

# Penerapan Clustering Bootstrap dengan Metode K-Means

Hendro Prasetyo<sup>1</sup>, Kuntoro<sup>2</sup>, Windhu Purnomo<sup>2</sup>, Soenarnatalina<sup>2</sup>, Merry Adriani<sup>3</sup>,  
Bambang Wijanarko<sup>4</sup>

<sup>1</sup>Prodi Kebidanan Jember Politeknik Kesehatan Kemenkes Malang

<sup>2</sup>Departemen Biostatistika dan Kependudukan  
Fakultas Kesehatan Masyarakat Universitas Airlangga

<sup>3</sup>Departemen Gizi Kesehatan  
Fakultas Kesehatan Masyarakat Universitas Airlangga

<sup>4</sup>Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam ITS Surabaya

Alamat korespondensi :

Hendro Prasetyo

hendroprasetyo53@yahoo.com

## ABSTRACT

*Cluster analysis was a process for grouping a set of objects based on data that have similar certain characteristic. K-Means was a method of cluster analysis which begins by determining the number of clusters desired. Bootstrap was a sampling technique with replacement from the original sample. Bootstrap was used to estimate the parameters based on minimal data using a computer. This method was useful to maximize relative difference and variation in the clusters. Malnutrition was a major problem in Indonesia and is still a concern in children under five. Infants with malnutrition would have a higher mortality rate. The purpose of this study was to assess the accuracy of K-Means and Bootstrap K-Means method to clustering nutritional status of children under five which was cross-tabulated with the nutritional status of children based on the WHO-2005 in the Ajung Public Health Center, Jember. The variable in this study was nutritional status based on WHO criteria 2005 as standard benchmarks, percentage and weight. This was non-reactive research, using secondary data in Ajung Public Health Center, without any direct interaction with the subject. This study concluded that the total accuracy rate (TAR) and Total Error Rate (TER) to determine nutritional status of K-Means method was TAR=0.9 and, TER=0.1; Bootstrap K-Means method (B=25) TAR=0.925 and TER=0.075; Bootstrap K-Means method (B=50) TAR=0.9417, TER=0.0583; and Bootstrap K-Means Bootstrap (B=75) TAR=0.9583 and TER=0.0417 after cross-tabulated with nutritional status based on WHO-2005 (weight for age). In conclusion general, the K-Means method and Bootstrap K-Means method and cross-tabulated with nutritional status based on WHO-2005 has shown very good accuracy to determine the nutritional status of children. The best method was Bootstrap K-Means (B=75). K-Means Bootstrap methods can be used as an alternative way to determine the nutritional status of children.*

*Keywords: cluster analysis with K-Means method, bootstrap, nutritional status*

## ABSTRAK

Analisis *Cluster* adalah sebuah proses untuk mengelompokkan satu set objek berdasarkan data yang telah similar certain karakteristik. *K-Means* adalah metode analisis cluster yang dimulai dengan menentukan jumlah cluster yang diinginkan. *Bootstrap* adalah teknik sampling dengan penggantian dari sampel asli. *Bootstrap* digunakan untuk memperkirakan parameter berdasarkan data minimal menggunakan komputer. Metode ini berguna untuk memaksimalkan relatif beda dan variasi dalam cluster. Malnutrisi merupakan masalah utama di Indonesia dan masih menjadi perhatian pada anak balita. Bayi dengan gizi buruk akan memiliki tingkat kematian yang lebih tinggi. Tujuan dari penelitian ini untuk menilai akurasi *K-Means* dan metode *Bootstrap K-Means* untuk pengelompokan status gizi anak under five yang cross-tabulated dengan status gizi anak berdasarkan WHO-2005 di Puskesmas Ajung, Jember. Variabel dalam penelitian ini adalah status gizi berdasarkan kriteria WHO 2005 tolok ukur sebagai standar, persentase dan berat. Ini adalah penelitian non-reaktif, menggunakan data sekunder di Ajung Puskesmas, tanpa interaksi langsung dengan subjek. Penelitian ini menyimpulkan bahwa total tingkat akurasi (TAR) dan *Total Error Rate* (TER) untuk menentukan status gizi metode *K-Means* adalah TAR=0,9 dan , TER=0,1 ; *Bootstrap K-Means* metode (B=25) TAR=0,925 dan TER=0,075 ; *Bootstrap K-Means* metode (B=50) TAR=0,9417 , TER=0,0583, dan *Bootstrap K-Means* (B=75) TAR=0,9583 dan TER=0,0417 setelah cross-tabulated

dengan status gizi berdasarkan WHO-2005 (berat badan untuk usia) . Dalam kesimpulan umum, metode K-Means dan metode Bootstrap K-Means dan crosstabulated dengan status gizi berdasarkan WHO-2005 telah menunjukkan akurasi yang sangat baik untuk menentukan status gizi anak . Metode terbaik adalah Bootstrap K-Means (B=75) . Metode K-Means Bootstrap dapat digunakan sebagai cara alternatif untuk menentukan status gizi anak.

Kata kunci : analisis cluster dengan metode , bootstrap , status gizi K-Means

## PENDAHULUAN

Kinerja komputer saat ini semakin pesat kemajuannya yang menyebabkan terjadinya perkembangan teori statistika yang mengarah pada kehandalan komputer, terutama kecepatan dan kemampuan menghitungnya. Dengan demikian berkembanglah metode iteratif, yang sering kali tidak memerlukan syarat-syarat ketat mengenai parameter sebaran populasinya. *Bootstrap* adalah suatu teknik penarikan sampel dengan pengembalian dari sebuah sampel asli. *Bootstrap* bertujuan memperoleh estimasi parameter berdasarkan data yang minimal dengan bantuan komputer (Teknomo, 2005).

Analisis klaster merupakan suatu proses untuk mengelompokkan kumpulan dari obyek data yang mempunyai kemiripan berdasarkan karakteristik tertentu (Gudono, 2011). Tujuan utama analisis klaster adalah mengklasifikasi obyek (kasus/elemen) seperti orang, produk atau barang, perusahaan kedalam kelompok yang relatif homogen didasarkan pada suatu set variabel yang dipertimbangkan untuk diteliti (Supranto, 2010).

K-Means adalah metode non hierarki dalam analisis cluster. Metode K-Means dimulai dengan menentukan terlebih dahulu jumlah klaster yang diinginkan. Metode ini berusaha meminimalkan variasi antar data dalam suatu klaster dan memaksimalkan variasi dengan data di klaster lainnya (Hill dan Lewicki, 2007).

Banyak kendala dalam mencapai pembangunan kesehatan, sehingga perlu adanya program kesehatan yang menyentuh langsung ke sasaran. Kendala itu diantaranya adalah masalah gangguan gizi yang masih banyak terjadi di daerah. Salah satu gangguan gizi adalah gizi buruk. Masalah gizi buruk pada balita bukan hanya disebabkan oleh kemiskinan, (masalah struktural) tapi juga karena aspek sosial dan budaya hingga menyebabkan tindakan yang tidak menunjang tercapainya gizi yang memadai untuk balita (Dwilistyowati, 2012).

Untuk memantau status gizi balita di Puskesmas Ajung digunakan standar klasifikasi WHO-2005 sebagai tolok ukur penentuan status gizi. Dalam penelitian ini diterapkan model klasifikasi status gizi balita menggunakan metode clustering bootstrap, dengan tujuan mendapatkan klaster yang sehomogen mungkin, sehingga digunakan dasar kesamaan skor nilai yang dianalisis. Data mengenai ukuran kesamaan tersebut, dianalisis dengan metode analisis cluster, sehingga akan dapat ditentukan siapa yang masuk klaster mana (Gudono, 2011). Karena jumlah klaster ditentukan terlebih dahulu, maka penerapan Clustering Bootstrap yang digunakan yaitu metode K-Means.

## METODE PENELITIAN

Jenis penelitian ini adalah penelitian non reaktif, yaitu peneliti tidak melakukan interaksi terhadap subyek penelitian, peneliti tidak memerlukan

respon dari responden, atau responden tidak ikut partisipasi aktif. Penelitian ini menggunakan sumber data sekunder dari data hasil kegiatan pemantauan status gizi di Puskesmas Ajung Kabupaten Jember. Status gizi diukur berdasarkan standar klasifikasi WHO-2005 yaitu berat badan menurut umur.

Sampel pada penelitian ini adalah balita yang tercatat datanya di bidang Gizi Puskesmas Ajung Kabupaten Jember pada bulan Pebruari 2013. Metode pengambilan sampelnya adalah *simple random sampling* yaitu sejumlah 120 balita. Variabel dalam penelitian ini adalah status gizi balita yang dihitung dari berat badan menurut umur sesuai standar yang diterapkan oleh WHO-2005.

Penelitian ini menelaah tentang penggunaan analisis kluster dengan Bootstrap yaitu mengklasifikasi responden ke dalam gugus yang relatif homogen yang didasarkan pada suatu set variabel dan mengklasifikasi setelah

dilakukan Bootstrap. Hasil analisis kluster akan diperbandingkan tingkat akurasi dengan klasifikasi status gizi standar WHO-2005 dengan cara tabulasi silang.

## HASIL PENELITIAN

Analisis deskripsi berat badan (BB) responden menunjukkan rerata BB balita 112,68 ons dengan standar deviasi 31,63 ons. Sedangkan umur balita rerata 32,37 bulan, dengan standar deviasi 18,96 bulan. Hal ini menunjukkan bahwa berat badan balita dan umur responden mempunyai kesenjangan yang cukup beragam.

Ketersebaran berat badan dan umur menurut penggolongan status gizinya berdasarkan WHO-2005 menunjukkan berat badan dan umur balita pada kelompok status gizinya mempunyai kesenjangan yang cukup beragam. Kondisi ini seperti yang ditunjukkan di Tabel 1.

Tabel 1. Analisis Deskriptif Umur dan Berat Menurut Status Gizinya dengan WHO-2005

	Berat Badan		Umur	
	Rerata	Standar Deviasi	Rerata	Standar Deviasi
Gizi Baik	130,65	12,36	44,58	10,98
Gizi Buruk	68,78	10,32	9,91	10,08
Gizi Kurang	93,23	8,95	19,58	8,42
Gizi Lebih	152,36	38,56	42,91	23,69

### Analisis Cluster Pada Status Gizi Balita Menggunakan Metode K-Means

Metode *clustering* nonhierarki dilakukan jika sebelumnya sudah ditentukan terlebih dahulu banyaknya kluster yang ingin dibentuk. Metode nonhierarki dikenal sebagai metode *K-means* (K-rataan). Pada metode ini, obyek dapat berpindah kluster pada setiap tahap *clustering*.

Algoritma untuk metode K-means ini adalah sebagai berikut:

a. Mengelompokkan obyek menjadi kluster awal sebanyak k.

b. Menghitung jarak *euclid* setiap obyek terhadap masing-masing kluster dan menempatkan kembali obyek ke dalam kluster berdasarkan jarak terdekat.

c. Mengulangi proses tersebut sampai tidak terjadi lagi perpindahan obyek ke kluster yang lain

Hasil pengklasteran berat dan umur berdasarkan status gizinya dengan menggunakan metode K-Means ditunjukkan pada Tabel 2.

Tabel 2. Pengelompokan Status Gizi Balita Menggunakan Metode *K-Means*

Status Gizi Balita	Jumlah Klaster	Berat Badan		Umur	
		Rerata	Standar Deviasi	Rerata	Standar Deviasi
Lebih	10	169,50	13,79	56,00	3,80
Kurang	31	96,68	8,94	19,52	10,34
Baik	54	131,69	9,72	46,20	8,33
Buruk	25	68,72	8,97	8,96	6,46

Tabel 2 menunjukkan untuk status gizi baik mempunyai anggota kluster paling banyak dengan rerata BB 131,69 ons dengan standar deviasi 9,72 ons, dan umur balita rerata 46,2 bulan dengan standar deviasi 8,33 bulan. Dan kluster status gizi lebih mempunyai anggota kluster paling sedikit dengan rerata BB 169,5 ons dengan standar deviasi 13,79 ons, dan umur balita rerata 56 bulan dengan standar deviasi 3,8 bulan. Kondisi ini menunjukkan variasi tertinggi terletak pada kelompok status gizi lebih.

#### **Analisis Cluster Pada Status Gizi Balita Menggunakan Metode K-Means Dengan Bootstrap (B=25, B=50 dan B=75)**

Pengelompokkan dengan metode *K-means* dengan *Bootstrap* dilakukan dengan resampling dengan pengembalian sebanyak 25 kali, 50 kali dan 75 kali dari sampel asli. Hal ini bertujuan untuk proses evaluasi dari pengelompokan status gizi menurut WHO-2005. Hasil analisis kluster status gizi berdasarkan berat badan menurut umur diuraikan dalam Tabel 3.

Tabel 3. Pengelompokan Status Gizi Balita Menggunakan Metode *K-Means* Dengan *Bootstrap* (B=25, B=50 dan B=75)

		B=25				B=50				
Status Gizi	Jumlah Balita klaster	Berat Badan		Umur		Jumlah klaster	Berat Badan		Umur	
		Rerata	Standar Deviasi	Rerata	Standar Deviasi		Rerata	Standar Deviasi	Rerata	Standar Deviasi
Lebih	11	157.09	34.47	47.91	19.93	11	157.09	34.47	47.91	19.93
Kurang	28	95.79	8.54	18.79	8.72	28	95.79	8.54	18.79	8.72
Baik	56	131.75	10.18	46.00	8.45	56	131.75	10.18	46.00	8.45
Buruk	25	69.32	10.18	10.20	9.79	25	69.32	10.18	10.20	9.79
		B=75								
Status Gizi	Jumlah Balita klaster	Berat Badan		Umur						
		Rerata	Standar Deviasi	Rerata	Standar Deviasi					
Lebih	10	157.60	36.29	46.90	20.71					
Kurang	29	95.28	8.82	18.76	8.56					
Baik	57	132.11	10.43	46.21	8.52					
Buruk	24	68.83	10.10	9.88	9.86					

Hasil pengklasteran status gizi berdasarkan berat badan dengan menggunakan metode *K-Means Bootstrap* B=25, B=50 dan B=75 menunjukkan hasil yang relatif sama dengan tingkat ketersebaran berat badan dan umur yang seragam. Jumlah anggota dalam kluster juga

menunjukkan kondisi yang relatif sama, dengan anggota kluster terbanyak pada kelompok status gizi baik. Kondisi ini menunjukkan jumlah replikasi *Bootstrap* tidak berbeda, sehingga digunakan replikasi 25, 50 maupun 75 akan menunjukkan kondisi sama.

### Ketepatan Hasil Klasifikasi Status Gizi Balita Menggunakan Metode K-Means dengan Klasifikasi Status Gizi Menurut WHO-2005

Tabel 4 menunjukkan tabulasi silang hasil analisis kluster dengan

metode *K-Means* dengan hasil pengelompokan dengan WHO-2005. Berdasarkan Tabel 4 diperoleh nilai *Total Accuracy Rate (TAR)* sebesar 0,9 dan nilai *Total Error Rate (TER)* sebesar 0,1.

Tabel 4. Tabulasi Silang Hasil Analisis Kluster Metode K-Means dengan WHO-2005

		WHO-2005				Total
		Lebih	Kurang	Baik	Buruk	
Metode K-Means	Lebih	8 6,7%	0 0%	2 1,7%	0 0%	10 8,3%
	Kurang	2 1,7%	24 20%	4 3,3%	1 0,8%	31 25,8%
	Baik	0 0%	0 0%	54 45%	0 0%	54 45%
	Buruk	1 0,8%	2 1,7%	0 0%	22 18,3%	25 20,8%
Total		11 9,2%	26 21,7%	60 50%	23 19,2%	120 100%

Berdasarkan informasi ada tabulasi silang Tabel 4 maka dihitung nilai *Total Accuracy Rate* dan *Total Error Rate* sebagai berikut :

$$\text{Total Accuracy Rate} = \frac{8 + 24 + 54 + 22}{120} = 0,9$$

$$\text{Total Error Rate} = \frac{2 + 2 + 4 + 1 + 1 + 2}{120} = 0,1$$

### Ketepatan Hasil Klasifikasi Status Gizi Balita Menggunakan Metode K-Means Bootstrap B=25, B=50 dan B=75 dengan Klasifikasi Status Gizi Menurut WHO-2005

Ketepatan hasil pengukuran antara metode *K-Means Bootstrap* dengan WHO-2005 diuraikan dalam Tabel 5.

Tabel 5. Perbandingan Ketepatan Pengelompokan Antar Metode K-Means Bootstrap

Indikator	Metode Pengelompokan		
	<i>K-Means Bootstrap</i> B=25	<i>K-Means Bootstrap</i> B=50	<i>K-Means Bootstrap</i> B=75
<i>Total Accuracy Rate (TAR)</i>	0,925	0,9417	0,9583
<i>Total Error Rate (TER)</i>	0,075	0,0583	0,0417

Tabel 5 menunjukkan bahwa ketepatan pengelompokan paling optimal menggunakan metode *K-Means* dengan *Bootstrap* 75 kali. Karena metode tersebut menunjukkan nilai TAR atau akurasi paling tinggi dengan nilai eror paling rendah.

## PEMBAHASAN

Analisis deskriptif status gizi balita di wilayah Puskesmas Ajung Kabupaten Jember tahun 2013, rerata BB balita 112,68 ons dan rerata usia balita 32,37 bulan. Berdasarkan standar WHO-2005 BB/U kategori status gizi

balita di wilayah Puskesmas Ajung Kabupaten Jember tahun 2013 termasuk normal.

Analisis klaster status gizi berdasarkan WHO-2005 menunjukkan rerata BB 130,65 ons. Rerata BB balitadengan analisis kluster metode *K-means* sebesar 169,50 ons, dengan metode *K-meansBootstrap* 25 kali menghasilkan rerata BB balita 164,80 ons, rerata BB balita 157,09 ons pada Bootstrap 50 kali dan rerata BB balita 157,60 ons pada Bootstrap 75 kali. Hal ini menunjukkan Rerata BB balita status gizi baik dengan Metode *K-means* lebih tinggi daripada rerata BB balita menurut WHO-2005.

Rerata BB balita gizi baik dengan metode *K-means* hampir sama dengan rerata BB balita dengan Metode *K-meansBootstrap* 25 kali, 50 kali maupun 75 kali. Hal ini sesuai dengan pendapat Efron dan Tibshirani (1993) bahwa replikasi *bootstrap* kecil ( $B=25$ ) sudah cukup informatif dan meningkat seiring dengan peningkatan replikasi *bootstrap*. Replikasi *bootstrap* ( $B=50$ ) sudah sangat cukup untuk memberikan estimasi yang akurat.

Status gizi buruk menurut klasifikasi WHO-2005 menunjukkan rerata BB balita sebesar 68,78 ons. Rerata BB balita dengan metode *K-means* sebesar 96,68 ons, dengan metode *K-meansBootstrap* 25 kali menghasilkan rerata BB balita sebesar 95,76 ons, rerata BB 95,79 ons pada Bootstrap 50 kali dan rerata BB 95,28 ons pada Bootstrap 75 kali. Hal ini menunjukkan rerata BB balita status gizi buruk dengan metode *K-means* lebih tinggi daripada rerata BB balita menurut WHO-2005. Rerata BB balita gizi buruk dengan metode *K-means* hampir sama dengan rerata BB balita dengan Metode *K-meansBootstrap* 25 kali, 50 kali maupun 75 kali dan kondisi

ini sesuai dengan pendapat Efron dan Tibshirani (1993).

Status gizi kurang menurut WHO-2005 menunjukkan rerata BB balita sebesar 93,23 ons. Rerata BB balitadengan metode *K-means* sebesar 131,69 ons, dengan metode *K-meansBootstrap* 25 kali menghasilkan rerata BB balita sebesar 131,75 ons, rerata BB 131,75 ons pada Bootstrap 50 kali dan rerata BB 132,11 ons pada Bootstrap 75 kali. Hal ini menunjukkan rerata BB balita status gizi kurang dengan metode *K-means* lebih tinggi daripada rerata BB balita menurut WHO-2005. Rerata BB balita gizi kurang dengan metode *K-means* hampir sama dengan rerata BB balita dengan Metode *K-meansBootstrap* 25 kali, 50 kali maupun 75 kali, kondisi ini sesuai dengan pendapat Efron dan Tibshirani (1993).

Status gizi lebih menurut WHO-2005 menunjukkan rerata BB balita sebesar 152,36 ons. Rerata BB balita dengan metode *K-means* sebesar 68,72 ons, dengan metode *K-means Bootstrap* 25 kali menghasilkan rerata BB balita sebesar 68,72 ons, rerata BB 69,32 ons pada Bootstrap 50 kali dan rerata BB 68,83 ons pada Bootstrap 75 kali. Hal ini menunjukkan rerata BB balita status gizi kurang dengan metode *K-means* lebih rendah daripada rerata BB balita menurut WHO-2005. Rerata BB balita gizi lebih dengan metode *K-means* hampir sama dengan rerata BB balita dengan Metode *K-meansBootstrap* 25 kali, 50 kali maupun 75 kali.

Ketepatan pengukuran status gizi balita dengan Metode *K-Means* dibandingkan dengan berdasarkan WHO-2005 sangat baik. Hal ini ditunjukkan oleh nilai TAR 0,9, dan TER 0,1. Ketepatan pengukuran dengan Metode *K-Means Bootstrap*  $B=25$  sangat baik yaitu TAR 0,925 dan TER 0,075. Ketepatan pengukuran

dengan Metode *K-Means Bootstrap* B=50 sangat baik TAR 0,9417, dan TER 0,0583. Ketepatan pengukuran dengan Metode *K-Means Bootstrap* B=75 sangat baik yaitu TAR 0,9583, dan TER 0,0417.

Nilai TAR semakin naik menunjukkan hasil semakin baik. Sedangkan nilai TER semakin naik menunjukkan hasil semakin tidak baik. Semakin besar resampling *bootstrap*, menunjukkan semakin besar nilai TAR dan semakin kecil nilai TER nya. Hal tersebut menunjukkan bahwa hasil semakin baik.

## SIMPULAN DAN SARAN

### Kesimpulan

Secara umum performance metode *K-Means* dan metode *K-Means* yang di-bootstrap berdasarkan hasil tabulasi silang dengan pengelompokan status gizi berdasarkan WHO-2005 menunjukkan sangat baik dalam mengukur ketepatan klasifikasi status gizi balita.

### Saran

Metode *K-Means* dan metode *K-Means* yang di-bootstrap dapat dipertimbangkan sebagai salah satu cara alternatif untuk menentukan klasifikasi status gizi balita.

## DAFTAR PUSTAKA

- Dwilityowati L. 2001. *Faktor penyebab gizi buruk pada balita*. Diakses 17 Maret 2013. <http://.blogspot.com/faktor-penyebab-gizi-buruk-pada-balita.html>
- Gudono. 2011. *Analisis Multivariat*, Edisi Pertama, Yogyakarta, BPFE
- Halim, S., Mallian, H. 2006. *Penggunaan Bootstrap Data Dependen untuk Membangun Selang Kepercayaan Pada Parameter Model Peramalan Data* Stationer, *Jurnal Teknik Industri*, Vol. 8, No. 1, hal. 54-60.
- Han K. 2007. *Data Mining Concepts and Techniques Second Edition*, San Fransisco, Elsevier Inc.
- Hill, T., Lewicki, P. 2007. *STATISTICS: Methods and Applications*, Statsoft, Tulsa, Diakses 12 April 2013, dari <http://statsoft/textbook/>
- Jelena, P. 2009. *Clustering & Bootstrapping*. University of Groningen, The Netherlands.
- Kuntoro., 2011. *Metode Statistik*, Edisi Revisi. Pustaka Melati, Surabaya.
- Sjahid, M., A., Adatul, M., Lalita P. 2010. *Bagging Regresi Logistik Ordinal Pada Status Gizi Balita*. *Media Statistika*, Vol. 3, No.2, Hal. 103-116
- Menteri Kesehatan RI .2010. *Standar Antropometri Penilaian Status Gizi Anak*, Dirjen Bina Gizi dan KIA, Jakarta.
- Payudho. 2008. *Analisis Cluster*. <http://.wordpress.com/2008/12/30/analisis-cluster/>. Diakses 15 Januari 2013
- Suparyanto. 2012. *Konsep Dasar Status Gizi Balita*. Diakses 15 Januari 2013. <http://beebidanindonesia.blogspot.com/2012/06/konsep-dasar-status-gizi-balita.html>
- Supranto J. 2010. *Analisis Multivariat, Arti & Interpretasi*. Jakarta, Rineka Cipta,.
- Zhu, W., Nancy, Z., Ning W. 2010. *Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS Implementations*, *Health Care and Life Science*