

CLUSTERING MENGGUNAKAN METODE K-MEANS UNTUK MENENTUKAN STATUS GIZI BALITA

Windha Mega Pradnya Duhita

¹Sistem Informasi STMIK AMIKOM Yogyakarta
Ring Road Utara Condong Catur Sleman Yogyakarta 55283
Telp. (0274) 884201 - 207 Fax. (0274) 884208
e-mail : windha.m@amikom.ac.id

ABSTRACT

Malnutrition is one of the health problems that quite often among toddlers in Indonesia. Trace data from the WHO, the number of infants who had died from malnutrition in Indonesia in 2012 was 29 out of 1000 births. Caring parents and village officials (in this case the officer Public Health Service Center - HEALTH) to monitor the nutritional indispensable. Research conducted to try to perform a grouping of 50 children in the village of Karang Songo into 5 clusters nutritional status. Grouping nutritional status of children in the village of Songo Flower using the K-Means method is done through several stages, namely: the determination of business objectives, data collection 50 children in the village of Songo coral, grouping balitake nutritional status in five clusters, namely cluster 1 - malnutrition; cluster 2 - malnutrition; cluster 3 - good nutrition; cluster 4 - nutrition; Cluster 5 - obesity, cluster calculations using SPSS software, analysis of the output data, grouping nutritional status of children using tables Growth Chart, and the latter tested by comparing the results of the grouping of K-means algorithm and tables Growth Chart. By comparing the results of grouping using a table Growth Chart and K-Means algorithm obtained 17 data have the same group. From this figure it can be concluded that the K-Means algorithm only has an accuracy score of 34% correct.

Keywords—Data Mining , Nutritional Status , K -Means

ABSTRAK

Kekurangan gizi atau yang biasa disebut malnutrisi merupakan salah satu masalah kesehatan yang cukup sering menimpa balita-balita di Indonesia. Merunut data dari WHO, jumlah balita yang meninggal dunia akibat kekurangan gizi di Indonesia tahun 2012 adalah 29 dari 1000 kelahiran. 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 50 balita di desa Karang Songo kedalam 5 cluster status gizi. Pengelompokan status gizi balita di Desa Kembang Songo menggunakan metode K-Means dilakukan melalui beberapa tahapan yaitu : penentuan tujuan bisnis, pengumpulan data 50 balita di Desa karang Songo, pengelompokan status gizi balitake dalam 5 cluster yaitu cluster 1 - gizi buruk; cluster 2 - gizi kurang; cluster 3 - gizi baik; cluster 4 - gizi lebih; cluster 5 - obesitas, perhitungan cluster menggunakan software SPSS, analisa hasil data output, pengelompokan status gizi balita menggunakan tabel Growth Chart, dan yang terakhir melakukan pengujian dengan membandingkan hasil pengelompokan algoritma K-means dan tabel Growth Chart.

Dengan membandingkan hasil pengelompokan menggunakan tabel Growth Chart dan algoritma K-Means didapat 17 data yang memiliki kelompok yang sama. Dari angka ini dapat disimpulkan bahwa algoritma K-Means hanya memiliki nilai akurasi 34% benar.

Kata Kunci—Data Mining, Status Gizi, K-Means

I. PENDAHULUAN

Kekurangan gizi atau yang biasa disebut malnutrisi merupakan salah satu masalah kesehatan yang cukup sering menimpa balita-balita di Indonesia. *The Government's Basic Health Research* (Riskesdas) menunjukkan bahwa jumlah balita di Indonesia usia 12 bulan sampai dengan 59 bulan penderita gizi buruk pada tahun 2013 mencapai 28,1 persen [1]. Jumlah balita yang meninggal dunia akibat kekurangan gizi adalah 29 dari 1000 kelahiran [2].

Malnutrisi pada balita di Indonesia disebabkan oleh beberapa faktor seperti [3] : 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 malnutrisi pada balita. Makanan yang 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.

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 [3]. 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 [4]. 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 [4]. Akibat lain yang disebabkan oleh malnutrisi adalah penurunan daya tahan tubuh [5]. 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 [4]. 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 [6]. Dampak yang lebih parah dari malnutrisi pada balita adalah timbulnya kecacatan, tingginya angka kesakitan dan percepatan kematian [3].

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.

Larose dalam buku yang ditulis oleh Kusri dan Luthfi mengelompokkan Data Mining dapat dibagi menjadi 6 kelompok yaitu deskripsi, estimasi, prediksi, klasifikasi, *clustering* (pengelompokan), dan asosiasi [7]. *Clustering* akan melakukan pengelompokan data-data ke dalam sejumlah kelompok (*cluster*) berdasarkan kesamaan karakteristik masing-masing data pada kelompok-kelompok yang ada [9]. Banyak metode yang bisa digunakan untuk melakukan *clustering* diantaranya : metode K-Means, metode LVQ (*Learning Vector Quantization*), FCM (*Fuzzy C-Means*), dan lain sebagainya.

Penelitian mengenai pemanfaatan Algoritma Data Mining telah banyak dilakukan sebelumnya, seperti terlihat pada tabel 1.

Tabel 1. Penelitian Terdahulu

No.	Judul	Kelebihan dan Kekurangan	Perbedaan Dengan Penelitian Yang Dilakukan
1.	Bagging Regresi Logistik Ordinal Pada Status Gizi Balita [10]	Kelebihan : Metode yang digunakan untuk melakukan pengelompokan adalah Regresi Logistik Ordinal. Metode ini mampu melakukan pengelompokan status gizi balita dengan standar WHO. Kelemahan : nilai akurasi yang dihasilkan oleh algoritma ini dibandingkan dengan data WHO memiliki nilai eror sebesar 22,046%.	Perbedaannya pada algoritma yang digunakan untuk melakukan pengelompokan. Penelitian ini menggunakan metode K-Means.
2.	Analisa Klasifikasi Status Gizi dengan Metode Fuzzy C-Means Menggunakan Aplikasi Berbasis Android [11]	Kelebihan : Metode yang digunakan untuk melakukan penentuan status nilai gizi balita adalah Fuzzy C-Means. Kekurangan : Presentase keakuratan hasil klasifikasi yang dihasilkan oleh sistem cukup rendah berkisar 22,81% hingga 28,07%.	Perbedaannya pada algoritma yang digunakan untuk melakukan pengelompokan. Penelitian ini menggunakan metode K-Means.
3.	Clustering Kualitas Beras Berdasarkan Ciri Fisik Menggunakan Metode K-Means [12]	Kelebihan : Metode yang digunakan untuk melakukan pengelompokan kualitas beras adalah K-Means. Metode K-Means mampu membantu petugas BULOG untuk mengelompokkan beras berdasarkan beberapa parameter kualitas dengan tepat, cepat. Validasi keakuratan yang dihasilkan sebesar 92,82% Kelemahan : jumlah parameter yang terlalu banyak membuat aplikasi berjalan cukup lambat.	Perbedaannya pada algoritma yang digunakan untuk melakukan pengelompokan. Penelitian ini menggunakan metode K-Means. Parameter yang digunakan sebanyak 2 yaitu berat badan (BB) dan tinggi badan (TB).

K-Means merupakan salah satu algoritma dalam data mining yang bisa digunakan untuk melakukan pengelompokan/clustering suatu data. Ada banyak pendekatan untuk membuat cluster, diantaranya adalah membuat aturan yang mendikte keanggotaan dalam group yang sama berdasarkan tingkat persamaan diantara anggota-anggotanya. Pendekatan lainnya adalah dengan membuat sekumpulan fungsi yang mengukur beberapa properti dari pengelompokan tersebut sebagai fungsi dari beberapa parameter dari sebuah clustering [8]. Metode K-Means adalah metode yang termasuk dalam algoritma clustering berbasis jarak yang membagi data ke dalam sejumlah cluster dan algoritma ini hanya bekerja pada atribut numerik [8].

Pengelompokan data dengan metode K-Means dilakukan dengan algoritma [9]:

1. Tentukan jumlah kelompok
2. Alokasikan data ke dalam kelompok secara acak
3. Hitung pusat kelompok (*centroid*/rata-rata) dari data yang ada di masing-masing kelompok. Lokasi centroid setiap kelompok diambil dari rata-rata (mean) semua nilai data pada setiap fiturnya. Jika M menyatakan jumlah data dalam sebuah kelompok, i menyatakan fitur ke- i dalam sebuah

kelompok, dan p menyatakan dimensi data, maka persamaan untuk menghitung centroid fitur ke- i digunakan persamaan 1.

$$C_i = \frac{1}{M} \sum_{j=1}^M x_j \quad (1)$$

persamaan 1 dilakukan sebanyak p dimensi dari $i=1$ sampai dengan $i=p$.

4. Alokasikan masing-masing data ke centroid/rata-rata terdekat. Ada beberapa cara yang dapat dilakukan untuk mengukur jarak data ke pusat kelompok, diantaranya adalah Euclidean [9]. Pengukuran jarak pada ruang jarak (*distance space*) Euclidean dapat dicari menggunakan persamaan 2.

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

Pengalokasian kembali data ke dalam masing-masing kelompok dalam metode K-Means didasarkan pada perbandingan jarak antara data dengan centroid setiap kelompok yang ada [9]. Data dialokasikan ulang secara tegas ke kelompok yang mempunyai centroid dengan jarak terdekat dari data tersebut. Pengalokasian data ini menurut MacQueen (1967) dapat ditentukan menggunakan persamaan 3 [9].

$$a_{i1} = \begin{cases} 1 & d = \min \{D(x_i, c_1)\} \\ 0 & \text{lainnya} \end{cases} \quad (3)$$

a_{i1} adalah nilai keanggotaan titik x_i ke pusat kelompok c_1 , d adalah jarak terdekat dari data x_i ke K kelompok setelah dibandingkan, dan c_1 adalah centroid (pusat kelompok) ke-1.

Fungsi objektif yang digunakan untuk metode K-Means ditentukan berdasarkan jarak dan nilai keanggotaan data dalam kelompok. Fungsi objektif menurut MacQueen (1967) dapat ditentukan menggunakan persamaan 4 [9].

$$J = \sum_{i=1}^n \sum_{l=1}^k a_{il} D(x_i, c_l)^2 \quad (4)$$

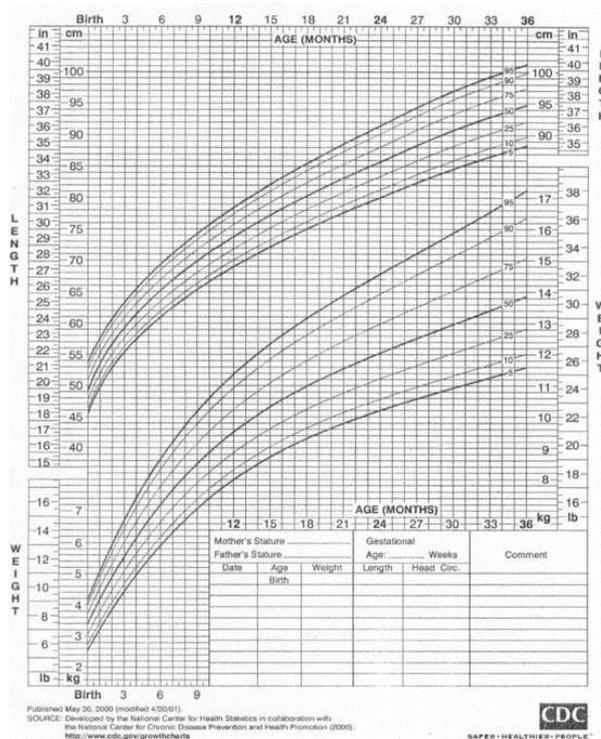
n adalah jumlah data, k adalah jumlah kelompok, a_{il} adalah nilai keanggotaan titik data x_i ke kelompok c_l yang diikuti. a mempunyai nilai 0 atau 1. Apabila data merupakan anggota suatu kelompok, nilai $a_{il} = 1$. Jika tidak, nilai $a_{il} = 0$.

5. 5. Kembali ke langkah 3, apabila masih ada data yang berpindah kelompok atau apabila ada perubahan nilai centroid di atas nilai ambang yang ditentukan, atau apabila perubahan nilai pada fungsi objektif yang digunakan masih di atas nilai ambang yang ditentukan.

2.1 Malnutrisi

Status gizi dapat didefinisikan sebagai bentuk ekspresi dari keadaan keseimbangan menggunakan variabel-variabel tertentu atau bisa juga perwujudan dari pemenuhan nutrisi menggunakan variabel-variabel tertentu [10]. Status gizi balita dapat diukur secara antropometri, atau bisa juga menggunakan Growth Chart yang dapat diunduh pada laman <http://cdc.gov/growthcharts/> atau standar WHO yang ada di laman <http://www.who.int/childgrowth/en/>.

Penelitian ini menggunakan Growth Chart dalam penentuan status gizi balita. Growth chart yang digunakan sebagai penentu status gizi balita dapat dilihat di gambar 1.

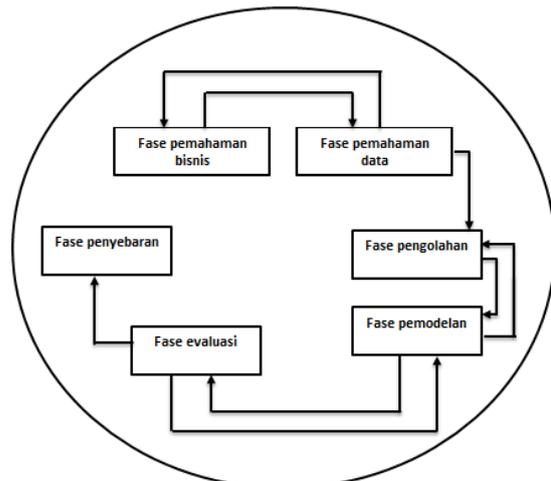


Gambar 1. Growth Chart

II. METODE PENELITIAN

Model yang digunakan untuk menyelesaikan penelitian ini adalah CRISP-DM (Cross Industry Standard Process for Data Mining). Model CRISP-DM diperkenalkan pertengahan tahun 1990 oleh sebuah perusahaan konsorsium Eropa [11].

Dalam CRISP-DM, sebuah proyek Data Mining memiliki siklus hidup yang terbagi dalam 6 fase seperti terlihat pada gambar 1.



Gambar 2 Siklus hidup CRISP-DM

Adapun langkah-langkah yang dilakukan pada penelitian ini menurut pada gambar 2 adalah sebagai berikut :

1. Fase Pemahaman Bisnis (*Business Understanding*) : meliputi penentuan tujuan bisnis, menilai situasi saat ini, menetapkan tujuan data mining, dan mengembangkan rencana proyek. Tujuan bisnis yang akan dilakukan pada penelitian ini adalah bagaimana melakukan pengelompokan nilai gizi balita di Desa Karang Songo, Jetis, Bantul menggunakan metode K-Means.
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 yang diusulkan ini menggunakan data primer, dengan respondennya adalah balita dengan umur dibawah 36 bulan sesuai data

yang ada di POSYANDU (Pos Pelayanan Terpadu) Desa Karang Songo. Perhitungan yang akan digunakan untuk menentukan status gizi balita adalah metode *Growth Chart* yang diambil dari laman <http://cdc.gov/growthcharts/>.

3. Fase Pengolahan Data (*Data Preparation Phase*) : Pada tahap ini dilakukan identifikasi dan pembangunan jawaban dari data yang telah dikumpulkan untuk bisa melakukan pengelompokan dan pemilahan ke dalam kelompok-kelompok yang telah ditentukan. Jumlah kelompok atau target yang digunakan pada penelitian ini dapat dilihat pada tabel 2.

Tabel 2. Status gizi balita

No.	Kriteria	Status Gizi
1.	BB saat ini/BB ideal < 70%	Gizi buruk
2.	BB saat ini/BB ideal 70% – 80%	Gizi kurang
3.	BB saat ini/BB ideal 80% – 100%	Gizi baik
4.	BB saat ini/BB ideal 100% – 110 %	Gizi lebih
5.	BB saat ini/BB idea > 110%	Obesitas/ <i>Obesity</i>

4. Fase Pemodelan (*Modeling Phase*) : Pada fase ini dilakukan pemilihan model yang akan digunakan untuk melakukan pengelompokan 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 di Desa Karang Songo.
5. Fase Evaluasi (*Evaluation Phase*) :

Pengujian akan dilakukan dengan membandingkan pengelompokan yang dilakukan oleh algoritma K-Means dengan pengelompokan yang dilakukan oleh Bidan Desa Karang songo.

6. Fase Penyebaran (*Deployment Phase*) : Fase ini dilakukan guna penemuan pengetahuan (identifikasi hubungan yang tak terduga dan berguna) untuk kemudian diterapkan pada operasi bisnis di berbagai tujuan, termasuk clustering.

III. HASIL DAN PEMBAHASAN

Tujuan bisnis yang akan dilakukan pada penelitian ini adalah bagaimana melakukan pengelompokan nilai gizi balita di Desa Karang Songo, Jetis, Bantul menggunakan metode K-Means. Penelitian dibuat menggunakan data primer, dengan respondennya adalah balita-balita di bawah usia 36 bulan yang ada di data POSYANDU Desa Karang Songo, Jetis, Bantul, Yogyakarta. Parameter yang digunakan untuk melakukan pengelompokan status gizi balita berjumlah 2 yaitu tinggi badan balita (TB) dan berat badan balita (BB).

Jumlah data yang akan digunakan sebanyak 50 data balita di Desa Karang Songo dengan usia kurang dari 36 bulan. Data ini dapat dilihat pada tabel 3.

Tabel 3. 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
Balita 21	99	7,8
Balita 22	97,5	10
Balita 23	88	9,4
Balita 24	75	10,1
Balita 25	95	12,8
Balita 26	72	10,2
Balita 27	50	6
Balita 28	67	5
Balita 29	68	8,2
Balita 30	65	9,4
Balita 31	61	7,1
Balita 32	62	5,8
Balita 33	53	3,5
Balita 34	55	5,8
Balita 35	54	3,5
Balita 36	52,5	6,8
Balita 37	77	4,7
Balita 38	73	5,8
Balita 39	72,5	6,9
Balita 40	71,5	8,1
Balita 41	55	6,7
Balita 42	59	5,5

Balita 43	54	4,9
Balita 44	46,5	4,2
Balita 45	95	7,4
Balita 46	87	9,1
Balita 47	75	6,5
Balita 48	92,5	9,4
Balita 49	93	8,4
Balita 50	97,5	7,9

Data yang ada di tabel 3 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 persamaan 5 [11].

$$\text{nilai normalisasi} = \frac{(\text{nilai awal} - \text{nilai minimal})}{(\text{nilai maksimal} - \text{nilai minimal})} \quad (5)$$

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 [11]. Adapun tahapan yang dilakukan untuk 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 5.

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

$$= \frac{(65 - 46,5)}{(99 - 46,5)} = 0,35$$

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

$$= \frac{(65 - 46,5)}{(99 - 46,5)} = 0,35$$

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

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

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

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

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

$$= \frac{(52 - 46,5)}{(99 - 46,5)} = 0,10$$

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

$$= \frac{(51 - 46,5)}{(99 - 46,5)} = 0,09$$

Perhitungan yang sama dilakukan hingga balita ke-50. Hasil dari normalisasi pada variabel tinggi badan dapat dilihat pada tabel 4.

Tabel 4. Data normalisasi variabel tinggi badan

Balita ke-	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
Balita 21	99	1,00
Balita 22	97,5	0,97
Balita 23	88	0,79
Balita 24	75	0,54
Balita 25	95	0,92
Balita 26	72	0,49
Balita 27	50	0,07
Balita 28	67	0,39
Balita 29	68	0,41
Balita 30	65	0,35
Balita 31	61	0,28
Balita 32	62	0,30
Balita 33	53	0,12
Balita 34	55	0,16
Balita 35	54	0,14
Balita 36	52,5	0,11
Balita 37	77	0,58
Balita 38	73	0,50

Tabel 4. Lanjutan

Balita ke-	TB	Normalisasi
Balita 39	72,5	0,50
Balita 40	71,5	0,48
Balita 41	55	0,16
Balita 42	59	0,24
Balita 43	54	0,14
Balita 44	46,5	0,00
Balita 45	95	0,92
Balita 46	87	0,77
Balita 47	75	0,54
Balita 48	92,5	0,88
Balita 49	93	0,89
Balita 50	97,5	0,97

Perhitungan dan persamaan yang sama digunakan untuk melakukan normalisasi variabel berat badan.

Setelah angka pada masing-masing variabel dilakukan normalisasi, langkah selanjutnya adalah menentukan jumlah kelompok atau cluster. Ke 50 data balita yang ada di tabel 5 akan dikelompokkan ke dalam 5 cluster yaitu : Gizi buruk, Gizi kurang, Gizi baik, Gizi lebih, Obesitas.

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 pengelompokkan berhenti. Misalkan nilai initial cluster centre yang diberikan adalah (ditunjukkan di tabel 5) :

Tabel 5. 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

Nilai initial *cluster centre* pada tabel 6 digunakan untuk menghitung jarak antara data dengan *centroid*. Persamaan yang digunakan untuk menghitung jarak pada penelitian ini adalah *Euclidean Distance* (persamaan 2).

Adapun contoh perhitungan jarak data ke-1 pada masing-masing *cluster* adalah :

$$\begin{aligned} d(x_1, c_1) &= \sqrt{(TB_1 - TB_{c1})^2 + (BB_1 - BB_{c1})^2} \\ &= \sqrt{(0,35 - 0,92)^2 + (0,25 - 1)^2} \\ &= 0,9427 \end{aligned}$$

$$\begin{aligned} d(x_1, c_2) &= \sqrt{(TB_1 - TB_{c2})^2 + (BB_1 - BB_{c2})^2} \\ &= \sqrt{(0,35 - 1)^2 + (0,25 - 0,46)^2} \\ &= 0,682 \end{aligned}$$

$$\begin{aligned} d(x_1, c_3) &= \sqrt{(TB_1 - TB_{c3})^2 + (BB_1 - BB_{c3})^2} \\ &= \sqrt{(0,35 - 0,35)^2 + (0,25 - 0,63)^2} \\ &= 0,383 \end{aligned}$$

$$d(x_1, c_4) = \sqrt{(TB_1 - TB_{c_4})^2 + (BB_1 - BB_{c_4})^2}$$

$$= \sqrt{(0,35 - 0,58)^2 + (0,25 - 0,13)^2}$$

$$= 0,256$$

$$d(x_1, c_5) = \sqrt{(TB_1 - TB_{c_5})^2 + (BB_1 - BB_{c_5})^2}$$

$$= \sqrt{(0,35 - 0)^2 + (0,25 - 0,08)^2}$$

$$= 0,390$$

Persamaan dan perhitungan yang sama diterapkan di 50 data untuk mendapatkan jarak tiap data pada masing-masing cluster seperti pada tabel 6.

Tabel 6. Jarak data pada tiap cluster

Nama Balita	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	0,478	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	0,538	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
Balita 21	0,5436	0,002	0,671	0,536	1,071
Balita 22	0,3054	0,241	0,625	0,691	1,152
Balita 23	0,3879	0,273	0,440	0,547	0,966
Balita 24	0,4759	0,521	0,209	0,581	0,831
Balita 25	0,0038	0,545	0,683	0,935	1,304
Balita 26	0,5165	0,576	0,163	0,598	0,804
Balita 27	1,1237	0,953	0,459	0,532	0,200
Balita 28	0,9919	0,679	0,470	0,192	0,399
Balita 29	0,7108	0,592	0,138	0,412	0,590
Balita 30	0,6752	0,671	0,005	0,553	0,657
Balita 31	0,8889	0,727	0,254	0,398	0,413
Balita 32	0,9782	0,736	0,387	0,308	0,339
Balita 33	1,1880	0,940	0,559	0,456	0,130
Balita 34	1,0683	0,865	0,426	0,434	0,233
Balita 35	1,2665	0,973	0,663	0,456	0,164
Balita 36	1,0322	0,892	0,362	0,517	0,298
Balita 37	0,9346	0,534	0,552	0,001	0,583
Balita 38	0,8596	0,539	0,413	0,139	0,532
Balita 39	0,7635	0,514	0,302	0,250	0,572
Balita 40	0,6726	0,525	0,185	0,379	0,631
Balita 41	1,0025	0,846	0,342	0,470	0,310
Balita 42	1,0398	0,800	0,430	0,352	0,274
Balita 43	1,1513	0,911	0,522	0,438	0,159
Balita 44	1,3044	1,071	0,656	0,583	0,005
Balita 45	0,5807	0,086	0,611	0,449	0,086
Balita 46	0,4247	0,269	0,422	0,509	0,269
Balita 47	0,7753	0,477	0,363	0,196	0,196
Balita 48	0,3682	0,214	0,526	0,585	0,214
Balita 49	0,4744	0,132	0,546	0,501	0,132
Balita 50	0,5294	0,031	0,641	0,521	0,031

Setelah masing-masing data dihitung jaraknya untuk tiap cluster, langkah selanjutnya adalah mengelompokkan data

sesuai 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 ke-5 cluster tersebut, data balita 1 memiliki jarak terpendek dengan cluster 4. Oleh karena itu data balita 1 masuk ke dalam cluster 4. Langkah yang sama diterapkan di ke-50 data untuk melakukan pengelompokan di iterasi 1. Hasil pengelompokan dapat dilihat pada tabel 7.

Tabel 7. Penempatan data pada cluster dengan jarak terdekat

Nama Balita	C1	C2	C3	C4	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	*				

Tabel 7. Lanjutan

Nama Balita	C1	C2	C3	C4	C5
Balita 21		*			
Balita 22		*			
Balita 23		*			
Balita 24			*		
Balita 25	*				
Balita 26			*		
Balita 27					*
Balita 28				*	
Balita 29			*		
Balita 30			*		
Balita 31			*		
Balita 32				*	
Balita 33					*
Balita 34					*
Balita 35					*
Balita 36					*
Balita 37				*	
Balita 38				*	
Balita 39				*	
Balita 40			*		
Balita 41					*
Balita 42					*
Balita 43					*
Balita 44					*
Balita 45		*			
Balita 46		*			
Balita 47				*	
Balita 48		*			
Balita 49		*			
Balita 50		*			

Setelah data dikelompokkan sesuai clusternya, langkah selanjutnya adalah menghitung nilai *centroid* baru di masing-masing *cluster* menggunakan persamaan 1.

$$c_{1TB} = \frac{(TB_{17} + TB_{20} + TB_{25})}{3} = 0,95$$

$$c_{1BB} = \frac{(BB_{17} + BB_{20} + BB_{25})}{3} = 0,91$$

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

centroid baru pada masing-masing *cluster* dapat dilihat pada tabel 8.

Tabel 8 Nilai *centroid* baru

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

Hasil pengelompokan ke-50 data balita dan jaraknya dengan pusat *cluster* hasil pengelompokan menggunakan software SPSS ditunjukkan pada tabel 9.

Tabel 9. Hasil cluster

Nama Balita	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

Tabel 9. Lanjutan

Nama Balita	Cluster	Jarak Dengan Pusat Cluster
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
Balita 21	2	0,165
Balita 22	2	0,163
Balita 23	2	0,105
Balita 24	3	0,213
Balita 25	1	0,097
Balita 26	3	0,199
Balita 27	5	0,085
Balita 28	4	0,111
Balita 29	3	0,026
Balita 30	3	0,117
Balita 31	3	0,200
Balita 32	4	0,158
Balita 33	5	0,088
Balita 34	5	0,054
Balita 35	5	0,208
Balita 36	5	0,144
Balita 37	4	0,170
Balita 38	4	0,042
Balita 39	4	0,129
Balita 40	3	0,077
Balita 41	5	0,137
Balita 42	5	0,113
Balita 43	5	0,059
Balita 44	5	0,180
Balita 45	2	0,155
Balita 46	2	0,109
Balita 47	4	0,109
Balita 48	2	0,062
Balita 49	2	0,041
Balita 50	2	0,137

Pada tabel 9, balita 1 dengan tinggi badan 65 cm dan berat badan 5,8 kg dimasukkan ke cluster 4 yaitu kelompok balita dengan status gizi berlebih dengan

jarak sebesar 0,108 dari pusat cluster 4. Berbeda dengan balita 2 yang memiliki tinggi badan 65 cm dengan berat badan 7,2 kg dikelompokkan ke dalam cluster 3 yaitu kelompok balita dengan status gizi baik dengan jarak sebesar 0,151 dari pusat cluster 3. Adapun jarak antar pusat cluster ditunjukkan pada gambar 3.

Cluster	1	2	3	4	5
1					
2	.346				
3	.646	.458			
4	.820	.525	.290		
5	1.077	.829	.438	.333	

Gambar 3 Jarak antar pusat cluster

Pengujian dilakukan dengan membandingkan pengelompokan yang dilakukan oleh algoritma K-Means dengan pengelompokan yang dilakukan oleh Bidan Desa Karang songo menggunakan tabel Growth Chart.

Perhitungan berat badan ideal balita menggunakan tabel Growth chart (gambar 1) dapat dilakukan dengan menarik angka tinggi badan balita ke kanan hingga menyentuh garis persentil. Selanjutnya tarik ke bawah menuju kurva berat badan dan didapatkan nilai berat badan ideal balita tersebut. Untuk mengetahui status gizi balita dilakukan dengan melakukan pembagian berat badan balita sekarang dibagi dengan berat idealnya. Hasil perhitungan lalu dicocokkan dengan tabel 2 untuk mendapatkan status gizi balita tersebut.

Tabel 10. Hasil pengelompokan menggunakan tabel Growth Chart

Nama Balita	Status Gizi	Nama Balita	Status Gizi
Balita 1	2	Balita 7	3
Balita 2	4	Balita 8	5
Balita 3	3	Balita 9	1
Balita 4	5	Balita 10	2
Balita 5	5	Balita 11	2
Balita 6	5	Balita 12	3

Nama Balita	Status Gizi	Nama Balita	Status Gizi
Balita 13	5	Balita 19	3
Balita 14	3	Balita 20	1
Balita 15	5	Balita 21	1
Balita 16	5	Balita 22	1
Balita 17	2	Balita 23	2
Balita 18	3	Balita 24	4

Nama Balita	Status Gizi	Nama Balita	Status Gizi
Balita 25	3	Balita 32	3
Balita 26	5	Balita 33	4
Balita 27	5	Balita 34	4
Balita 28	1	Balita 35	3
Balita 29	4	Balita 36	5
Balita 30	5	Balita 37	1
Balita 31	5	Balita 38	1

Nama Balita	Status Gizi	Nama Balita	Status Gizi
Balita 39	2	Balita 46	2
Balita 40	3	Balita 47	1
Balita 41	5	Balita 48	2
Balita 42	4	Balita 49	1
Balita 43	4	Balita 50	1
Balita 44	5		
Balita 45	1		

Dengan membandingkan tabel 9 (tabel pengelompokan status gizi balita menggunakan Growth Chart) dan tabel 10 (pengelompokan dengan algoritma K-Means) didapatkan hasil sebagai berikut :

a. Terdapat 17 balita dengan hasil pengelompokan yang sama yaitu balita ke-

5,6,12,13,15,16,19,20,23,27,36,40,41,44, 46 dan balita ke-48

b. 33 balita lainnya terdapat perbedaan dalam penentuan kelompok atau cluster nya.

Dari hasil tersebut dapat disimpulkan bahwa kebenaran algoritma K-Means dalam mengelompokkan status gizi balita memiliki nilai akurasi sebesar 34% (17/50).

IV.SIMPULAN

Kesimpulan dari penelitian ini adalah dengan membandingkan hasil pengelompokan menggunakan tabel Growth Chart dan algoritma K-Means didapat 17 data yang memiliki kelompok yang sama. Dari angka ini dapat disimpulkan bahwa algoritma K-Means hanya memiliki nilai akurasi 34% benar. Nilai ini bisa berubah seiring dengan penambahan data latih.

PENELITIAN LANJUTAN

Saran dari penelitian ini adalah perlu dicoba algoritma lain untuk melakukan pengelompokan status gizi balita sehingga diharapkan pengelompokan yang dilakukan memiliki nilai akurasi yang lebih baik.

DAFTAR PUSTAKA

- [1] Jakarta Post, 2015, Indonesia's Newborns Face Future Challenges Due Malnutrition, <http://www.thejakartapost.com/news/2015/06/24/indonesia-s-newborns-face-future-challenges-due-malnutrition.html>, diakses tanggal 25 Juli 2015.
- [2] World Bank, 2014, World Development Indicators: Millennium Development Goals: eradicating poverty and saving lives, <http://wdi.worldbank.org/table/1.2>, 25 Juli 2015.
- [3] Rahim, F. K. 2014. Faktor Risiko Underweight Balita Umur 7-59 Bulan. *Jurnal Kesehatan Masyarakat*, No. 2, Vol. 9, 115-121.
- [4] Fatimah, S., et al. 2008. Faktor-faktor Yang Berkontribusi Terhadap Status Gizi Pada Balita di Kecamatan Ciawi Kabupaten Tasikmalaya. *Nourning Journal of Padjajaran University*, No. 17, Vol. 10, 37.
- [5] Adisasmito, W. 2007. Faktor Risiko Diare Pada Bayi dan Balita di Indonesia : Systematic Review Penelitian Akademik Bidang Kesehatan. *Jurnal MAKARA KESEHATAN*, No. 1, Vol. 11, 1-10.
- [6] Welasasih, B.D., Wirjatmadi, R.B. 2012. Beberapa Faktor Yang Berhubungan Dengan Status Gizi Balita Stunting. *The Indonesian Journal of Public Health*, No. 3, Vol. 8, 99-104.
- [7] Kusrini dan E.T. Luthfi., 2009, *Algoritma Data Mining*, Andi, Yogyakarta.
- [8] Witten, et al., 2012, *Data Mining Practical Machine Learning Tools and Technique*, 2nd Edition, Morgan Kaufmann, San Fransisco.
- [9] Prasetyo, E., 2012, *Data Mining : Konsep dan Aplikasi Menggunakan MATLAB*, Andi, Yogyakarta.
- [10] Anggraeni, R., Indrarti, A. 2010. Klasifikasi Status Gizi Balita Berdasarkan Indeks Antropometri (BB/U) Menggunakan Jaringan Syaraf Tiruan. *Jurnal SNASTI, ICCS*, 14-18.
- [11] Atthina, N., Iswari, I. 2015. Klasterisasi Data Kesehatan Penduduk untuk Menentukan Rentang Derajat Kesehatan Daerah dengan Metode K-Means. *Jurnal SNASTI*, B-52 - B-59.