok, yakni data gizi baik, buruk, dan sedang. Berikut ini merupakan diagram flowchart dari algoritma k-means clustering dengan asumsi bahwa parameter input adalah jumlah data set sebanyak n data dan jumlah inisialisasi centroid k=3 sesuai dengan penelitian yang dilakukan oleh penulis.



Gambar 5: Flowchart Algoritma *K-Means Clustering*

Pada gambar 5 dijelaskan proses algoritma k-means untuk menyelesaikan masalah sebagai berikut :

1. Tentukan jumlah cluster, untuk menentukan banyaknya jumlah *cluster* atau *cluster* dilakukan dengan beberapa pertimbangan teoritis dan konseptual yang mungkin diusulkan untuk menentukan beberapa banyak *cluster* yang harus dibentuk.
2. Hitung jarak setiap data ke pusat *cluster*, untuk menghitung jarak antara objek dengan *centroid* menggunakan rumus *euclidean distance*.
3. Kelompokkan data ke dalam *cluster* yang dengan jarak yang paling pendek, untuk menghitung jarak *cluster* yang pendek menggunakan rumus *cluster* jarak.
4. Hitung pusat *cluster* baru, untuk menghitung pusat *cluster* baru menggunakan rumus *cluster* pusat.
5. Jika hasil dari proses diatas telah menunjukkan bahwa tidak ada lagi objek yang berpindah *cluster* dan algoritma telah stabil, maka proses dinyatakan selesai. Jika belum ulangi langkah 2 sampai dengan langkah 4 hingga sudah tidak ada lagi data yang berpindah.

c. Pembahasan

Dalam menganalisa kebutuhan dengan menggunakan algoritma *K-Means Clustering* yakni mengelompokkan data berdasarkan nilai rata-rata data yang diteliti dibutuhkan *cluster* yang merupakan kelompok data dan titik awal *cluster*. Pada penelitian ini, penulis melakukan 3 *cluster*. Hal ini merupakan tahap awal untuk menentukan hasil yakni kelompok. Menentukan titik pusat awal *cluster* pada metode *Data Mining* algoritma *K-Means Clustering* adalah dipilih secara random. Adapun titik pusat awal *cluster* yang dipilih oleh penulis secara random dapat dilihat pada tabel 3 berikut:

Tabel 3. Data dengan titik awal *cluster*

NO	NAMA BALITA	BERAT BADAN (KG)	TINGGI BADAN (cm)
1	Natalia Pasaribu	2.5	30.2
2	Mira Marpaung	2.8	48.8
3	Andreas Sormin	3.1	50.8
4	Alfredo sormin	3.2	50.8
5	Anroi Sipahutar	3.4	50.3
6	Enjelika Sipahutar	2.7	4.1
7	Muhammad Pirmansha Pasaribu	3.4	49.8
8	Sartika Goltom	3.2	41.8
9	Kristop Lubis	3.7	50.1
10	Ayu Pane	3.6	45.6
11	Nurhidayah	4.2	53.7
12	Dewi Pakpahan	4.3	50.5
13	Esti Sibarani	4.1	48.7
14	Exael Situmorang	4.2	52.7
15	Uba Wahyu Pakpahan	4.5	54.7
16	Konida Munthe	4.1	52.8
17	Enjel sidabutar	4.6	54.5
18	Wanris Ritonga	4.3	51.6
19	Gomgom Sabar Hasibuan	4.8	56.7
20	Halasson	5.1	56.7
21	Wiwin Pasaribu	4.8	54.2
22	Laura Lara Hasibuan	4.8	55.7
23	Joel Hasibuan	4.9	50.4
24	Rodami Parsaoran Silitonga	4.2	53.7
25	Putra Sudiantono	5.4	55.5

Dari table 3 tersebut, dapat diambil kesimpulan masing-masing titik awal *Cluster* berdasarkan urutan yang akan dilakukan penulis yakni daerah rawan tindak kriminal, sedang dan tidak rawan tindak kriminal adalah sebagai berikut:

- a. Pusat *Cluster I*: Data ke-3 dengan BB=3.1 dan TB=50.8
- b. Pusat *Cluster II*: Data ke-25 dengan BB=5.4 dan TB= 55.5
- c. Pusat *Cluster III*: Data ke-1 dengan BB=2.5 dan TB=30.2

Pada algoritma *K-Means Clustering*, menghitung jarak setiap data ke pusat *cluster* adalah bertujuan untuk menentukan titik pusat *cluster* yang baru menggunakan rumus *Euclidian Distance* sebagai berikut:

$$D(ik) = \sqrt{\sum_j^m (Cij - Ckj)^2}$$

Dimana: Cij = Pusat *cluster*
Ckj = Data

Pada tahapan ini, perhitungan jarak setiap data ke titik pusat *cluster* dilakukan secara perulangan. Dan apabila nilai pusat *cluster* pada iterasi tersebut sama dengan nilai pusat *cluster* pada iterasi sebelumnya, perulangan dihentikan karena hasil telah ditemukan. Adapun perhitungan jarak data ke pusat *cluster* pada masing-masing iterasi adalah sebagai berikut:

- a. Iterasi 1

$$\begin{aligned} \text{Data-1 ke Pusat Cluster I} &= \sqrt{(2.5 - 3.1)^2 + (30.2 - 50.8)^2} \\ &= 20.61 \end{aligned}$$

Dan seterusnya sampai

$$\begin{aligned} \text{Data-25 ke Pusat Cluster I} &= \sqrt{(5.4 - 3.1)^2 + (55.5 - 50.8)^2} \\ &= 5.23 \end{aligned}$$

$$\begin{aligned} \text{Data-1 ke Pusat Cluster II} &= \sqrt{(2.5 - 5.4)^2 + (30.2 - 55.5)^2} \\ &= 25.5 \end{aligned}$$

Dan seterusnya sampai

$$\begin{aligned} \text{Data-25 ke Pusat Cluster III} &= \sqrt{(5.4 - 2.5)^2 + (55.5 - 30.2)^2} \\ &= 25.47 \end{aligned}$$

Hasil perhitungan menggunakan *Euclidian Distance* pada iterasi I tersebut diatas dapat dilihat pada table 4. berikut:

Tabel 4 Hasil Iterasi I

NAMA BALITA	BERAT BADAN (KG)	TINGGI BADAN (cm)	C0	C1	C2
Natalia Pasaribu	2.5	30.2	20.61	25.5	0
Mira Marpaung	2.8	48.8	2.02	7.2	18.60242
Andreas Sormin	3.1	50.8	0.00	5.2	20.60874
Alfredo sormin	3.2	50.8	0.10	5.2	20.61189
Anroi Sipahutar	3.4	50.3	0.58	5.6	20.12014
Enjelika Sipahutar	2.7	4.1	46.70	51.5	26.10077
Muhammad Pirmansha Pasaribu	3.4	49.8	1.04	6.0	19.62065
Sartika Goltom	3.2	41.8	9.00	13.9	11.6211
Kristop Lubis	3.7	50.1	0.92	5.7	19.93615
Ayu Pane	3.6	45.6	5.22	10.1	15.43924
Nurhidayah	4.2	53.7	3.10	2.2	23.56141
Dewi Pakpahan	4.3	50.5	1.24	5.1	20.37965
Esti Sibarani	4.1	48.7	2.33	6.9	18.56906
Exael Situmorang	4.2	52.7	2.20	3.0	22.56413
Uba Wahyu Pakpahan	4.5	54.7	4.14	1.2	24.5815
Konida Munthe	4.1	52.8	2.24	3.0	22.65657
Enjel sidabutar	4.6	54.5	3.99	1.3	24.39057
Wanris Ritonga	4.3	51.6	1.44	4.1	21.47557
Gomgom Sabar Hasibuan	4.8	56.7	6.14	1.3	26.59962
Halasson	5.1	56.7	6.23	1.2	26.62724
Wiwin Pasaribu	4.8	54.2	3.80	1.4	24.10996
Laura Lara Hasibuan	4.8	55.7	5.19	0.6	25.60352
Joel Hasibuan	4.9	50.4	1.84	5.1	20.34207
Rodami Parsaoran Silitonga	4.2	53.7	3.10	2.2	23.56141
Putra Sudiantono	5.4	55.5	5.23	0.0	25.46566

Tahap selanjutnya adalah menentukan jarak terpendek data terhadap pusat *cluster* awal yang dipilih secara random sebelumnya. Menentukan jarak terpendek dengan pusat *cluster* data dapat menggunakan rumus $=\text{MIN}(E10,F10,G10)$. Adapun hasil penentuan jarak terpendek terhadap pusat *cluster* yang telah ditabulasikan dapat dilihat pada tabel 5 berikut:

Tabel 5. Tabulasi Binner Jarak Terdekat Data terhadap *Cluster*

NAMA BALITA	C0	C1	C2	Cluster
Natalia Pasaribu	0	0	1	C2
Mira Marpaung	1	0	0	C0
Andreas Sormin	1	0	0	C0
Alfredo sormin	1	0	0	C0
Anroi Sipahutar	1	0	0	C0
Enjelika Sipahutar	0	0	1	C2
Muhammad Pirmansha Pasaribu	1	0	0	C0
Sartika Goltom	1	0	0	C0
Kristop Lubis	1	0	0	C0
Ayu Pane	1	0	0	C0
Nurhidayah	0	1	0	C1
Dewi Pakpahan	1	0	0	C0
Esti Sibarani	1	0	0	C0

NAMA BALITA	C0	C1	C2	Cluster
Exael Situmorang	1	0	0	C0
Uba Wahyu Pakpahan	0	1	0	C1
Konida Munthe	1	0	0	C0
Enjel sidabutar	0	1	0	C1
Wanris Ritonga	1	0	0	C0
Gomgom Sabar Hasibuan	0	1	0	C1
Halasson	0	1	0	C1
Wiwin Pasaribu	0	1	0	C1
Laura Lara Hasibuan	0	1	0	C1
Joel Hasibuan	1	0	0	C0
Rodami Parsaoran Silitonga	0	1	0	C1
Putra Sudiantono	0	1	0	C1

Berdasarkan nilai minimum yang telah ditentukan menggunakan rumus dan hasil terlihat pada table 5 tersebut di atas, maka diperoleh hasil pengelompokan seperti terlihat pada tabel 6 berikut :

Tabel 6 Hasil Pengelompokan

Kelompok (Cluster)	Anggota Cluster	Jumlah Anggota
0	{2,3,4,5,7,8,9,10,12,13,14,16,18,23}	14
1	{11,15,17,19,20,21,22,24,25}	9
2	{1,6}	2

Tahapan selanjutnya adalah menentukan titik pusat *cluster* baru yakni dengan mencari rata-rata semua anggota *cluster* dengan cara menjumlahkan seluruh *value* anggota *cluster* dibagi jumlah anggota. Adapun titik pusat *cluster* baru beserta *value* setiap variable adalah sebagai berikut:

$$\begin{aligned} \sum \text{value}_{C_0}(X1) &= \frac{\sum \text{value}_{C_0} BB}{\text{jumlah anggota } C_0} \\ &= \frac{(52.3)}{14} \\ &= 3.73571429 \end{aligned}$$

$$\begin{aligned} \sum \text{value}_{C_0}(X2) &= \frac{\sum \text{value}_{C_0} TB}{\text{jumlah anggota } C_0} \\ &= \frac{694.7}{14} \\ &= 49.62142857 \end{aligned}$$

$$\begin{aligned} \sum \text{value}_{C_1}(X1) &= \frac{\sum \text{value}_{C_1} BB}{\text{jumlah anggota } C_1} \\ &= \frac{(42.4)}{9} \\ &= 4.71111111 \end{aligned}$$

$$\begin{aligned} \sum \text{value}_{C_1}(X2) &= \frac{\sum \text{value}_{C_1} TB}{\text{jumlah anggota } C_1} \\ &= \frac{(495.4)}{9} \\ &= 55.04444444 \end{aligned}$$

$$\begin{aligned} \sum \text{value}_{C_2}(X1) &= \frac{\sum \text{value}_{C_2} BB}{\text{jumlah anggota } C_2} \\ &= \frac{(5.2)}{2} \\ &= 2.6 \end{aligned}$$

$$\begin{aligned} \sum \text{value}_{C_2}(X2) &= \frac{\sum \text{value}_{C_2} TB}{\text{jumlah anggota } C_2} \\ &= \frac{34.3}{2} \\ &= 17.15 \end{aligned}$$

Berdasarkan proses perhitungan di atas, maka diperoleh *centroid* baru sebagai tolak ukur pada iterasi selanjutnya dengan nilai sebagai berikut:

$$C_0 = [3.73571429 ; 49.62142857]$$

$$C_1 = [4.71111111 ; 55.04444444]$$

$$C_2 = [2.6 ; 17.15]$$

b. iterasi 2

Data-1 ke Pusat Cluster 1

$$= \sqrt{(2.5 - 3.73571429)^2 + (30.2 - 49.62142857)^2} = 19.46$$

Dan seterusnya sampai

Data-25 ke Pusat *Cluster I*

$$= \sqrt{(5.4 - 3.73571429)^2 + (55.5 - 49.62142857)^2} = 6.11$$

Data-1 ke Pusat *Cluster II*

$$= \sqrt{(2.5 - 4.71111111)^2 + (30.2 - 55.0444444)^2} = 24.9$$

Dan seterusnya sampai

Data-25 ke Pusat *Cluster II*

$$= \sqrt{(5.4 - 4.71111111)^2 + (55.5 - 55.0444444)^2} = 0.8$$

Data-1 ke Pusat *Cluster III* $= \sqrt{(2.5 - 2.6)^2 + (30.2 - 17.15)^2}$
 $= 13.05038$

sampai

Data-25 ke Pusat *Cluster III* $= \sqrt{(5.4 - 2.6)^2 + (55.5 - 17.15)^2}$
 $= 38.45208$

Hasil perhitungan menggunakan *Euclidian Distance* pada iterasi II tersebut diatas dapat dilihat pada tabel 7 berikut:

Tabel 7 Hasil Iterasi II

NAMA BALITA	BERAT BADAN (KG)	TINGGI BADAN (cm)	C1	C2	c3
Natalia Pasaribu	2.5	30.2	19.46	24.9	13.05038
Mira Marpaung	2.8	48.8	1.25	6.5	31.65063
Andreas Sormin	3.1	50.8	1.34	4.5	33.65371
Alfredo sormin	3.2	50.8	1.29	4.5	33.65535
Anroi Sipahutar	3.4	50.3	0.76	4.9	33.15965
Enjelika Sipahutar	2.7	4.1	45.53	51.0	13.05038
Muhammad Pirmansha Pasaribu	3.4	49.8	0.38	5.4	32.6598
Sartika Goltom	3.2	41.8	7.84	13.3	24.6573
Kristop Lubis	3.7	50.1	0.48	5.0	32.96836
Ayu Pane	3.6	45.6	4.02	9.5	28.46757
Nurhidayah	4.2	53.7	4.10	1.4	36.585
Dewi Pakpahan	4.3	50.5	1.04	4.6	33.3933
Esti Sibarani	4.1	48.7	0.99	6.4	31.58564
Exael Situmorang	4.2	52.7	3.11	2.4	35.58599
Uba Wahyu Pakpahan	4.5	54.7	5.14	0.4	37.59804
Konida Munthe	4.1	52.8	3.20	2.3	35.68154
Enjel sidabutar	4.6	54.5	4.95	0.6	37.40351
Wanris Ritonga	4.3	51.6	2.06	3.5	34.49192
Gomgom Sabar Hasibuan	4.8	56.7	7.16	1.7	39.61114
Halasson	5.1	56.7	7.21	1.7	39.62894
Wiwin Pasaribu	4.8	54.2	4.70	0.8	37.11526
Laura Lara Hasibuan	4.8	55.7	6.17	0.7	38.61272
Joel Hasibuan	4.9	50.4	1.40	4.6	33.32945
Rodami Parsaoran Silitonga	4.2	53.7	4.10	1.4	36.585
Putra Sudiantono	5.4	55.5	6.11	0.8	38.45208

Tahap selanjutnya adalah menentukan jarak terpendek data terhadap pusat *cluster* awal yang dipilih secara random sebelumnya. Menentukan jarak terpendek dengan pusat *cluster* data dapat menggunakan rumus $=\text{MIN}(K32, L32, M132)$. Adapun hasil penentuan jarak terpendek terhadap pusat *cluster* yang telah ditabulasikan dapat dilihat pada tabel 8 berikut:

Tabel 8 Tabulasi Binner Jarak Terdekat Data terhadap *Cluster* Baru

NO	NAMA BALITA	C0	C1	C2
1	Natalia Pasaribu	0	0	1
2	Mira Marpaung	0	0	1
3	Andreas Sormin	1	0	0
4	Alfredo sormin	1	0	0
5	Anroi Sipahutar	1	0	0
6	Enjelika Sipahutar	1	0	0
7	Muhammad Pirmansha Pasaribu	0	0	1
8	Sartika Goltom	1	0	0
9	Kristop Lubis	1	0	0
10	Ayu Pane	1	0	0
11	Nurhidayah	1	0	0

NO	NAMA BALITA	C0	C1	C2
12	Dewi Pakpahan	0	1	0
13	Esti Sibarani	1	0	0
14	Exael Situmorang	1	0	0
15	Uba Wahyu Pakpahan	0	1	0
16	Konida Munthe	0	1	0
17	Enjel sidabutar	0	1	0
18	Wanris Ritonga	0	1	0
19	Gomgom Sabar Hasibuan	1	0	0
20	Halasson	0	1	0
21	Wiwin Pasaribu	0	1	0
22	Laura Lara Hasibuan	0	1	0
23	Joel Hasibuan	0	1	0
24	Rodami Parsaoran Silitonga	1	0	0
25	Putra Sudiantono	0	1	0

Berdasarkan hasil table tersebut di atas, adapun titik pusat *cluster* baru adalah sebagai berikut:

$$\begin{aligned}\sum \text{value} C_0(X1) &= \frac{\sum \text{value} C_0 BB}{\text{jumlah anggota } c0} \\ &= \frac{(44)}{12} \\ &= 3.666666667\end{aligned}$$

$$\begin{aligned}\sum \text{value} C_0(X2) &= \frac{\sum \text{value} C_0 TB}{\text{jumlah anggota } c0} \\ &= \frac{589.2}{12} \\ &= 49.1\end{aligned}$$

$$\begin{aligned}\sum \text{value} C_1(X1) &= \frac{\sum \text{value} C_1 BB}{\text{jumlah anggota } c1} \\ &= \frac{(50.7)}{11} \\ &= 4.609090909\end{aligned}$$

$$\begin{aligned}\sum \text{value} C_1(X2) &= \frac{\sum \text{value} C_1 TB}{\text{jumlah anggota } c1} \\ &= \frac{(600.9)}{11} \\ &= 54.62727273\end{aligned}$$

$$\begin{aligned}\sum \text{value} C_2(X1) &= \frac{\sum \text{value} C_2 BB}{\text{jumlah anggota } c2} \\ &= \frac{(5.2)}{2} \\ &= 2.6\end{aligned}$$

$$\begin{aligned}\sum \text{value} C_2(X2) &= \frac{\sum \text{value} C_2 TB}{\text{jumlah anggota } c2} \\ &= \frac{34.3}{2} \\ &= 17.15\end{aligned}$$

Berdasarkan proses perhitungan di atas, maka diperoleh *centroid* baru sebagai tolak ukur pada iterasi selanjutnya dengan nilai sebagai berikut:

$$C_0 = [3.666666667 ; 49.1]$$

$$C_1 = [4.609090909 ; 54.62727273]$$

$$C_2 = [2.6 ; 17.15]$$

c. Iterasi 3

Data-1 ke Pusat *Cluster I*

$$= \sqrt{(2.5 - 3.666666667)^2 + (30.2 - 49.1)^2} = 18.94$$

Dan seterusnya sampai

Data-25 ke Pusat *Cluster I*

$$= \sqrt{(5.4 - 3.666666667)^2 + (55.5 - 49.1)^2} = 6.63$$

Data-1 ke Pusat *Cluster II*

$$= \sqrt{(2.5 - 4.609090909)^2 + (30.2 - 54.62727273)^2} = 24.5$$

Dan seterusnya sampai

Data-25 ke Pusat *Cluster II*

$$= \sqrt{(5.4 - 4.609090909)^2 + (55.5 - 54.62727273)^2} = 1.2$$

$$\begin{aligned} \text{Data-1 ke Pusat Cluster III} &= \sqrt{(2.5 - 2.6)^2 + (30.2 - 17.15)^2} \\ \text{sampai} &= 13.05038 \end{aligned}$$

$$\begin{aligned} \text{Data-25 ke Pusat Cluster III} &= \sqrt{(5.4 - 2.6)^2 + (55.5 - 17.15)^2} \\ &= 38.45208 \end{aligned}$$

Hasil perhitungan menggunakan *Euclidian Distance* pada iterasi III tersebut diatas dapat dilihat pada tabel 9 berikut:

Tabel 9 Hasil Iterasi III

NAMA BALITA	BERAT BADAN (KG)	TINGGI BADAN (cm)	C0	C1	C2
Natalia Pasaribu	2.5	30.2	18.94	24.5	13.05038314
Mira Marpaung	2.8	48.8	0.92	6.1	31.65063191
Andreas Sormin	3.1	50.8	1.79	4.1	33.65371451
Alfredo sormin	3.2	50.8	1.76	4.1	33.65534876
Anroi Sipahutar	3.4	50.3	1.23	4.5	33.15965169
Enjelika Sipahutar	2.7	4.1	45.01	50.6	13.05038314
Muhammad Pirmansha Pasaribu	3.4	49.8	0.75	5.0	32.65979945
Sartika Goltom	3.2	41.8	7.31	12.9	24.65730115
Kristop Lubis	3.7	50.1	1.00	4.6	32.96835604
Ayu Pane	3.6	45.6	3.50	9.1	28.46756927
Nurhidayah	4.2	53.7	4.63	1.0	36.58500376
Dewi Pakpahan	4.3	50.5	1.54	4.1	33.39330023
Esti Sibarani	4.1	48.7	0.59	5.9	31.58563756
Exael Situmorang	4.2	52.7	3.64	2.0	35.58598741
Uba Wahyu Pakpahan	4.5	54.7	5.66	0.1	37.59803851
Konida Munthe	4.1	52.8	3.73	1.9	35.68154285
Enjel sidabutar	4.6	54.5	5.48	0.1	37.40350919
Wanris Ritonga	4.3	51.6	2.58	3.0	34.49191934
Gomgom Sabar Hasibuan	4.8	56.7	7.68	2.1	39.61114111
Halasson	5.1	56.7	7.73	2.1	39.62893514
Wiwini Pasaribu	4.8	54.2	5.22	0.5	37.11525966
Laura Lara Hasibuan	4.8	55.7	6.70	1.1	38.61272459
Joel Hasibuan	4.9	50.4	1.79	4.2	33.32945394
Rodami Parsaoran Silitonga	4.2	53.7	4.63	1.0	36.58500376
Putra Sudiantono	5.4	55.5	6.63	1.2	38.45208057

Tahap selanjutnya adalah menentukan jarak terpendek data terhadap pusat *cluster* awal yang dipilih secara random sebelumnya. Menentukan jarak terpendek dengan pusat *cluster* data dapat menggunakan rumus $=\text{MIN}(K32, L32, M132)$. Adapun hasil penentuan jarak terpendek terhadap pusat *cluster* yang telah ditabulasikan dapat dilihat pada tabel 10 berikut:

Tabel 10. Tabulasi Binner Jarak Terdekat Data terhadap *Cluster* Baru

NO	NAMA BALITA	C0	C1	C2
1	Natalia Pasaribu	0	0	1
2	Mira Marpaung	1	0	0
3	Andreas Sormin	1	0	0
4	Alfredo sormin	1	0	0
5	Anroi Sipahutar	1	0	0
6	Enjelika Sipahutar		0	1
7	Muhammad Pirmansha Pasaribu	1	0	0
8	Sartika Goltom	1	0	0
9	Kristop Lubis	1	0	0
10	Ayu Pane	1	0	0
11	Nurhidayah	0	1	0
12	Dewi Pakpahan	1	0	0
13	Esti Sibarani	1	0	0
14	Exael Situmorang	0	1	0
15	Uba Wahyu Pakpahan	0	1	0
16	Konida Munthe	0	1	0
17	Enjel sidabutar	0	1	0
18	Wanris Ritonga	1		0
19	Gomgom Sabar Hasibuan	0	1	0
20	Halasson	0	1	0
21	Wiwini Pasaribu	0	1	0
22	Laura Lara Hasibuan	0	1	0
23	Joel Hasibuan	1	0	0

NO	NAMA BALITA	C0	C1	C2
24	Rodami Parsaoran Silitonga	0	1	0
25	Putra Sudiantono	0	1	0

Hasil tersebut di atas dibandingkan dengan hasil iterasi ke-2 sebelumnya dan hasilnya adalah sama. Maka oleh karena itu, pencarian iterasi dihentikan dan hasil telah ditemukan. Dimana hasil dari pada pengelompokan dapat dilihat pada table 10, yaitu C0 adalah kelompok Gizi baik, dan C1 gizi sedang dan C2 gizi buruk.

5. KESIMPULAN

Berdasarkan analisa dan pembahasan maka penulis dapat menarik kesimpulan sebagai berikut :

1. Pembagian pengelompokan data gizi balita belum efektif dan efisien, dikarenakan belum adanya metode khusus untuk mendukung keberhasilan pembagian pengelompokan data gizi balita, dan masih dilakukan secara manual.
2. Melakukan pembagian kategori berdasarkan data-data berat badan dan tinggi badan, kemudian akan dikategorikan menjadi gizi baik, sedang, dan buruk dengan nilai yang didapat pada setiap cluster.
3. Mengelompokkan pola gizi balita dengan *Algoritma k-means* memang sangat rumit dan dibutuhkan ketelitian yang sangat tinggi dalam penyesuaiannya dengan menggunakan *Software Rapidminer*.

DAFTAR PUSTAKA

- [1] "UPT PUSKESMAS GAROGA." [Online]. Available: <http://garoganews.blogspot.com/2017/09/upt-puskesmas-garoga.html>. [Accessed: 05-Feb-2020].
- [2] C. Purnamaningsih, R. Saptono, and A. Aziz, "Pemanfaatan Metode K-Means Clustering dalam Penentuan Penjurusan Siswa SMA," *J. Teknol. Inf. ITSmart*, vol. 3, no. 1, p. 27, 2016.
- [3] "Puskesmas Garoga Serius Meningkatkan Kesehatan Warga – Borsak Mengatasi Nababan." [Online]. Available: <https://nababan.wordpress.com/2015/08/21/puskesmas-garoga-serius-meningkatkan-kesehatan-warga/>. [Accessed: 05-Feb-2020].
- [4] P. D. . Silitonga, "Clustering of Patient Disease Data by Using K-Means Clustering," *Int. J. Comput. Sci. Inf. Secur. IJCSIS*, vol. 15, no. 7, pp. 219–233, 2017.
- [5] W. Nengsih, "Clustering K-Means Analysis (Studi Kasus : Koleksi Perpustakaan)," *J. Aksara Komput. Terap.*, vol. 6, no. 2, 2017.
- [6] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019.
- [7] M. Taufik, "Makalah Gizi Balita." [Online]. Available: https://www.academia.edu/31979678/Makalah_Gizi_Balita. [Accessed: 05-Feb-2020].
- [8] Y. Pasambo, "Gambaran Status Gizi Balita Di Rt 03 / Rw 09 Kelurahan Barombong Kecamatan Tamalate Kota Makassar," *Media Kesehat. Politek. Kesehat. Makassar*, vol. 13, no. 1, p. 1, 2018.