

Pengukuran Keakuratan Metode K-Means untuk Menentukan Status Gizi Balita

Adimas Ketut Nalendra

Universitas Nusantara PGRI Kediri

Jl. KH. Achmad Dahlan No. 76 Mojoroto Kota Kediri

Email : dimasketutuchiha@gmail.com¹⁾

Abstract

Infancy is an optimal growth and development that occurs at the age of 0 to 2 years. According to The Government's Basic Health Research shows that the number of children under five in Indonesia with malnutrition in 2013 reached 28.1 percent of the population. To analyze malnutrition can use the K-Means method with a variable weight and height. For verification, the Growth Chart table is used to measure the accuracy of the K-Means method. Research uses the CRIPS-DM cycle model that is used to handle a data mining project. From the results of the test, the accuracy rate is 34%. Using the K-Means method still has a small level of accuracy so that other methods are needed that have higher accuracy.

Keywords: k-means, growth chart, gizi buruk, CRIPS-DM.

1. Pendahuluan

1.1 Latar Belakang

Masa balita merupakan masa tumbuh kembang yang optimal terjadi pada umur 0 s.d 2 tahun. Pertumbuhan fisik dan mental terjadi pada masa ini. Pertumbuhan fisik meliputi pertumbuhan tinggi badan, berat badan sehingga pada kondisi ini anak bisa terampil dalam menggunakan tangan, jari-jari, kaki, dapat berjalan berbicara, berlari dan dapat makan sendiri. Sedangkan pertumbuhan mental anak memiliki tingkat keingin tahun tinggi, suka meniru berani belajar dll.

Menurut The Government's Basic Health Research (Riskesdas) menunjukkan bahwa jumlah balita di Indonesia usia 12 sampai 59

bulan penderita gizi buruk pada tahun 2013 mencapai 28,1 persen (Post, 2015). Dari data tersebut menunjukkan bahwa masih banyak balita di Indonesia yang menderita gizi buruk. Gizi buruk akan berpengaruh negatif pada masa yang akan datang.

Ada beberapa faktor yang dapat menyebabkan balita gizi buruk. Gizi buruk balita di Indonesia disebabkan oleh beberapa faktor seperti (Rahim, 2014) : 1) 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 gizi buruk pada balita. Makanan yang tidak bergizi tidak selalu harus mahal. Orang tua hanya harus pandai memilih jenis makanan yang bisa mencukupi 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.

2) 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.

Kurang gizi pada balita membawa dampak yang negatif pada pertumbuhan fisik dan mental. Yang dikemudian hari akan menghambat beberapa proses belajar yang dilakukan oleh balita seperti belajar berbicara, berjalan, makan dan lain-lain. Kecerdasan intelektual (IQ) pada balita penderita gizi buruk juga lebih rendah dibandingkan balita yang sehat. Akibat lain balita yang menderita gizi buruk adalah balita memiliki sistem imun yang rendah sehingga akan mudah terinfeksi penyakit dibandingkan dengan balita yang memiliki sistem imun tinggi.

Kepedulian pemerintah dan masyarakat untuk memantau gizi balita sangat diperlukan. Gizi buruk pada balita tidak terjadi secara tiba-tiba seperti penyakit umumnya tetapi memiliki tanda-tanda seperti berat badan kurang dari standart, tinggi badan kurang dari standart. Dari gejala tersebut dapat menjadi indikator awal terjadinya gizi buruk pada balita.

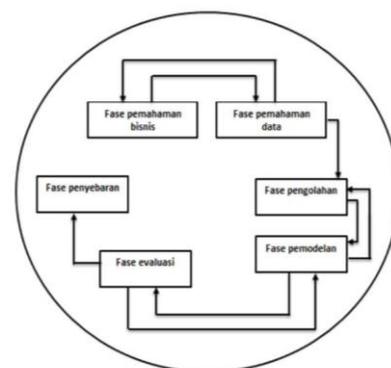
Dari gejala yang terjadi pada balita penderita gizi buruk dapat dilakukan pengelompokan. Kelompok tersebut dikelompokkan berdasarkan karakteristik data yang sama (cluster). Berdasarkan kelompok data yang sama dapat dianalisis

menggunakan metode yang memanfaatkan algoritma data mining yang telah banyak dilakukan pada peneliti sebelumnya. Metode yang banyak digunakan untuk memanfaatkan algoritma data mining adalah metode K-Means. K-Means adalah metode yang termasuk dalam algoritma clustering berbasis jarak yang membagi data ke dalam sejumlah cluster dan algoritma hanya bekerja pada atribut numerik. (Witten, 2012).

Metode K-Means banyak digunakan karena memiliki beberapa kelebihan. Mulai dari mudah untuk diimplementasikan dan dijalankan, waktu yang dibutuhkan untuk menjalankan relatif cepat, mudah di adaptasi dan umum digunakan. Dari beberapa kelebihan tersebut perlu di ukur seberapa akurat metode ini digunakan untuk menentukan status gizi balita.

1.2 Metode Penelitian

Dalam penelitian ini menggunakan model CRIPS-DM (Cross Industry Standart Process for Data Mining). Dalam sebuah proyek data mining memiliki siklus hidup yang terbagi dalam 6 fase yang terlihat pada gambar 1.



Gambar 1. Siklus Model CRIPS-DM

Langkah yang dilakukan dalam penelitian ini adalah :

1. Fase Bisnis (Business Understanding) : Meliputi penentuan tujuan bisnis, menilai situasi saat ini, menetapkan tujuan data mining dan mengembangkan proyek. Tujuan bisnis yang dilakukan dalam penelitian ini adalah pengelompokkan nilai gizi balita dengan menggunakan metode K-Means lalu di ukur tingkat keakuratannya.
2. Fase Pemahaman Data (Data Understanding Phase) : Setelah tujuan bisnis dan rencana proyek ditetapkan, langkah selanjutnya melakukan pengumpulan data awal, deskripsi data, eksplorasi data dan verifikasi kualitas data. Penelitian ini menggunakan data primer dengan respondennya adalah balita dengan umur 0 – 36 bulan sesuai data posyandu di Desa Badas Kecamatan Badas Kabupaten Kediri. Perhitungan yang akan digunakan untuk menentukan status gizi balita adalah metode Growth Chart.
3. Fase Pengolahan Data (Data Preparation Phase) : Pada tahap ini dilakukan identifikasi dan pembangunan jawaban dari data yang telah dikumpulkan untuk bisa melakukan pengelompokkan dan pemilihan ke dalam kelompok-kelompok yang telah ditentukan. Jumlah kelompok atau target yang digunakan pada penelitian ini terlihat pada tabel berikut :

Tabel 1. Status Gizi Balita

No.	Kriteria	Status Gizi
1.	BB ideal < 70%	Gizi Buruk
2.	BB ideal 70% - 80%	Gizi Kurang
3.	BB ideal 80% - 100%	Gizi Baik
4.	BB ideal 100% – 110%	Gizi Lebih
5.	BB ideal >110%	Obesitas

4. Fase Pemodelan (Modeling Phase) : Pada fase ini dilakukan pemilihan model yang akan digunakan untuk melakukan

pengelompokkan status gizi balita. Model atau metode yang akan digunakan pada penelitian ini adalah metode K-Means. Jumlah data latih yang akan digunakan sebanyak ± 50 data balita.

5. Fase Evaluasi (Evaluation Phase) : Pengujian akan dilakukan dengan membandingkan pengelompokkan yang dilakukan oleh algoritma K-Means dengan pengelompokkan oleh bidan desa di posyandu.

2. Hasil dan Pembahasan

Tujuan bisnis yang akan dilakukan pada penelitian ini adalah bagaimana melakukan pengelompokkan nilai gizi balita dengan metode K-Means lalu diukur tingkat keakuratannya. Data yang digunakan merupakan data primer dengan responden balita dengan umur dibawah 36 bulan. Parameter yang digunakan untuk pengelompokkan status gizi berjumlah 2 yaitu tinggi badan (TB) dan berat badan (BB).

Jumlah data yang digunakan berjumlah 50 data balita di posyandu desa Badas dengan usia kurang dari 36 bulan. Data tabel dapat dilihat pada tabel 2.

Tabel 2. Sample Data Balita

Balita ke-	Tinggi Badan (TB)	Berat Badan (BB)
Balita 1	65	5,8
Balita 2	65	7,2
Balita 3	60	5
Balita 4	60	8
Balita 5	52	5,8
Balita 6	51	5
Balita 7	54	3,5
Balita 8	52,5	7,8
Balita 9	70	4,2
Balita 10	71	6,2
Balita 11	72,5	7
Balita 12	71,5	8,5
Balita 13	55	5,5

Balita 14	57	4,8
Balita 15	52	6,5
Balita 16	46,5	5,7
Balita 17	95	12
Balita 18	82	9,7
Balita 19	75	8
Balita 20	99	11

Data diatas tidak dapat langsung dilakukan pemrosesan karena besaran angka yang jauh antara variabel tinggi badan dan berat badan. Perbedaan jarak ini akan menyulitkan dalam proses pengelompokkan. Salah satu solusi digunakan untuk memperkecil besaran angka antar variabel adalah dengan melakukan normalisasi angka yang ada di variabel tinggi badan dan berat badan menggunakan persamaan di bawah ini.

Nilai Normalisasi =

$$\frac{\text{Nilai Awal} - \text{Nilai Minimal}}{\text{Nilai Maksimal} - \text{Nilai Minimal}}$$

Nilai variabel tinggi 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. Adapapun proses normalisasi adalah :

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

Nilai Maksimum (X_{maks}) = 99

Nilai Minimum (X_{min}) = 46,5

- b. Menghitung nilai normalisasi menggunakan persamaan di bawah.

$$X_{11} = \frac{X_{balita\ 1} - X_{min}}{X_{maks} - X_{min}} = \frac{65 - 46,5}{99 - 46,5} = 0,35$$

$$X_{12} = \frac{X_{balita\ 1} - X_{min}}{X_{maks} - X_{min}} = \frac{65 - 46,5}{99 - 46,5} = 0,35$$

$$X_{13} = \frac{X_{balita\ 1} - X_{min}}{X_{maks} - X_{min}} = \frac{60 - 46,5}{99 - 46,5} = 0,26$$

$$X_{14} = \frac{X_{balita\ 1} - X_{min}}{X_{maks} - X_{min}}$$

$$= \frac{60 - 46,5}{99 - 46,5} = 0,26$$

$$X_{15} = \frac{X_{balita\ 1} - X_{min}}{X_{maks} - X_{min}} = \frac{52 - 46,5}{99 - 46,5} = 0,10$$

$$X_{16} = \frac{X_{balita\ 1} - X_{min}}{X_{maks} - X_{min}} = \frac{51 - 46,5}{99 - 46,5} = 0,09$$

Perhitungan yang sama dilakukan hingga balita ke-50. Sample hasil dari normalisasi pada variabel tinggi badan dapat dilihat pada tabel 3.

Tabel 3. Sample Normalisasi Data Balita

Balita ke-	Tinggi Badan (TB)	Normalisasi
Balita 1	65	0,35
Balita 2	65	0,35
Balita 3	60	0,26
Balita 4	60	0,26
Balita 5	52	0,10
Balita 6	51	0,09
Balita 7	54	0,14
Balita 8	52,5	0,11
Balita 9	70	0,45
Balita 10	71	0,47
Balita 11	72,5	0,50
Balita 12	71,5	0,48
Balita 13	55	0,16
Balita 14	57	0,20
Balita 15	52	0,10
Balita 16	46,5	0,00
Balita 17	95	0,92
Balita 18	82	0,68
Balita 19	75	0,54
Balita 20	99	1,00

Perhitungan dan persamaan yang sama digunakan untuk menormalisasi variabel berat badan. Setelah angka pada masing-masing variabel dilakukan normalisasi, langkah selanjutnya adalah menentukan jumlah kelompok atau cluster. Ke 50 data balita akan dikelompokkan kedalam 5 cluster yaitu : Gizi Buruk, Gizi kurang, Gizi Baik, Gizi Lebih dan Obesitas.

Setelah jumlah cluster ditentukan, langkah selanjutnya adalah melakukan penentuan

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 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 maka proses iterasi dilanjutkan hingga nilai sama atau sampai dengan nilai maksimum iterasi yang telah ditetapkan sebelumnya (misal 100). Namun jika nilai initial cluster centre yang baru sama dengan initial cluster centre yang lama, maka proses pengelompokkan berhenti.

Tabel 4. Initial Cluster Centre

No.	Status Gizi	TB (cm)	BB (kg)
1.	Gizi Buruk	0,92	1,00
2.	Gizi Kurang	1,00	0,46
3.	Gizi Baik	0,35	0,63
4.	Gizi Lebih	0,58	0,13
5.	Obesitas	0,00	0,08

Persamaan yang digunakan untuk menghitung jarak pada penelitian ini adalah Euclidean Distance yang memiliki persamaan sebagai berikut :

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Contoh perhitungan jarak data ke-1 pada masing-masing cluster adalah :

$$\begin{aligned} D(x_{1,C1}) &= \sqrt{(TB1 - TBc1)^2 + (BB1 - BBc1)^2} \\ &= \sqrt{(0,35 - 0,92)^2 + (0,25 - 1)^2} \\ &= 0,9427 \end{aligned}$$

$$\begin{aligned} D(x_{1,C2}) &= \sqrt{(TB1 - TBc1)^2 + (BB1 - BBc1)^2} \\ &= \sqrt{(0,35 - 1)^2 + (0,25 - 0,46)^2} \\ &= 0,682 \end{aligned}$$

$$\begin{aligned} D(x_{1,C3}) &= \sqrt{(TB1 - TBc1)^2 + (BB1 - BBc1)^2} \\ &= \sqrt{(0,35 - 0,35)^2 + (0,25 - 0,63)^2} \\ &= 0,383 \end{aligned}$$

$$\begin{aligned} D(x_{1,C4}) &= \sqrt{(TB1 - TBc1)^2 + (BB1 - BBc1)^2} \\ &= \sqrt{(0,35 - 0,58)^2 + (0,25 - 0,13)^2} \\ &= 0,256 \end{aligned}$$

$$\begin{aligned} D(x_{1,C5}) &= \sqrt{(TB1 - TBc1)^2 + (BB1 - BBc1)^2} \\ &= \sqrt{(0,35 - 0)^2 + (0,25 - 0,08)^2} \\ &= 0,390 \end{aligned}$$

Persamaan dan perhitungan yang sama diterapkan di 50 data untuk mendapatkan jarak tiap data pada masing-masing cluster seperti pada tabel 5.

Tabel 5. Jarak data pada tiap cluster

Nama	Jarak C1	Jarak C2	Jarak C3	Jarak C4	Jarak C5
Balita 1	0,9427	0,682	0,383	0,256	0,390
Balita 2	0,8275	0,651	0,232	0,352	0,475
Balita 3	1,0690	0,801	0,378	0,324	0,270
Balita 4	0,8401	0,743	0,173	0,479	0,479
Balita 5	1,1096	0,920	0,455	0,490	0,197
Balita 6	1,1830	0,962	0,548	0,495	0,118
Balita 7	1,2665	0,973	0,663	0,456	0,164
Balita 8	0,9686	0,886	0,289	0,572	0,399
Balita 9	1,0384	0,673	0,563	0,143	0,448
Balita 10	0,8421	0,560	0,359	0,196	0,512
Balita 11	0,7546	0,512	0,292	0,261	0,577
Balita 12	0,6409	0,530	0,156	0,421	0,660
Balita 13	1,0913	0,873	0,456	0,427	0,211
Balita 14	1,1218	0,862	0,513	0,380	0,209
Balita 15	1,0600	0,906	0,393	0,513	0,264
Balita 16	1,1955	1,025	0,527	0,590	0,157
Balita 17	0,0861	0,460	0,640	0,856	1,245
Balita 18	0,4130	0,384	0,328	0,545	0,895
Balita 19	0,6392	0,458	0,242	0,356	0,677
Balita 20	0,2094	0,346	0,674	0,796	1,236

Setelah masing-masing data dihitung jaraknya untuk setiap cluster, langkah selanjutnya adalah mengelompokkan data sesuai dengan clusternya. Kelompok cluster suatu data diambil dari jarak terpendek data tersebut terhadap suatu cluster. Misalnya untuk data balita 1 memiliki jarak 0,9427 terhadap cluster 1. Pada cluster 2 memiliki jarak 0,682. Pada cluster 3 memiliki jarak 0,383. Pada cluster 4 memiliki jarak 0,256 dan pada cluster 5 memiliki jarak 0,390. Dari kelima cluster tersebut data balita 1 memiliki jarak terpendek dengan cluster 4. Oleh karena itu data balita 1 masuk dalam cluster 4. Langkah yang sama di terapkan sampai data ke 50 untuk melakukan pengelompokkan

data iterasi 1. Contoh hasil dapat dilihat pada tabel 6.

Tabel 6. Penempatan data pada cluster dengan jarak terdekat.

Nama	Jarak C1	Jarak C2	Jarak C3	Jarak C4	Jarak C5
Balita 1				*	
Balita 2			*		
Balita 3					*
Balita 4			*		
Balita 5					*
Balita 6					*
Balita 7					*
Balita 8			*		
Balita 9				*	
Balita 10				*	
Balita 11				*	
Balita 12			*		
Balita 13					*
Balita 14					*
Balita 15					*
Balita 16					*
Balita 17	*				
Balita 18			*		
Balita 19			*		
Balita 20	*				

Setelah data dikelompokkan sesuai clusternya, langkah berikutnya adalah menghitung nilai centroid baru di masing-masing cluster menggunakan persamaan berikut :

$$C_i = \frac{1}{M} \sum_{j=1}^M X_j$$

$$C_{1TB} = \frac{(TB17+TB20+TB25)}{3} = 0,95$$

$$C_{1BB} = \frac{(BB17+BB20+BB25)}{3} = 0,91$$

Rumusan perhitungan yang sama dilakukan untuk menghitung nilai centroid baru pada masing-masing cluster. Nilai centroid baru pada masing-masing cluster pada dilihat pada tabel 7.

Tabel 7. Initial Cluster Centre

No.	Status Gizi	TB (cm)	BB (kg)
1.	Gizi Buruk	0,95	0,91
2.	Gizi Kurang	0,90	0,56

No.	Status Gizi	TB (cm)	BB (kg)
3.	Gizi Baik	0,41	0,54
4.	Gizi Lebih	0,46	0,25
5.	Obesitas	0,13	0,19

Setelah nilai centroid baru dihitung, langkah selanjutnya adalah dibandingkan dengan nilai centroid sebelumnya (pada iterasi ini dibandingkan dengan nilai initial cluster centre). Jika nilainya sama maka proses iterasi dihentikan. Namun jika nilainya tidak sama, maka proses pengelompokkan diulangi kembali.

Tabel 8. Hasil Cluster

Nama	Cluster	Jarak Dengan Pusat Cluster
Balita 1	4	0,108
Balita 2	3	0,151
Balita 3	5	0,140
Balita 4	3	0,167
Balita 5	5	0,051
Balita 6	5	0,60
Balita 7	5	0,208
Balita 8	5	0,253
Balita 9	4	0,168
Balita 10	4	0,044
Balita 11	4	0,139
Balita 12	3	0,062
Balita 13	5	0,035
Balita 14	5	0,099
Balita 15	5	0,116
Balita 16	5	0,132
Balita 17	1	0,027
Balita 18	2	0,220
Balita 19	3	0,134
Balita 20	1	0,110

Setelah ditemukan cluster dan jarak antar cluster. Pengujian akurasi dilakukan dengan membandingkan pengelompokkan hasil K-Means dengan pengelompokkan yang dilakukan oleh bidan desa dengan menggunakan tabel Growth Chart.

Tabel 9. Contoh Hasil Menggunakan Growth Chart

Nama	Cluster
------	---------

Nama	Cluster
Balita 1	2
Balita 2	4
Balita 3	3
Balita 4	5
Balita 5	5
Balita 6	5
Balita 7	3
Balita 8	5
Balita 9	1
Balita 10	2
Balita 11	2
Balita 12	3
Balita 13	5
Balita 14	3
Balita 15	5
Balita 16	5
Balita 17	2
Balita 18	3
Balita 19	3
Balita 20	1

Dengan membandingkan hasil cluster menggunakan metode K-Means dengan menggunakan tabel Growth Chart diperoleh hasil sebagai berikut :

- Terdapat 17 balita dengan hasil pengelompokkan yang sama yaitu balita ke-5,6,12,13,15,16,19,20,23,27,36,40,41,44,46 dan balita ke-48.
- 33 balita lainnya terdapat perbedaan dalam penentuan cluster.
- Tingkat keakuratan kebenaran algoritma K-Means dalam pengelompokkan status gizi balita sebesar $\frac{17}{50} = 34\%$

Kesimpulan

Hasil perbandingan metode K-Means dengan menggunakan tabel Growth Chart diperoleh 17 data yang memiliki kelompok yang sama. Dari hasil tersebut dapat disimpulkan bahwa tingkat keakuratan algoritma K-Means adalah sebesar 34% dengan data sample sebanyak 50 balita.

Tingkat keakuratan metode K-Means untuk menentukan status gizi balita termasuk rendah. Oleh karena itu perlu di uji cobakan metode yang lain yang memiliki tingkat keakuratan yang lebih baik.

Daftar Pustaka

- Jakarta Post, 2015. The Jakarta Post. "http://www.thejakartapost.com/news/2015/06/24/indonesia-s-newsborns-face-future-challenges-due-malnutrition.html"
- Rahim, F.K., 2014. Faktor Risiko Underweight Balita Umur 7-59 Bulan. *Jurnal Kesehatan Masyarakat*, Volume 9, 115-121.
- Witten, E.A., 2012. *Data Mining Practical Mechine Learning*. 2nd Edition. San Fransico : Morgan Kaufmann.
- Kusrini dan E.T. Luthfi., 2009, *Algoritma Data Mining*, Andi, Yogyakarta.
- Anggraeni, R., Indrarti, A. 2010. Klasifikasi Status Gizi Balita Berdasarkan Indeks Antropometri (BB/U) Menggunakan Jaringan Syarat Tiruan. *Jurnal SNASTI, ICCS*, 14-18
- Larose, Daniel. 2005. *Discovering Knowledge in Data : An Introduction to Data Mining*. Canada : Wiley Publising.