

PENERAPAN METODE K-MEANS CLUSTERING UNTUK MENENTUKAN STATUS GIZI BAIK DAN GIZI BURUK PADA BALITA (STUDI KASUS KABUPATEN ROKAN HULU)

¹⁾Dona, ²⁾Mi'rajul Rifqi

^{1,2)} Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pasir Pengaraian
Jl. Tuanku Tambusai, Kumu Rambah Hilir
E-mail: dona201804@gmail.com, mirajulrifqi@gmail.com

ABSTRAK

Kekurangan gizi atau yang biasa disebut malnutrisi merupakan salah satu masalah kesehatan yang cukup sering menimpa balita-balita di Indonesia. Kepedulian orang tua dan aparat desa (dalam hal ini petugas Pusat Pelayanan Kesehatan Masyarakat – PUSKESMAS) untuk memantau gizi balita sangat diperlukan. Penelitian yang dilakukan mencoba untuk melakukan pengelompokan 15 balita di Kab. Rokan Hulu kedalam 2 *cluster* status gizi. Pengelompokan status gizi balita di Kab. Rokan Hulu menggunakan metode *K-Means* dilakukan melalui beberapa tahapan yaitu : penentuan tujuan bisnis, pengumpulan data 15 balita di Kab. Rokan Hulu, pengelompokan status gizi balita ke dalam 2 *cluster* yaitu *cluster* 1 - gizi baik; *cluster* 2 -Gizi Baik, pengelompokan status gizi balita menggunakan algoritma *K-Means*, dan yang terakhir melakukan pengujian dengan membandingkan hasil pengelompokan algoritma *K-means* dan *Rapid Miner*.

Kata kunci : Data Mining , Clustering , K -Means

ABSTRACT

Malnutrition is one of the health problems that quite often afflicts toddlers in Indonesia. The care of parents and village officials (in this case the Community Health Service Center staff - PUSKESMAS) to monitor the nutrition of children under five is very necessary. The research was conducted to try to group 15 toddlers in Kab. Rokan Hulu into 2 clusters of nutritional status. Grouping the nutritional status of children under five in Kab. Rokan Hulu using the K-Means method is carried out through several stages, namely: determining business goals, collecting data for 15 toddlers in Kab. Rokan Hulu, grouping the nutritional status of children under five into 2 clusters, namely cluster 1 - good nutrition; cluster 2 -Good nutrition, grouping the nutritional status of children under five using the K-Means algorithm, and the last one doing a test by comparing the results of the grouping of the K-means and Rapid Miner algorithms.

Keywords: Data Mining, Clustering, K-Means

PENDAHULUAN

Kekurangan gizi atau yang biasa disebut malnutrisi merupakan salah satu masalah kesehatan yang cukup sering menimpa balita-balita di Indonesia.

Malnutrisi pada balita di Indonesia disebabkan oleh beberapa faktor seperti, Konsumsi makanan yang diberikan kepada balita. Banyak orang tua yang tidak mengerti mengenai kandungan gizi makanan yang diberikan kepada balitanya menjadi salah satu faktor yang cukup dominan menjadi penyebab malnutrisi pada balita. Makanan yang bergizi tidak selalu harus mahal. Orang tua hanya harus pandai memilih jenis makanan yang bisa mencukup nilai gizi

balitanya. Pengetahuan tentang gizi makanan inilah yang terkadang tidak dimiliki oleh banyak orang tua di Indonesia. Pendidikan yang rendah dan kemiskinan kerap menjadi alasan orang tua kurang bisa memperhatikan asupan gizi makanan yang dikonsumsi oleh balitanya, Lingkungan yang tidak sehat. Rendahnya kepedulian masyarakat pada kebersihan lingkungan di beberapa kota di Indonesia membuat banyak masyarakat terutama balita rentan terinfeksi berbagai macam penyakit. Balita yang mudah terserang penyakit cenderung memiliki gizi yang kurang dibandingkan dengan balita yang jarang menderita sakit.

Gizi kurang atau malnutrisi pada balita membawa dampak negatif terhadap pertumbuhan fisik maupun mental, yang selanjutnya akan menghambat beberapa proses belajar yang dilakukan oleh balita seperti belajar berbicara, berjalan, makan dan lain-lain. Kecerdasan Intelektual (IQ) balita penderita malnutrisi cenderung lebih rendah dibandingkan balita yang sehat. Hal ini disebabkan karena kurang terpenuhinya gizi pada anak akan menghambat sintesis protein DNA sehingga menyebabkan terhambatnya pembentukan sel otak yang selanjutnya akan menghambat perkembangan otak. Jika hal ini terjadi setelah masa divisi sel otak terhenti, hambatan sintesis protein akan menghasilkan otak dengan jumlah sel yang normal tetapi dengan ukuran yang lebih kecil. Akibat lain yang disebabkan oleh malnutrisi adalah penurunan daya tahan tubuh. Balita yang memiliki sistem imun yang rendah akan mudah terinfeksi penyakit dibandingkan balita yang memiliki sistem imun tinggi. Diare dan ISPA (Infeksi Saluran Pernapasan) dan Tuberculosis merupakan tiga dari beberapa penyakit yang sering diderita oleh balita . Kekurangan gizi juga dapat menyebabkan stunting pada balita. Stunting (tubuh pendek) merupakan suatu kondisi terlambatnya pertumbuhan anak yang ditandai dengan tinggi badan anak lebih pendek dibandingkan dengan tinggi badan anak-anak lain di usia yang sama. Dampak yang lebih parah dari malnutrisi pada balita adalah timbulnya kecacatan, tingginya angka kesakitan dan percepatan kematian, Kepedulian orang tua dan aparat desa (dalam hal ini petugas Pusat Pelayanan Kesehatan Masyarakat – PUSKESMAS) untuk memantau gizi balita sangat diperlukan. Malnutrisi pada balita tidak terjadi secara tiba-tiba seperti penyakit pada umumnya. Tanda-tanda seperti berat badan kurang dari

standar, stunting bisa menjadi indikator awal terjadinya malnutrisi pada balita.

METODE

Algoritma *K-Means Clustering*

K-Means clustering adalah teknik pengelompokan data yang mana keberadaan tiap-tiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaan. Teknik ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981.[4]*K-Means* merupakan algoritma *clustering* yang berulang-ulang. Algoritma dimulai dengan pemilihan secara acak K , K disini merupakan banyaknya *cluster* yang ingin dibentuk. Kemudian menetapkan nilai secara random yang disebut dengan *centroid*.

Pada penelitian terdahulu yang dilakukan oleh Elly Muningsih dan Sri Kiswati (2015), metode *K-Means* merupakan salah satu metode dalam fungsi *clustering* atau pengelompokan. *Clustering* mengacu pada pengelompokan data, observasi atau kasus berdasarkan kemiripan objek yang diteliti.[5]

Metodologi Dasar Analisa *K-Means Clustering*

K-Means merupakan salah satu metode dalam fungsi *clustering* atau pengelompokan. *Clustering* mengacu pada pengelompokan data, observasi atau kasus berdasar kemiripan objek yang diteliti. Sebuah *cluster* adalah suatu kumpulan data yang mirip dengan lainnya atau ketidakmiripan data pada kelompok lain. *Clustering* diartikan dengan membagi objek data (bentuk, entitas, contoh, ketaatan, unit) ke dalam beberapa jumlah kelompok (grup, bagian atau kategori). Sedangkan tujuan proses *clustering* dijelaskan oleh yaitu untuk meminimalkan terjadinya *objective function* yang diset dalam proses *clustering*. [6]

Algoritma dasar dalam *K-mean* adalah: [7]

- a. Tentukan nilai k sebagai jumlah *cluster* yang ingin dibentuk.
- b. Bangkitkan k *centroid* (titik pusat cluster) awal secara random.
- c. Hitung jarak setiap data ke masing-masing *centroid* menggunakan rumus korelasi antar dua objek yaitu *Euclidean Distance*

$$De = \sqrt{(M_{ix} - C_{ix})^2 + (M_{iy} - C_{iy})^2}$$

- d. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan *centroid* nya.
- e. Tentukan posisi *centroid* baru (C_k) dengan cara menghitung nilai rata-rata dari data-data yang ada pada *centroid* yang sama.

$$C_k = \left(\frac{1}{n_k}\right) \sum d_i$$

Dimana n_k adalah jumlah dokumen dalam *cluster* k dan d_i adalah dokumen dalam *cluster* k.

- f. Kembali ke langkah 3 jika posisi *centroid* baru dengan *centroid* lama tidak sama. Normalisasi Data

Suatu teknik untuk mengorganisasikan data ke dalam tabel-tabel untuk

Melakukan pengamatan terhadap data yang diteliti, dengan menggunakan alat bantu komputer terutama pada kualitas gizi buruk yang ada pada puskesmas di kabupaten Rokan Hulu provinsi Riau.

1. Analisa Algoritma *K-Means Clustering*
 Pada proses ini penulis memulainya dengan melakukan analisis dan pembelajaran terhadap algoritma-algoritma yang ada pada data mining terkhusus pada algoritma *K-Means Clustering*. Perancangan Pada tahap ini penulis mulai membuat rancangan-rancangan dari aplikasi yang akan dibuat, antara lain desain dari tampilan,

memenuhi kebutuhan pemakai di dalam suatu organisasi . Data-data yang ada dilakukan normalisasi dengan membagi nilai data tersebut dengan nilai *range* data (nilai data maksimum-nilai data minimum). [8] Tujuan dari normalisasi yaitu untuk menghilangkan kerangkapan data, untuk mengurangi kompleksitas, dan untuk mempermudah pemodifikasian. [9] berikut ini merupakan salah satu rumus normalisasi yang akan digunakan dalam normalisasi data algoritma *K-mean*.

Nilai Normalisasi

$$= \frac{(\text{Nilai Awal} - \text{Nilai Minimal})}{(\text{Nilai Maksimal} - \text{Nilai Minimal})}$$

Keterangan:

Nilai normalisasi = nilai data normal

Nilai awal = nilai data awal atau nilai data ke-1

Nilai minimal = nilai minimum data keseluruhan

Nilai maximum = nilai maximum data keseluruhan

KERANGKA KERJA PENELITIAN

perancangan menu-menu, serta penentuan dari fungsi setiap tombol yang ada.

Pada tahap ini penulis mulai menuangkan konsep-konsep yang ada kedalam bentuk

:



2. Implementasi

Pada tahap ini, implementasi dibuat berdasarkan hasil analisa dalam perancangan sistem yang dikembangkan menggunakan penerapan aplikasi rapidminer.

HASIL

Pada tabel di bawah ini merupakan sampel data balita yang digunakan untuk melakukan percobaan perhitungan manual.

Tabel 1. Data Balita

No	Balita ke-	Tinggi Badan (Cm)	Berat Badan (Kg)
1	Balita 1	72,5	7
2	Balita 2	71,5	8,7
3	Balita 3	52	6,5
4	Balita 4	75	10,1
5	Balita 5	99	7,8
6	Balita 6	75	8
7	Balita 7	72	10,2
8	Balita 8	50	6
9	Balita 9	82	9,7
10	Balita 10	95	12
11	Balita 11	88	9,4
12	Balita 12	70	4,2
13	Balita 13	51	5
14	Balita 14	60	8
15	Balita 15	54	3,5

Data yang ada di tabel 4.1 tidak dapat langsung dilakukan pemrosesan dikarenakan terdapat besaran angka yang cukup jauh antara variabel tinggi badan dan berat badan. Perbedaan jarak atau besaran angka yang cukup jauh ini dapat menyulitkan dalam proses pengelompokan. Salah satu solusi yang digunakan untuk memperkecil besaran angka antar variabel adalah dengan melakukan normalisasi angka-angka yang ada di variabel tinggi badan dan berat badan menggunakan rumus:

Nilai Normalisasi

$$= \frac{(\text{Nilai Awal} - \text{Nilai Minimal})}{(\text{Nilai Maksimal} - \text{Nilai Minimal})}$$

Nilai variabel tinggi badan dan berat badan akan dinormalisasi ke dalam rentang 0 – 1. Normalisasi angka pada tiap variabel ini sangat dibutuhkan sebelum proses perhitungan nilai centroid oleh algoritma *K-Means* agar tidak ada parameter yang mendominasi dalam perhitungan jarak antar data. Adapun tahapan yang dilakukan untuk proses normalisasi adalah :

- Mencari nilai maksimum dan minimum untuk variabel tinggi badan (X)

$$X_{11} = (X_{\text{balita1}} - X_{\text{min}}) / (X_{\text{maks}} - X_{\text{min}})$$

$$= (72,5 - 50) / (99 - 50) = 0,459$$

$$X_{12} = (X_{\text{balita2}} - X_{\text{min}}) / (X_{\text{maks}} - X_{\text{min}})$$

$$= (71,5 - 50) / (99 - 50) = 0,439$$

$$X_{13} = (X_{\text{balita3}} - X_{\text{min}}) / (X_{\text{maks}} - X_{\text{min}})$$

$$= (52 - 50) / (99 - 50) = 0,041$$

$$X_{14} = (X_{\text{balita4}} - X_{\text{min}}) / (X_{\text{maks}} - X_{\text{min}})$$

$$= (75 - 50) / (99 - 50) = 0,510$$

$$X_{15} = (X_{\text{balita5}} - X_{\text{min}}) / (X_{\text{maks}} - X_{\text{min}})$$

$$= (99 - 50) / (99 - 50) = 1$$

$$X_{16} = (X_{\text{balita6}} - X_{\text{min}}) / (X_{\text{maks}} - X_{\text{min}})$$

$$= (75 - 50) / (99 - 50) = 0,510$$

Perhitungan yang sama dilakukan hingga balita ke-15 dan juga digunakan untuk melakukan normalisasi variabel berat badan. Hasil dari normalisasi pada variabel tinggi badan dan berat badan dapat dilihat pada tabel 2.

Tabel 2. Data Normalisasi Variabel

No	Balita	Tinggi	Berat
----	--------	--------	-------

ke-	Badan (Cm)	Badan (Kg)
1	Balita 1	0,459
2	Balita 2	0,439
3	Balita 3	0,041
4	Balita 4	0,510
5	Balita 5	1,000
6	Balita 6	0,510
7	Balita 7	0,449
8	Balita 8	0,000
9	Balita 9	0,653
10	Balita 10	0,918
11	Balita 11	0,776
12	Balita 12	0,408
13	Balita 13	0,020
14	Balita 14	0,204
15	Balita 15	0,082

Setelah angka pada masing-masing variabel dilakukan normalisasi, langkah selanjutnya adalah menentukan jumlah kelompok atau *cluster*. Ke 15 data balita yang ada di tabel 4.2 akan dikelompokkan ke dalam 2 *cluster* yaitu : Gizi buruk dan Gizi Baik.

Setelah jumlah *cluster* ditentukan, langkah selanjutnya adalah melakukan menentukan nilai initial *cluster centre* untuk masing-masing cluster pada setiap variabelnya. Nilai initial *cluster centre* pada iterasi yang pertama (perhitungan pertama kali) diberikan secara acak. Pada iterasi selanjutnya, nilai initial *cluster centre* (pengulangan ke-1 sampai dengan posisi normal/maksimal iterasi) diberikan dengan menghitung nilai rata-rata data pada setiap *clusternya*. Jika nilai initial *cluster centre* yang baru sama dengan nilai initial *cluster centre* yang baru maka proses iterasi dilanjutkan hingga nilai sama atau sampai dengan nilai maksimal iterasi yang telah ditetapkan sebelumnya (misalnya 100). Namun jika nilai initial *cluster centre* yang baru sama dengan initial cluster centre yang lama, proses pengelompokan akan berhenti. Berikut langkah-langkah yang akan dilakukan:

1. Tentukan initial *cluster centre* “Centroid”

Untuk penentuan pusat awal diambil dari nilai tabel normalisasi

- Pusat *cluster* ke-1 : { 0,449 | 1 }
- Pusat *cluster* ke-2 : { 0,204 | 0,672 }

2. Perhitungan jarak pusat *cluster*
 Untuk mengukur jarak antara data dengan pusat *cluster* digunakan *Euclidian Distance*, kemudian akan didapatkan matrik jarak sebagai berikut :

$$De = \sqrt{(M_{ix} - C_{ix})^2 + (M_{iy} - C_{iy})^2}$$

Keterangan:

De= jarak antara data dengan pusat *cluster*

M_{ix}=nilai variabel x

C_{ix}=nilai *cluster* variabel x

M_{iy}=nilai variabel y

C_{iy}=nilai *cluster* untuk variabel y

Iterasi Ke-1

1. Perhitungan jarak dari data terhadap pusat *cluster* ke-1

$$D1 = \sqrt{(0,459 - 0,449)^2 + (0,522 - 1)^2} = 0,478$$

$$D2 = \sqrt{(0,439 - 0,449)^2 + (0,776 - 1)^2} = 0,224$$

$$D3 = \sqrt{(0,041 - 0,449)^2 + (0,448 - 1)^2} = 0,687$$

$$D4 = \sqrt{(0,510 - 0,449)^2 + (0,985 - 1)^2} = 0,063$$

$$D5 = \sqrt{(1 - 0,449)^2 + (0,642 - 1)^2} = 0,657$$

$$D6 = \sqrt{(0,510 - 0,449)^2 + (0,672 - 1)^2} = 0,334$$

$$D7 = \sqrt{(0,449 - 0,449)^2 + (1 - 1)^2} = 0$$

$$D8 = \sqrt{(0 - 0,449)^2 + (0,373 - 1)^2} = 0,771$$

$$D9 = \sqrt{(0,653 - 0,449)^2 + (0,925 - 1)^2} = 0,217$$

$$D10 = \sqrt{(0,918 - 0,449)^2 + (1,269 - 1)^2} = 0,541$$

$$D11 = \sqrt{(0,776 - 0,449)^2 + (0,881 - 1)^2} = 0,348$$

$$D12 = \sqrt{(0,408 - 0,449)^2 + (0,104 - 1)^2} = 0,896$$

$$D13 = \sqrt{(0,020 - 0,449)^2 + (0,224 - 1)^2} = 0,887$$

$$D14 = \sqrt{(0,204 - 0,449)^2 + (0,672 - 1)^2} = 0,410$$

$$D15 = \sqrt{(0,082 - 0,449)^2 + (0 - 1)^2} = 1,065$$

2. Perhitungan jarak dari data terhadap pusat *cluster* ke-2

$$D21 = \sqrt{(0,459 - 0,204)^2 + (0,522 - 0,672)^2} = 0,296$$

$$D22 = \sqrt{(0,439 - 0,204)^2 + (0,776 - 0,672)^2} = 0,257$$

$$D23 = \sqrt{(0,041 - 0,204)^2 + (0,448 - 0,672)^2} = 0,277$$

$$D24 = \sqrt{(0,510 - 0,204)^2 + (0,985 - 0,672)^2} = 0,438$$

$$D25 = \sqrt{(1 - 0,204)^2 + (0,642 - 0,672)^2} = 0,797$$

$$D26 = \sqrt{(0,510 - 0,204)^2 + (0,672 - 0,672)^2} = 0,306$$

$$D27 = \sqrt{(0,449 - 0,204)^2 + (1 - 0,672)^2} = 0,409$$

$$D28 = \sqrt{(0 - 0,204)^2 + (0,373 - 0,672)^2} = 0,362$$

$$D29 = \sqrt{(0,653 - 0,204)^2 + (0,925 - 0,672)^2} = 0,516$$

$$D210 = \sqrt{(0,918 - 0,204)^2 + (1,269 - 0,672)^2} = 0,931$$

$$D211 = \sqrt{(0,776 - 0,204)^2 + (0,881 - 0,672)^2} = 0,608$$

$$D212 = \sqrt{(0,408 - 0,204)^2 + (0,104 - 0,672)^2} = 0,603$$

$$D213 = \sqrt{(0,020 - 0,204)^2 + (0,224 - 0,672)^2} = 0,484$$

$$D214 = \sqrt{(0,204 - 0,204)^2 + (0,672 - 0,672)^2} = 0$$

$$D215 = \sqrt{(0,082 - 0,204)^2 + (0 - 0,672)^2} = 0,683$$

Pada iterasi pertama dimulai dengan menentukan nilai *centroid* awal dengan mengambil nilai secara acak pada tabel normalisasi pada perhitungan di atas adalah pada data ke-7 dan data ke-14 setelah diperoleh pusat *cluster* 1 dan 2, selanjutnya dilakukan perhitungan jarak dari data terhadap pusat *cluster* 1 dan *cluster* 2 dengan menggunakan *Euclidian Distance*, maka diperoleh hasil untuk nilai tertinggi diinitalkan nilai *clusternya*

adalah= 0, sedangkan untuk nilai *cluster* terendah di inisialkan adalah=1, sehingga diperoleh hasil seperti tabel 3.

Tabel 3. Hasil Perhitungan Iterasi 1

No	Balita ke-	Nilai C1	Nilai C2	C1	C2
1	Balita 1	0.478	0.296	0	1
2	Balita 2	0.224	0.257	1	0
3	Balita 3	0.687	0.277	0	1
4	Balita 4	0.063	0.438	1	0
5	Balita 5	0.657	0.797	1	0
6	Balita 6	0.334	0.306	0	1
7	Balita 7	0.000	0.409	1	0
8	Balita 8	0.771	0.362	0	1
9	Balita 9	0.217	0.516	1	0
10	Balita 10	0.541	0.931	1	0
11	Balita 11	0.348	0.608	1	0
12	Balita 12	0.896	0.603	0	1
13	Balita 13	0.887	0.484	0	1
14	Balita 14	0.410	0.000	0	1
15	Balita 15	1.065	0.683	0	1

Pada pencarian iterasi pertama didapatkan hasil sebagai berikut :

{ Balita 2, Balita 4, Balita 5, Balita 7, Balita 9, Balita 10, Balita 11 } = 7 Anggota C1 (Gizi Buruk)

{ Balita 1, Balita 3, Balita 6, Balita 8, Balita 12, Balita 13, Balita 14, Balita 15 }

= 8 Anggota C2 (Gizi

Baik)

Iterasi Ke-2

Hitung Titik Pusat Baru

Menentukan posisi *centroid* baru (C_k) dengan cara menghitung nilai rata-rata dari data yang ada pada setiap *centroid* yang sama.

Maka didapatkan nilai *centroid* 2 baru yakni={0,377 | 0,216 }

1. Perhitungan jarak dari data terhadap pusat *cluster* ke-1 baru

D_1

$$= \sqrt{(0,459 - 0,678)^2 + (0,522 - 0,925)^2}$$

$$= 0,458$$

D_2

$$= \sqrt{(0,439 - 0,678)^2 + (0,776 - 0,925)^2}$$

$$= 0,282$$

D_3

$$= \sqrt{(0,041 - 0,678)^2 + (0,448 - 0,925)^2}$$

$$= 0,796$$

D_4

$$= \sqrt{(0,510 - 0,678)^2 + (0,985 - 0,925)^2}$$

$$= 0,178$$

D_5

$$= \sqrt{(1 - 0,678)^2 + (0,642 - 0,925)^2}$$

$$= 0,429$$

D_6

$$= \sqrt{(0,510 - 0,678)^2 + (0,672 - 0,925)^2}$$

$$= 0,304$$

D_7

$$= \sqrt{(0,449 - 0,678)^2 + (1 - 0,925)^2}$$

$$= 0,241$$

D_8

$$= \sqrt{(0 - 0,678)^2 + (0,373 - 0,925)^2}$$

$$= 0,874$$

D_9

$$= \sqrt{(0,653 - 0,678)^2 + (0,925 - 0,925)^2}$$

$$= 0,025$$

D_{10}

$$= \sqrt{(0,918 - 0,678)^2 + (1,269 - 0,925)^2}$$

$$= 0,419$$

$$D11 = \sqrt{(0,776 - 0,678)^2 + (0,881 - 0,925)^2} = 0,107$$

$$D12 = \sqrt{(0,408 - 0,678)^2 + (0,104 - 0,925)^2} = 0,864$$

$$D13 = \sqrt{(0,020 - 0,678)^2 + (0,224 - 0,925)^2} = 0,961$$

$$D14 = \sqrt{(0,204 - 0,678)^2 + (0,672 - 0,925)^2} = 0,537$$

$$D15 = \sqrt{(0,082 - 0,678)^2 + (0 - 0,925)^2} = 1,101$$

3. Perhitungan jarak dari data terhadap pusat *cluster* ke-2

$$D21 = \sqrt{(0,459 - 0,377)^2 + (0,522 - 0,216)^2} = 0,317$$

$$D22 = \sqrt{(0,439 - 0,377)^2 + (0,776 - 0,216)^2} = 0,564$$

$$D23 = \sqrt{(0,041 - 0,377)^2 + (0,448 - 0,216)^2} = 0,408$$

$$D24 = \sqrt{(0,510 - 0,377)^2 + (0,985 - 0,216)^2} = 0,781$$

$$D25 = \sqrt{(1 - 0,377)^2 + (0,642 - 0,216)^2} = 0,755$$

$$D26 = \sqrt{(0,510 - 0,377)^2 + (0,672 - 0,216)^2} = 0,475$$

$$D27 = \sqrt{(0,449 - 0,377)^2 + (1 - 0,216)^2} = 0,787$$

$$D28 = \sqrt{(0 - 0,377)^2 + (0,373 - 0,216)^2} = 0,408$$

$$D29 = \sqrt{(0,653 - 0,377)^2 + (0,925 - 0,216)^2} = 0,761$$

$$D210 = \sqrt{(0,918 - 0,377)^2 + (1,269 - 0,216)^2} = 1,184$$

$$D211 = \sqrt{(0,776 - 0,377)^2 + (0,881 - 0,216)^2} = 0,775$$

$$D212 = \sqrt{(0,408 - 0,377)^2 + (0,104 - 0,216)^2} = 0,116$$

$$D213 = \sqrt{(0,020 - 0,377)^2 + (0,224 - 0,216)^2} = 0,357$$

$$D214 = \sqrt{(0,204 - 0,377)^2 + (0,672 - 0,216)^2} = 0,487$$

$$D215 = \sqrt{(0,082 - 0,377)^2 + (0 - 0,216)^2} = 0,366$$

Iterasi ke 2 dilakukan karena jumlah anggota C1 dan C2 tidak sama pada iterasi ke-1, sehingga dilakukan proses perhitungan dimulai dengan menentukan posisi *centroid* baru dengan menghitung nilai rata-rata dari data yang ada pada setiap *centroid* yang sama, nilai yang diambil adalah dari tabel normalisasi. Setelah didapat nilai *centroid* baru selanjutnya melakukan perhitungan jarak dari data terhadap pusat *cluster* 1 dan 2 dengan menggunakan *Euclidian Distance* sehingga diperoleh hasil seperti tabel 4.4 untuk nilai *cluster* tertinggi pada *cluster* diinitalkan = 0, sementara terendah diinitalkan=1.

Tabel 4. Hasil Perhitungan Iterasi 2

No	Balita ke-	Nilai C1	Nilai C2	C1	C2
1	Balita 1	0.458	0.317	0	1

2	Balita 2	0.282	0.564	1	0
3	Balita 3	0.796	0.408	0	1
4	Balita 4	0.178	0.781	1	0
5	Balita 5	0.429	0.755	1	0
6	Balita 6	0.304	0.475	1	0
7	Balita 7	0.241	0.787	1	0
8	Balita 8	0.874	0.408	0	1
9	Balita 9	0.025	0.761	1	0
10	Balita 10	0.419	1.184	1	0
11	Balita 11	0.107	0.775	1	0
12	Balita 12	0.864	0.116	0	1
13	Balita 13	0.961	0.357	0	1
14	Balita 14	0.537	0.487	0	1
15	Balita 15	1.101	0.366	0	1

Pada pencarian iterasi kedua didapatkan hasil sebagai berikut :

{ Balita 2, Balita 4, Balita 5, Balita 6, Balita 7, Balita 9, Balita 10, Balita 11 } = 8 Anggota C1 (Gizi Buruk)

{ Balita 1, Balita 3, Balita 8, Balita 12, Balita 13, Balita 14, Balita 15 } = 7 Anggota C2 (Gizi Baik)

Karena hasil anggota iterasi – 1 dan iterasi – 2 berbeda maka akan dihitung nilai *centroid* baru.

Iterasi-3

1. Hitung Titik Pusat Baru

Mencari nilai *centroid* 1 baru :

Maka didapatkan nilai *centroid* 1 baru yakni = {0,657 | 0,95 }

Maka didapatkan nilai *centroid* 2 baru yakni = {0,335 | 0,173 }

2. Perhitungan jarak dari data terhadap pusat *cluster* ke-1 baru

D1

$$= \sqrt{(0,459 - 0,657)^2 + (0,522 - 0,95)^2}$$

$$= 0,471$$

D2

$$= \sqrt{(0,439 - 0,657)^2 + (0,776 - 0,95)^2}$$

$$= 0,279$$

D3

$$= \sqrt{(0,041 - 0,657)^2 + (0,448 - 0,95)^2}$$

$$= 0,795$$

D4

$$= \sqrt{(0,510 - 0,657)^2 + (0,985 - 0,95)^2}$$

$$= 0,151$$

$$D5 = \sqrt{(1 - 0,657)^2 + (0,642 - 0,95)^2}$$

$$= 0,461$$

D6

$$= \sqrt{(0,510 - 0,657)^2 + (0,672 - 0,95)^2}$$

$$= 0,315$$

$$D7 = \sqrt{(0,449 - 0,657)^2 + (1 - 0,95)^2}$$

$$= 0,214$$

$$D8 = \sqrt{(0 - 0,657)^2 + (0,373 - 0,95)^2}$$

$$= 0,874$$

D9

$$= \sqrt{(0,653 - 0,657)^2 + (0,925 - 0,95)^2}$$

$$= 0,025$$

D10

$$= \sqrt{(0,918 - 0,657)^2 + (1,269 - 0,95)^2}$$

$$= 0,412$$

D11

$$= \sqrt{(0,776 - 0,657)^2 + (0,881 - 0,95)^2}$$

$$= 0,137$$

D12

$$= \sqrt{(0,408 - 0,657)^2 + (0,104 - 0,95)^2}$$

$$= 0,881$$

D13

$$= \sqrt{(0,020 - 0,657)^2 + (0,224 - 0,95)^2}$$

$$= 0,966$$

D14

$$= \sqrt{(0,204 - 0,657)^2 + (0,672 - 0,95)^2}$$

$$= 0,532$$

D15

$$= \sqrt{(0,082 - 0,657)^2 + (0 - 0,95)^2}$$

$$= 1,111$$

4. Perhitungan jarak dari data terhadap pusat *cluster* ke-2

D21

$$= \sqrt{(0,459 - 0,335)^2 + (0,522 - 0,173)^2}$$

$$= 0,371$$

D22

$$= \sqrt{(0,439 - 0,335)^2 + (0,776 - 0,173)^2}$$

$$= 0,612$$

D23

$$= \sqrt{(0,041 - 0,335)^2 + (0,448 - 0,173)^2}$$

$$= 0,403$$

D24

$$= \sqrt{(0,510 - 0,335)^2 + (0,985 - 0,173)^2}$$

$$= 0,831$$

D25

$$= \sqrt{(1 - 0,335)^2 + (0,642 - 0,173)^2}$$

$$= 0,814$$

D26

$$= \sqrt{(0,510 - 0,335)^2 + (0,672 - 0,173)^2}$$

$$= 0,529$$

D27 =

$$\sqrt{(0,449 - 0,335)^2 + (1 - 0,173)^2} =$$

$$0,835$$

D28

$$= \sqrt{(0 - 0,335)^2 + (0,373 - 0,173)^2}$$

$$= 0,390$$

D29

$$= \sqrt{(0,653 - 0,335)^2 + (0,925 - 0,173)^2}$$

$$= 0,817$$

D210

$$= \sqrt{(0,918 - 0,335)^2 + (1,269 - 0,173)^2}$$

$$= 1,241$$

D211

$$= \sqrt{(0,776 - 0,335)^2 + (0,881 - 0,173)^2}$$

$$= 0,834$$

D212

$$= \sqrt{(0,408 - 0,335)^2 + (0,104 - 0,173)^2}$$

$$= 0,100$$

D213

$$= \sqrt{(0,020 - 0,335)^2 + (0,224 - 0,173)^2}$$

$$= 0,319$$

D214

$$= \sqrt{(0,204 - 0,335)^2 + (0,672 - 0,173)^2}$$

$$= 0,516$$

D215

$$= \sqrt{(0,082 - 0,335)^2 + (0 - 0,173)^2}$$

$$= 0,307$$

Iterasi ke 3 dilakukan Karena hasil anggota iterasi – 1 dan iterasi – 2 berbeda, sehingga dilakukan proses perhitungan dimulai dengan menentukan posisi *centroid* baru dengan menghitung nilai rata-rata dari data yang ada pada setiap *centroid* yang sama, nilai yang diambil adalah dari tabel normalisasi. Setelah didapat nilai *centroid* baru selanjutnya melakukan perhitungan jarak dari data terhadap pusat *cluster* 1 dan 2 dengan menggunakan *Euclidian Distance* sehingga diperoleh hasil seperti tabel 4.5 untuk nilai *cluster* tertinggi pada *cluster* diinialkan = 0, sementara terendah diinialkan=1

Tabel 5 Hasil Perhitungan Iterasi 3

No	Balita ke-	Nilai C1	Nilai C2	C1	C2
1	Balita 1	0.471	0.371	0	1
2	Balita 2	0.279	0.612	1	0
3	Balita 3	0.795	0.403	0	1
4	Balita 4	0.151	0.831	1	0
5	Balita 5	0.461	0.814	1	0
6	Balita 6	0.315	0.529	1	0
7	Balita 7	0.214	0.835	1	0
8	Balita 8	0.874	0.390	0	1

9	Balita 9	0.025	0.817	1	0
10	Balita 10	0.412	1.241	1	0
11	Balita 11	0.137	0.834	1	0
12	Balita 12	0.881	0.100	0	1
13	Balita 13	0.966	0.319	0	1
14	Balita 14	0.532	0.516	0	1
15	Balita 15	1.111	0.307	0	1

Pada pencarian iterasi ketiga didapatkan hasil sebagai berikut :

{ Balita 2, Balita 4, Balita 5, Balita 6, Balita 7, Balita 9, Balita 10, Balita 11 } = 8 Anggota C1 (Gizi Buruk)

{ Balita 1, Balita 3, Balita 8, Balita 12, Balita 13, Balita 14, Balita 15 } = 7 Anggota C2 (Gizi Baik)

Karena hasil anggota iterasi – 2 dan iterasi – 3 sama maka dalam perhitungan manual dapat dihentikan.

Hasil Pengujian Clustering

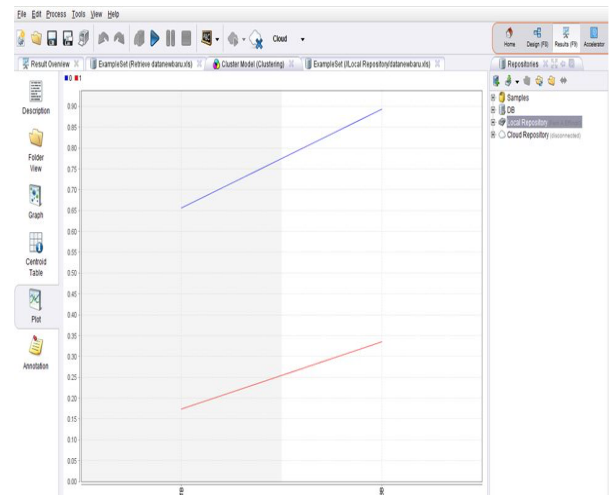
Hasil penentuan gizi buruk dengan metode *Clustering K-Means* dengan menggunakan aplikasi *RapidMiner* dapat dilihat pada gambar 1.

ID	Name	cluster_0	cluster_1	
1	Balita 1	cluster_1	0.439	0.522
2	Balita 2	cluster_0	0.439	0.779
3	Balita 3	cluster_1	0.841	0.448
4	Balita 4	cluster_0	0.553	0.985
5	Balita 5	cluster_0	1	0.542
6	Balita 6	cluster_0	0.515	0.972
7	Balita 7	cluster_0	0.448	1
8	Balita 8	cluster_1	0	0.373
9	Balita 9	cluster_0	0.853	0.825
10	Balita 10	cluster_0	0.918	0.289
11	Balita 11	cluster_0	0.776	0.881
12	Balita 12	cluster_1	0.408	0.154
13	Balita 13	cluster_1	0.303	0.224
14	Balita 14	cluster_1	0.204	0.972
15	Balita 15	cluster_1	0.882	0

Gambar 1. Tampilan Hasil Cluster (ExampleSet Retrieve 15 Data)

ExampleSet (Retrieve 15 Data)

Pada *ExampleSet* dapat dilihat beberapa tampilan hasil *cluster* yaitu berupa *Plot View* yang terlihat pada gambar 5.10.



Gambar 2. Tampilan Hasil Cluster (Plot View)

b. Cluster Model (Clustering)

Pada *Cluster Model (Clustering)* ini dapat dilihat jumlah dari ke-2 *cluster* adalah *cluster* 1 sebanyak 8 *items*. *Cluster* 2 sebanyak 7 *items*. Keseluruhan data berjumlah 15 *items*. Hasil pengelompokan dapat dilihat pada gambar 5.11.

Kesimpulan

Berdasarkan analisa dan pengujian menggunakan *Software RapidMiner 6.5* pada Penderita Gizi Buruk di Kabupaten Rokan Hulu, maka dapat diambil beberapa kesimpulan sebagai berikut :

1. Teknik *Clustering* pada data balita di Rokan Hulu menghasilkan 2 *cluster* yaitu *cluster* Gizi Baik dan *cluster* gizi buruk.
2. Penerapan teknik *Clustering* dengan algoritma *K-Means* yang dilakukan dapat menghasilkan

sebuah pengetahuan mengenai data balita di Kab. Rokan Hulu.

3. Hasil metode Algoritma K-means Clustering dengan menggunakan Aplikasi *RapidMiner 6.5* dapat menghasilkan pengetahuan untuk mengambil keputusan yang lebih optimal.

DAFTAR PUSTAKA

- [1] E. N. Wahyudi, A. Jananto, and Narwati, "Analisa Profil Data Mahasiswa Baru terhadap Program Studi yang dipilih di Perguruan Tinggi Swasta Jawa Tengah dengan Menggunakan Teknik Data Mining," *J. Teknol. Inf. Din.*, vol. 16, no. 1, pp. 29–43, 2011.
- [2] A. Prajana, "Aplikasi Data Mining Untuk Perbandingan Manajemen Laba Terhadap Tingkat Resiko Investasi Pada Perusahaan Perbankan Di Indonesia Stock Exchange," *Int. J. Nat. Sci. Eng.*, vol. 1, no. 1, p. 28, 2017, doi: 10.23887/ijnse.v1i1.12437.
- [3] F. Nasari and S. Darma, "Seminar Nasional Teknologi Informasi dan Multimedia 2015 Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Potensi Utama)," pp. 6–8, 2015.
- [4] D. D. C. Nugraha, Z. Naimah, M. Fahmi, and N. Setiani, "Klasterisasi Judul Buku dengan Menggunakan Metode K-Means," *Semin. Nas. Apl. Teknol. Inf. Yogyakarta*, vol. 21, no. 1, pp. 1907–5022, 2014.
- [5] E. Muningsih and S. Kiswati, "Penerapan Metode K-Means Untuk Clustering Produk Online Shop Dalam Penentuan Stok Barang," *J. Bianglala Inform.*, vol. 3, no. 1, pp. 10–17, 2015.
- [6] B. M. Metisen and H. L. Sari, "Analisis clustering menggunakan metode K-Means dalam pengelompokkan penjualan produk pada Swalayan Fadhila," *J. Media Infotama*, vol. 11, no. 2, pp. 110–118, 2015.
- [7] Asroni and R. Adrian, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," *Semesta Tek.*, vol. 18, no. 1, pp. 76–82, 2016, doi: 10.18196/st.v18i1.708.
- [8] D. Puspitasari, C. Rahmad, and M. Astiningrum, "Normalisasi Tabel Pada Basisdata Relasional," *J. Pros. SENTIA | ISSN 2085-2347*, vol. 8, no. 1, pp. 340–345, 2016.
- [9] G. Triyono, "Pertimbangan Melakukan Denormalisasi Pada Model Basis Data Relasi," *J. Telemat. MKOM | ISSN 2085-725X*, vol. 3, no. 2, pp. 19–25, 2016.