

PENGLASTERAN KABUPATEN/KOTA DI INDONESIA BERDASARKAN MASALAH GIZI BALITA DENGAN MENGGUNAKAN METODE TWO STEP CLUSTER DAN ENSEMBLE K-MODES

Cichi Chelchillya Candra, Ferra Yanuar, Dodi Devianto

Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas, Padang, Indonesia

E-mail: cichichelchillya@gmail.com, ferrayanuar@sci.unand.ac.id, ddevianto@sci.unand.ac.id

Abstrak

Pemerintah selalu berupaya untuk mengurangi prevalensi masalah gizi yang terjadi pada anak melalui berbagai program kesehatan. Namun karena berbagai masalah gizi tersebut memiliki penanganan yang berbeda, maka pemerintah suatu daerah perlu mengetahui masalah gizi anak apa yang dominan ditemui di daerahnya sehingga program yang akan dijalankan tepat. Tujuan dari penelitian ini adalah mengklasterkan kabupaten/kota di Indonesia berdasarkan masalah gizi balita dengan menggunakan metode Two Step Cluster dan Ensemble K-Modes serta mendeskripsikan karakteristik masalah gizi pada masing-masing klaster akhir yang terbentuk. Objek yang diamati pada penelitian ini terdiri dari 492 kabupaten/kota di Indonesia. Data yang digunakan yaitu data masalah gizi balita di semua kabupaten/kota di Indonesia. Hasil penelitian ini yaitu pengklasteran kabupaten/kota di Indonesia berdasarkan masalah gizi balita menggunakan metode two step cluster dan ensemble k-modes menghasilkan klaster yang berbeda. Metode ensemble k-modes lebih baik dalam mengklasterkan data gizi balita daripada metode two step cluster. Hal ini dapat dilihat dari nilai keragaman pengklasteran yang lebih kecil, yaitu 0,569015. Karakteristik klaster 1 menunjukkan bahwa sebagian besar kabupaten/kota dalam klaster ini terletak di wilayah Indonesia bagian barat, dengan total 229 kabupaten/kota.

Kata Kunci: Masalah gizi balita, Klaster kabupaten/kota, Metode Two Step Cluster, Ensemble K-Modes.

How to cite:	Cichi Chelchillya Candra, Ferra Yanuar, Dodi Devianto (2022) Pengklasteran Kabupaten/Kota di Indonesia Berdasarkan Masalah Gizi Balita Dengan Menggunakan Metode Two Step Cluster dan Ensemble K-Modes, (7) 10. Doi: 10.36418/syntax-literate.v7i10.13108
E-ISSN:	2548-1398
Published by:	Ridwan Institute

Pengklasteran Kabupaten/Kota di Indonesia Berdasarkan Masalah Gizi Balita Dengan Menggunakan Metode Two Step Cluster dan Ensemble K-Modes

Abstract

The government always strives to reduce the prevalence of nutritional problems in children through various health programs. However, due to the different handling of these nutritional problems, regional governments need to know the dominant nutritional problem found in their area in order to implement appropriate programs. The objective of this study is to cluster the districts/cities in Indonesia based on the nutritional problems of toddlers using the Two Step Cluster and Ensemble K-Modes methods, and to describe the characteristics of the nutritional problems in each final cluster formed. The objects observed in this study consist of 492 districts/cities in Indonesia. The data used is the nutritional problem data of toddlers in all districts/cities in Indonesia. The results of this study show that clustering the districts/cities in Indonesia based on the nutritional problems of toddlers using the Two Step Cluster and Ensemble K-Modes methods resulted in different clusters. The Ensemble K-Modes method is better at clustering nutritional data of toddlers compared to the Two Step Cluster method. This can be seen from the smaller diversity value of clustering, which is 0.569015. The characteristics of Cluster 1 indicate that the majority of districts/cities in this cluster are located in the western part of Indonesia, with a total of 229 districts/cities.

Keywords: *Nutritional problems in toddlers, District/city clusters, Two Step Cluster, Ensemble K-Modes methods.*

Pendahuluan

Usia di bawah lima tahun adalah usia emas dalam perkembangan seorang individu. Pada usia ini anak mengalami tumbuh kembang yang luar biasa baik dari segi fisik maupun kecerdasan. Untuk mengoptimalkan proses tumbuh kembang tersebut tentu dibutuhkan gizi yang berkualitas dalam jumlah yang cukup. Pentingnya pemenuhan gizi ini juga disebabkan karena masa balita merupakan masa yang rentan terhadap masalah gizi. Kasus kematian yang terjadi pada balita merupakan salah satu akibat dari gizi buruk. Gizi buruk dimulai dari penurunan berat badan ideal seorang anak sampai akhirnya berada jauh di bawah rata-rata.

Balita merupakan kelompok umur yang paling rentan mengalami malnutrisi. Jika terjadi kekurangan nutrisi pada tahap awal kehidupan dapat meningkatkan risiko infeksi, mortalitas, dan morbiditas bersamaan dengan penurunan perkembangan mental dan kognitif. Kekurangan nutrisi pada anak bisa bertahan lama dan melampaui masa kanak-kanak. Kekurangan nutrisi pada usia dini menurunkan prestasi pendidikan dan produktivitas tenaga kerja dan meningkatkan risiko penyakit kronis di usia lanjut.

Seribu hari pertama kehidupan merupakan masa kritis bagi perkembangan saraf anak. Kekurangan nutrisi merupakan kontributor utama gangguan perkembangan saraf anak, terutama di rangkaian sumber daya yang rendah. Anak-anak dengan nutrisi yang seimbang memiliki peluang yang lebih baik untuk berkembang. Sebaliknya, anak-anak yang dibesarkan dalam kondisi buruk, kekurangan gizi dan kelebihan gizi berisiko terhadap kesehatan dan hasil sosial yang negatif sepanjang perjalanan hidup mereka.

Penentuan status gizi merupakan suatu upaya yang dilakukan dalam rangka meningkatkan derajat kesehatan balita. Indeks yang biasa dipakai untuk menilai status gizi anak pada usia ini adalah berat badan terhadap umur (BB/U), tinggi badan terhadap umur (TB/U), dan berat badan terhadap tinggi badan (BB/TB). Ketiga indeks ini dapat menunjukkan masalah status gizi apa yang diderita oleh seorang balita. Salah satu contoh masalahnya adalah gizi kurang (*underweight*), pendek (*stunting*), kurus (*wasting*) dan gemuk (*overweight*).

Pemerintah selalu berupaya untuk mengurangi prevalensi masalah gizi yang terjadi pada anak melalui berbagai program kesehatan. Namun karena berbagai masalah gizi tersebut memiliki penanganan yang berbeda, maka pemerintah suatu daerah perlu mengetahui masalah gizi anak apa yang dominan ditemui di daerahnya sehingga program yang akan dijalankan tepat. Untuk memudahkan pemerintah dalam menyusun program untuk mengatasi masalah gizi balita di Indonesia, maka salah satu cara yang dilakukan adalah dengan mengelompokkan kabupaten/kota berdasarkan kemiripan masalah gizi yang ada di masing-masing kabupaten/kota. Dengan pengklasteran ini, maka kabupaten/kota yang mengalami masalah yang sama akan diberikan program yang sama juga. Sehingga dapat disusun program pemerintah yang lebih efektif, efisien dan lebih tepat sasaran.

Pengklasteran merupakan suatu proses untuk mengelompokkan objek-objek sehingga objek yang berada dalam suatu klaster memiliki karakteristik yang mirip. Teknik statistika yang digunakan untuk mengelompokkan objek-objek ke dalam klaster adalah analisis klaster. Terdapat banyak metode pengklasteran. Untuk mengelompokkan objek dalam jumlah besar, seperti untuk mengelompokkan banyaknya kabupaten/kota di Indonesia. Pada penelitian ini akan dibahas bagaimana pengklasteran kabupaten/kota di Indonesia berdasarkan masalah gizi balita dengan menggunakan metode *Two Step Cluster* dan Ensemble K-Modes.

Mongi dalam penelitiannya mengungkapkan bahwa analisis *Two Step Cluster* adalah metode pengklasteran yang dapat memberikan solusi untuk mengelompokkan suatu objek ke dalam kelompok-kelompok yang memiliki kemiripan (homogen) dengan permasalahan pada skala pengukuran, objek yang diteliti berukuran cukup besar dengan peubah yang berbeda yaitu kategorik dan numerik sehingga hasil akhir untuk penyelesaian dari metode tersebut dapat diketahui cluster optimal yang terbentuk. Andrew melakukan perbandingan antara algoritma *Two Step Cluster* dengan

Pengklasteran Kabupaten/Kota di Indonesia Berdasarkan Masalah Gizi Balita Dengan Menggunakan Metode Two Step Cluster dan Ensemble K-Modes

algoritma K-Prototype, diperoleh kesimpulan bahwa algoritma *Two Step Cluster* lebih baik karena nilai rasio keragaman algoritma *Two Step Cluster* lebih kecil dibandingkan dengan algoritma K-Prototypes. Saputra pernah menggunakan Ensemble Cluster dengan metode K-Modes untuk mengelompokkan aksesori jeruk dan membandingkan hasil pengklasterannya dengan metode numerik murni (Fuzzy C-Means) serta kategorik murni (K-Modes) sehingga diketahui bahwa metode Ensemble K-Modes mampu menerangkan dan menggabungkan kedua tipe karakter menjadi satu karakter data yaitu kategorik dengan hasil yang lebih baik. Larasati melakukan perbandingan antara algoritma Ensemble K-Modes dengan algoritma *Similarity Weight and Filter Method* (SWFM), diperoleh kesimpulan bahwa algoritma Ensemble K-Modes lebih baik karena nilai akurasi pengklasteran yang lebih besar dibandingkan dengan algoritma SWFM.

Berdasarkan latar belakang yang telah dipaparkan sebelumnya, masalah yang akan dibahas dalam penelitian ini adalah bagaimana mendapatkan solusi optimal pada pengklasteran kabupaten/kota di Indonesia berdasarkan masalah gizi balita dengan menggunakan metode *Two Step Cluster* dan Ensemble K-Modes serta bagaimana karakteristik masalah gizi yang dihadapi kabupaten/kota pada masing-masing klaster tersebut.

Pada penelitian ini data yang diambil merupakan data sekunder yaitu data masalah gizi pada kabupaten/kota di Indonesia tahun 2022 yang bersumber dari data Study Status Gizi Indonesia (SSGI) Tahun 2022 yang dilakukan oleh Kementerian Kesehatan. Metode klasterisasi yang dipilih *Two Step Cluster* dan Ensemble K-Modes.

Adapun tujuan dari penelitian ini adalah mengklasterkan kabupaten/kota di Indonesia berdasarkan masalah gizi balita dengan menggunakan metode *Two Step Cluster* dan Ensemble K-Modes serta mendeskripsikan karakteristik masalah gizi pada masing-masing klaster akhir yang terbentuk.

Metode Penelitian

Objek yang diamati pada penelitian ini terdiri dari 492 kabupaten/kota di Indonesia. Data yang digunakan yaitu data masalah gizi balita di semua kabupaten/kota di Indonesia. Data yang digunakan bersumber dari situs resmi Study Status Gizi Indonesia (SSGI) (www.kesmas.kemkes.go.id). Data yang diamati terdiri dari data numerik dan kategorik.

Peubah penelitian yang digunakan adalah sebagai berikut.

A. Peubah Numerik.

1. Persentase balita yang memiliki berat badan kurang/*underweight* (X_1).

Persentase balita *underweight* dapat ditentukan dengan rumus

$$X_1 = \frac{\text{jumlah balita } \textit{underweight}}{\text{jumlah balita}} \times 100\%.$$

2. Persentase balita yang memiliki tinggi badan jauh lebih pendek/*stunting* (X_2).

Persentase balita *stunting* dapat ditentukan dengan rumus

$$X_2 = \frac{\text{jumlah balita } \textit{stunting}}{\text{jumlah balita}} \times 100\%.$$

3. Persentase balita yang memiliki badan kurus/*wasting* (X_3).

Persentase balita *wasting* dapat ditentukan dengan rumus

$$X_3 = \frac{\text{jumlah balita } \textit{wasting}}{\text{jumlah balita}} \times 100\%.$$

4. Persentase balita yang memiliki berat badan berlebih/gemuk (X_4).

Persentase balita gemuk dapat ditentukan dengan rumus

$$X_4 = \frac{\text{jumlah balita gemuk}}{\text{jumlah balita}} \times 100\%.$$

B. Peubah Kategorik

1. Pembagian wilayah di Indonesia (X_5).

Pembagian wilayah di Indonesia terdiri dari 3 kategori dengan kode sebagai berikut,

$$X_5 = \begin{cases} 1, & \text{jika berada di wilayah Indonesia bagian barat,} \\ 2, & \text{jika berada di wilayah Indonesia bagian tengah,} \\ 3, & \text{jika berada di wilayah Indonesia bagian timur.} \end{cases}$$

2. Status wilayah di setiap provinsi (X_6).

Status wilayah di setiap provinsi terdiri dari 2 kategori dengan kode sebagai berikut,

$$X_6 = \begin{cases} 1, & \text{untuk kabupaten,} \\ 2, & \text{untuk kota.} \end{cases}$$

Tahapan analisis yang dilakukan pada penelitian ini adalah sebagai berikut.

1. Mendeskripsikan masalah gizi balita yang dialami setiap kabupaten/kota di Indonesia. Pendeskripsian data dilakukan dengan menggunakan statistika deskriptif untuk data numerik dan diagram lingkaran untuk data kategorik.
2. Mengklasterkan kabupaten/kota di Indonesia berdasarkan masalah gizi balita.
3. Mendeskripsikan karakteristik dari masing-masing klaster yang dihasilkan.

Hasil dan Pembahasan

A. Pengklasteran Kabupaten/kota di Indonesia Berdasarkan Masalah Gizi Balita menggunakan Ensemble K-Modes.

1. Pengklasteran dengan Fuzzy C-Means

Pengklasteran data numerik pada penelitian ini dilakukan menggunakan metode Fuzzy C-Means (FCM) dan data kategorik menggunakan metode K-Modes. Hasil dari kedua pengklasteran akan menghasilkan data bertipe

Pengklasteran Kabupaten/Kota di Indonesia Berdasarkan Masalah Gizi Balita Dengan Menggunakan Metode Two Step Cluster dan Ensemble K-Modes

kategorik. Selanjutnya akan dilakukan proses ensemble yaitu dengan menggabungkan hasil pengklasteran kedua metode menggunakan algoritma K-Modes.

Dalam algoritma Fuzzy C-Means yang pertama kali dilakukan adalah menentukan jumlah kluster awal (c) yang diinginkan yaitu $2 \leq c \leq 9$, pangkat pembobot yang digunakan adalah 2 ($w = 2$), error terkecil yang diharapkan (ξ) adalah 10^{-5} . Selanjutnya untuk nilai $c = 2$ akan dibangkitkan bilangan random U_{ik} , $i = 1, 2, \dots, 492$, $k = 1, 2$, sebagai elemen-elemen matriks partisi awal U. Matriks partisi awal U (U_0) yang terbentuk adalah :

Tabel 1
Derajat Keanggotaan Awal

No	Kabupaten/Kota	c_1	c_2
1	Kab. Gayo Lues	0.788079736	0.211920264
2	Kab. Aceh Tengah	0.727725522	0.272274478
3	Kota Banda Aceh	0.854593787	0.145406213
4	Kota Sabang	0.718183358	0.281816642
5	Kab. Aceh Barat	0.954993607	0.045006393
⋮	⋮	⋮	⋮
488	Kab. Asmat	0.658693500	0.341306500
489	Kab. Supiori	0.782094300	0.217905700
490	Kab. Waropen	0.664335659	0.335664341
491	Kab. Dogiyai	0.854899116	0.145100884
492	Kota Batu	0.368486230	0.631513770

Data yang digunakan adalah data masalah gizi balita tahun 2022 yang terdiri dari 492 data.

Tabel 2
Data yang digunakan (x_{ij})

No	Kabupaten/kota	x_1	x_1	x_3	x_4
1	Kab. Gayo Lues	17,1	34,6	5,4	2,2
2	Kab. Aceh Tengah	15,9	32	3,6	2,1
3	Kota Banda Aceh	21,4	25,1	12,4	3,3
4	Kota Sabang	20,4	23,4	11	3,3
5	Kab. Aceh Barat	22,5	30,4	10,5	1,4
⋮	⋮	⋮	⋮	⋮	⋮
488	Kab. Asmat	36,3	54,4	14,1	4,8

489	Kab. Supiori	26	40,2	7,9	2
490	Kab. Waropen	22,6	22,2	13,8	2,1
491	Kab. Dogiyai	24,1	35,1	12,8	2,7
492	Kota Batu	12,6	25,2	3,4	5

Dengan cara yang sama, akan diperoleh derajat keanggotaan baru setiap objek yang hasilnya akan disajikan pada Tabel 3 berikut:

Tabel 3
Derajat Keanggotaan Awal Akhir Iterasi 1

No	Kabupaten/Kota	c_1	c_2
1	Kab. Gayo Lues	0.64727058	0.35272942
2	Kab. Aceh Tengah	0.62902683	0.37097317
3	Kota Banda Aceh	0.75101496	0.24898504
4	Kota Sabang	0.41287766	0.58712234
5	Kab. Aceh Barat	0.74219859	0.25780141
⋮	⋮	⋮	⋮
488	Kab. Asmat	0.43382344	0.56617656
489	Kab. Supiori	0.62448225	0.37551775
490	Kab. Waropen	0.38563794	0.61436206
491	Kab. Dogiyai	0.66234558	0.33765442
492	Kota Batu	0.58004614	0.41995386

Akan diperiksa kondisi berhenti $|P_t - P_{t-1}| < \xi = 10^{-5}$. Karena $|P_1 - P_0| = 579642,98653 - 0 = 579642,98653 > \xi = 10^{-5}$ maka diulangi lagi dari langkah keempat sampai pada iterasi terakhir dimana nilai $|P_t - P_{t-1}| < \xi$. Untuk kasus pada penelitian ini, proses iterasi baru akan berhenti pada iterasi ke-41 dengan nilai:

$|P_{41} - P_{40}| = 38645,585764 - 38645,585758 = 0,6 \times 10^{-5} < \xi = 10^{-5}$
dengan nilai pusat klaster pada iterasi ke-41 adalah sebagai berikut:

Tabel 4
Hasil Pusat Klaster Pada Iterasi 41

Cluster	x_1	x_1	x_3	x_4
1	23.04757	30.26641	9.919163	2.939091
2	14.13559	17.93949	6.880213	3.710104

serta nilai U partisi pada iterasi ke-41 dengan derajat keanggotaan baru adalah sebagai berikut:

Tabel 5
Derajat Keanggotaan Akhir Iterasi 41

No	Kabupaten/Kota	c_1	c_2
1	Kab. Gayo Lues	0.794721090	0.205278910
2	Kab. Aceh Tengah	0.693326714	0.306673286
3	Kota Banda Aceh	0.790510914	0.209489086
4	Kota Sabang	0.608520408	0.391479592
5	Kab. Aceh Barat	0.987742332	0.012257668
⋮	⋮	⋮	⋮
488	Kab. Asmat	0.706363967	0.293636033
489	Kab. Supiori	0.850717001	0.149282999
490	Kab. Waropen	0.633846008	0.366153992
491	Kab. Dogiyai	0.929046576	0.070953424
492	Kota Batu	0.274937675	0.725062325

Selanjutnya dengan cara yang sama, pengklasteran menggunakan fuzzy c-means akan dilakukan untuk jumlah klaster 3, 4, 5, 6, 7, 8 dan 9. Pada jumlah klaster 3, proses iterasi akan berhenti pada iterasi ke-91 dengan nilai fungsi objektif minimum yang didapatkan adalah 32564,965434 dengan derajat keanggotaan yang dapat dilihat pada lampiran 2. Pada saat jumlah klaster 4, proses iterasi akan berhenti pada iterasi ke-101 dengan nilai fungsi objektif 28765,456876 dan derajat keanggotaan dapat dilihat pada lampiran 3. Selanjutnya pada jumlah klaster 5, proses iterasi berhenti pada iterasi ke-216 dengan fungsi objektif 21456,974563 dan derajat keanggotaan yang dapat dilihat pada lampiran 4. Pada jumlah klaster 6, proses iterasi berhenti pada iterasi ke-1000 dengan fungsi objektif 16845,467587 dan nilai derajat keanggotaan dapat dilihat pada lampiran 5. Pada jumlah klaster 7, proses iterasi akan berhenti pada iterasi ke-538 dengan fungsi objektif 13871,574324 dengan derajat keanggotaan yang dapat dilihat pada lampiran 6. Pada jumlah klaster 8, proses iterasi akan berhenti pada iterasi ke-855 dengan nilai fungsi objektif 11654,764432 dan derajat keanggotaan dapat dilihat di lampiran 7. Serta pada jumlah klaster 9, proses iterasi akan berhenti pada iterasi ke-1000 dengan nilai fungsi objektif 9654.326747 dan derajat keanggotaan dapat dilihat di lampiran 8.

Jumlah klaster optimum ditentukan dengan melihat indeks validitas. Indeks validitas yang digunakan adalah Partition Entropy. Jumlah klaster optimum apabila memiliki nilai indeks Partition Entropy yang terkecil.

Indeks validitas PE pada jumlah klaster 3, 4, 5, 6, 7, 8 dan 9 dihitung dengan cara yang sama dan akan disajikan pada Tabel 5 berikut:

Tabel 5
Penentuan Klater Optimum

Klaster	Iterasi	Fungsi Objektif	PE
2	41	38645,585764	1.3872724
3	91	32564,965434	1.06606669
4	101	28765,456876	0.8585695
5	216	21456,974563	1.018743
6	1000	16845,467587	0.6606669
7	538	13871,574324	1.271923
8	855	11654,764432	1.384695
9	1000	9654.326747	1.406311

Berdasarkan Tabel 5, dapat dilihat bahwa jumlah klaster yang optimum diberikan ketika jumlah klaster = 6 dengan nilai Partition Entropy lebih kecil dibanding klaster lain. Banyak Iterasi yang diperoleh adalah 1000 dengan nilai fungsi objektifnya 16845,467587. Anggota klaster Optimum ditentukan dari derajat keanggotaan klaster pada iterasi terakhir. Berikut diberikan derajat keanggotaan dari dua klaster provinsi-provinsi di Indonesia:

Tabel 6
Derajat Keanggotaan Klaster Optimum (1)

No	Kabupaten/kota	Klaster 1	Klaster 2	Klaster 3
1	Kab. Gayo Lues	0.19801758	0.1793566	0.07890313
2	Kab. Aceh Tengah	0.31695342	0.2010471	0.10858627
3	Kota Banda Aceh	0.08481329	0.7372701	0.04827765
4	Kota Sabang	0.12260231	0.7162455	0.07246862
5	Kab. Aceh Barat	0.08044206	0.2691019	0.03656757
⋮	⋮	⋮	⋮	⋮
488	Kab. Asmat	0.11894761	0.14876940	0.09081180
489	Kab. Supiori	0.04090668	0.05846294	0.02415608
490	Kab. Waropen	0.14232067	0.48773458	0.12861442
491	Kab. Dogiyai	0.14232067	0.07382402	0.02109069
492	Kota Batu	0.52378068	0.11517213	0.18852867

Derajat keanggotaan terbesar menunjukkan bahwa kecenderungan tertinggi untuk masuk menjadi anggota klaster tersebut.

2. Pengklausteran dengan K-Modes

Metode K-Modes akan digunakan untuk mengelompokkan variabel dengan skala data kategorik. Metode ini menempatkan sebuah objek pengamatan

Pengklastran Kabupaten/Kota di Indonesia Berdasarkan Masalah Gizi Balita Dengan Menggunakan Metode Two Step Cluster dan Ensemble K-Modes

tepat menjadi anggota suatu klaster. Ukuran kemiripan pada metode ini diukur menggunakan nilai modus. Dalam penelitian ini jumlah klaster yang akan dibentuk yaitu 2, 3, 4, 5 dan 6 cluster. Penentuan centroid awal pada penelitian ini yaitu dengan mengambil data yang tersedia secara random atau acak seperti yang dapat dilihat pada Tabel 7 berikut ini:

Tabel 7
Tabel Penentuan Centroid Awal

No	KABUPATEN/KOTA	Daerah	Waktu
1	Kab. Gayo Lues	1	1
2	Kab. Aceh Tengah	1	1
3	Kota Banda Aceh	2	1
4	Kota Sabang	2	1
5	Kab. Aceh Barat	1	1
⋮	⋮	⋮	⋮
488	Kab. Asmat	1	3
489	Kab. Supiori	1	3
490	Kab. Waropen	1	3
491	Kab. Dogiyai	1	3
492	Kota Batu	2	1

Hasil dari perhitungan jarak antara data dengan centroid dapat dilihat seperti pada tabel 8 berikut ini:

Tabel 8
Tabel Jarak Data dengan Centroid

No	Kabupaten/Kota	Jarak Antar Cluster	
		Cluster 1	Cluster 2
1	Kab. Gayo Lues	1	0
2	Kab. Aceh Tengah	1	0
3	Kota Banda Aceh	0	1
4	Kota Sabang	0	1
5	Kab. Aceh Barat	1	0
⋮	⋮	⋮	⋮
488	Kab. Asmat	2	1
489	Kab. Supiori	2	1
490	Kab. Waropen	2	1
491	Kab. Dogiyai	2	1

492	Kota Batu	0	1
-----	-----------	---	---

Pengelompokkan data cluster didapatkan dengan melihat jarak terpendek. Setelah itu klusterkan data berdasarkan clusternya, dapat dilihat pada tabel 9 berikut ini:

Tabel 9
Memilih Jarak Terpendek

No	Kabupaten/Kota	Jarak Antar Cluster		Cluster
		C1	C2	
1	Kab. Gayo Lues	1	0	2
2	Kab. Aceh Tengah	1	0	2
3	Kota Banda Aceh	0	1	1
4	Kota Sabang	0	1	1
5	Kab. Aceh Barat	1	0	1
⋮	⋮	⋮	⋮	⋮
488	Kab. Asmat	2	1	2
489	Kab. Supiori	2	1	2
490	Kab. Waropen	2	1	2
491	Kab. Dogiyai	2	1	2
492	Kota Batu	0	1	1

Mencari Modus (nilai yang sering muncul) untuk menentukan pusat centroid baru. Centroid baru didapatkan dengan melihat data yang sering muncul (modus), dimana pada anggota klaster tiap cluster dibandingkan nilai apa saja yang sering muncul per atribut. Jika terdapat nilai yang sama maka diambil mana yang paling pertama ditemui. Jadi nilai centroid baru yang didapatkan pada iterasi 1 adalah seperti tabel 10 berikut ini:

Tabel 10
Data Centroid Baru

Cluster	Daerah	Waktu
1	2	1
2	1	3

Setelah Centroid baru didapatkan proses kembali ke langkah nomor 3 yaitu menghitung ketidak miripan setiap centroid dengan menggunakan centroid baru untuk lanjut ke iterasi selanjutnya. Proses akan berhenti jika anggota pada iterasi sebelumnya bernilai sama (tidak ada perubahan nilai sama sekali). Tabel

Pengklasteran Kabupaten/Kota di Indonesia Berdasarkan Masalah Gizi Balita Dengan Menggunakan Metode Two Step Cluster dan Ensemble K-Modes

semua pola yang didapatkan dari proses Data Mining dengan menggunakan algoritma K-Modes beserta nama untuk masing-masing cluster dapat dilihat pada tabel 11 dibawah ini:

Tabel 11
Memilih Jarak Terpendek

No	Kabupaten/Kota	Jarak Antar Cluster		Cluster
		C1	C2	
1	Kab. Gayo Lues	1	0	2
2	Kab. Aceh Tengah	0	1	1
3	Kota Banda Aceh	0	1	1
4	Kota Sabang	1	2	1
5	Kab. Aceh Barat	0	1	1
⋮	⋮	⋮	⋮	⋮
488	Kab. Asmat	2	1	2
489	Kab. Supiori	1	1	1
490	Kab. Waropen	0	1	1
491	Kab. Dogiyai	2	2	1
492	Kota Batu	0	1	1

Selanjutnya dengan cara yang sama, pengklasteran menggunakan K-Modes akan dilakukan untuk jumlah kluster 3, 4, 5 dan 6.

Tabel 12
Perbandingan Nilai Akurasi Pengklasteran .

Banyak Cluster	Nilai Akurasi
2	0,84045
3	0,85163
4	0,97561
5	0,98374
6	1

Hasil perbandingan nilai akurasi pada Tabel 12 diperoleh hasil bahwa pengklasteran K-Modes dengan jumlah kluster sebanyak 6 memiliki nilai akurasi yang lebih besar dibandingkan dengan jumlah kluster lainnya. Pengklasteran tersebut membagi kabupaten/kota di Indonesia menjadi 6 kluster dengan masing-masing anggota sebanyak 222 daerah, 129 daerah, 50 daerah, 39 daerah, 16 daerah dan 36 daerah. Distribusi anggota kabupaten/kota pada masing-masing kluster yang terbentuk dapat dilihat pada Tabel 13.

Tabel 13
Distribusi Kabupaten/Kota dalam Klaster

Klaster	N
1	229
2	129
3	50
4	39
5	45
Total	492

Setelah terbentuk klaster, selanjutnya akan diambil rata-rata dari masing-masing peubah.

Tabel 14
Nilai Rata-Rata Peubah berdasarkan Klaster Optimal yang Terbentuk.

Peubah	Klaster					Rata-Rata
	1	2	3	4	5	Keseluruhan
<i>Underweight</i>	21,428	21,84109	14,5131	26,55128	15,73778	25,01781
<i>Stunting</i>	25,908	26,75969	19,0804	27,79778	34,31538	33,4653
<i>Wasting</i>	7,16507	12,41282	5,41778	9,50388	9,46	10,98989
<i>Overweight</i>	4,09913	2,74574	2,53333	2,928	3,04	3,836549

Dari Tabel 14 dapat dilihat bahwa untuk persentase balita *underweight*, klaster 4 memiliki nilai rata-rata tertinggi dari klaster lainnya dan nilainya berada di atas rata-rata keseluruhan klaster. Untuk persentase balita *stunting*, klaster 5 memiliki nilai rata-rata tertinggi dari klaster lainnya dan nilainya berada di atas rata-rata keseluruhan klaster. Untuk persentase balita *wasting*, klaster 2 memiliki nilai rata-rata tertinggi dari klaster lainnya dan nilainya berada di atas rata-rata keseluruhan klaster. Kemudian, untuk persentase balita *overweight*, klaster 1 memiliki nilai rata-rata tertinggi dari klaster lainnya dan nilainya berada di atas rata-rata keseluruhan klaster. Sedangkan klaster 3 merupakan klaster dengan nilai rata-rata terendah dari klaster lainnya dan nilainya berada di bawah rata-rata keseluruhan klaster. Sebaran kabupaten/kota pada masing-masing klaster untuk wilayah Indonesia bagian barat, tengah dan timur, serta status wilayah di setiap provinsi dapat dilihat pada Tabel berikut.

Tabel 15
Sebaran Klaster pada Pembagian Wilayah di Indonesia dan Status

Wilayah di setiap Provinsi

Klaster	Pembagian Wilayah di Indonesia			Status Wilayah di setiap Provinsi	
	bagian Barat	bagian Tengah	bagian Timur	Kabupaten	Kota
1	180	31	18	157	72
2	11	97	21	109	20
3	50	0	0	50	0
4	20	0	19	39	0
5	40	0	5	42	3
Total	301	128	63	397	95

Dari Tabel 15 dapat dilihat bahwa pada kaster 1 sebagian besar daerah berada di wilayah Indonesia bagian barat dengan 157 wilayah berstatus kabupaten dan 72 wilayah berstatus kota. Pada kaster 2 sebagian besar daerah berada di wilayah Indonesia bagian tengah dengan 109 wilayah berstatus kabupaten dan 20 wilayah berstatus kota. Pada kaster 3 semua daerah berada di wilayah Indonesia bagian barat dengan 50 wilayah berstatus kabupaten. Pada kaster 4 semua daerah berada di wilayah Indonesia bagian barat dan timur dengan 39 wilayah berstatus kabupaten. Kemudian pada kaster 5 sebagian besar daerah berada di wilayah Indonesia bagian barat dengan 42 wilayah berstatus kabupaten dan 3 wilayah berstatus kota.

3. Perbandingan antara Two Step Cluster (TSC) dan Ensemble K-Modes

Hasil pengklasteran terbaik diperoleh dengan membandingkan nilai rasio keragaman dalam klaster (S_w) dan keragaman antar klaster (S_B) menggunakan metode *Two Step Cluster* dan metode Ensemble K-Modes. Berikut merupakan nilai perbandingan untuk kedua metode pengklasteran data campuran.

Tabel 16
Perbandingan Nilai Akurasi Kedua Metode

	<i>Two Step Cluster</i>		Ensemble K-Modes	
	Kontinu	Kategorik	Kontinu	Kategorik
S_w	53413,51	117,3314	37716,09	239,8779
S_B	129142,8	146,3764	143681,1	273,9799
Rasio (S_w/S_B)	0,4136	0,801574	0,262499	0,875531
Keragaman	0,607587		0,569015	

Berdasarkan Tabel 16 metode pengklasteran data campuran Ensemble K-Modes dengan jumlah kluster optimum sebanyak 5 memiliki nilai keragaman terkecil yaitu sebesar 0,569015. Sehingga dapat diketahui bahwa metode pengklasteran data campuran terbaik untuk mengklasterkan kabupaten/kota di Indonesia berdasarkan masalah gizi balita adalah dengan menggunakan metode Ensemble K-Modes.

Kesimpulan

Kesimpulan yang diperoleh dalam penelitian ini adalah : (1) Pengklasteran kabupaten/kota di Indonesia berdasarkan masalah gizi balita dengan menggunakan metode *two step cluster* terdiri dari 2 kluster dengan nilai keragaman 0,607587, sedangkan pengklasteran dengan ensemble k-modes menghasilkan 5 kluster optimum dengan nilai keragaman 0,569015. (2) Metode pengklasteran data campuran terbaik menunjukkan bahwa ensemble k-modes menghasilkan pengklasteran yang lebih baik dibandingkan dengan metode *two step cluster* karena nilai keragaman pengklasteran yang dimiliki ensemble k-modes lebih kecil yaitu 0,569015. Dengan demikian pengklasteran kabupaten/kota di Indonesia berdasarkan masalah gizi balita menghasilkan 5 kluster. (3) Berdasarkan karakteristik masing-masing kluster, diketahui bahwa kluster 1 yang terdiri dari 229 kabupaten/kota dengan sebagian besar daerah berada di wilayah Indonesia bagian barat. Kluster 1 merupakan kluster dengan masalah gizi *overweight* tertinggi dari kluster lainnya. Kluster 2 terdiri dari 129 kabupaten/kota dengan sebagian besar daerah berada di wilayah Indonesia bagian tengah. Kluster 2 merupakan kluster dengan masalah gizi *wasting* tertinggi dari kluster lainnya. Kluster 3 terdiri dari 50 kabupaten dengan semua daerah berada di wilayah Indonesia bagian barat. Kluster 3 merupakan kluster dengan semua masalah gizi terendah dari kluster lainnya. Kluster 4 terdiri dari 39 kabupaten dengan semua daerah berada di wilayah Indonesia bagian barat dan timur. Kluster 4 merupakan kluster dengan masalah gizi *underweight* tertinggi dari kluster lainnya. Kluster 5 terdiri dari 45 kabupaten/kota dengan sebagian besar daerah berada di wilayah Indonesia bagian barat. Kluster 5 merupakan kluster dengan masalah gizi *stunting* tertinggi dari kluster lainnya.

BIBLIOGRAFI

- Anonimous. 2001. *The SPSS TwoStep Cluster Component*. Technical Report SPSS Inc. Chicago.
- Bacher, J., K. Wenzig and M. Vogler. 2004. *SPSS TwoStep Cluster : A First Evaluation*. FAU, Jerman.
- Bunkers, M., Miller, J., & DeGaetano, A. (1996). Definition of climate regions in the northern plains using an objective cluster modification technique. *Journal of Climate*, 130-146.
- Chan, Y. H. 2005. Biostatistik 304 – ClusterAnalysis. *Medical Journal*. **46** : 153.
- Chiu, T., D. Fang, J. Chen, Y. Wang, and C. Jeris. 2001. A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment. *In Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2001*. Pp.263-268.
- Endris N, Asefa H, Dube L. 2017. *Prevalence of Malnutrition and Associated Factors among Children in Rural Ethiopia*. 2017:6587853.
- Hair, J., Black, W., Babin, J., & Anderson, R. (2010). *Multivariate Data Analysis "7th ed"*. Pearson Prentice Hall, New Jersey(US).
- Huang, Z. (1997). Clustering large data sets with mixed numeric and categorical values. *In Proceedings of the 1st pacific-asia conference on knowledge discovery and data mining,(PAKDD)* (pp. 21-34).
- John CC, Black MM, Nelson 3rd CA. 2017. Neurodevelopment: The Impact of Nutrition and Inflammation During Early to Middle Childhood in Low Resource Settings. *Pediatrics* 2017;139 (suppl1):S59– S71.
- Johnson RA, Wichern DW. 2007. *Applied Multivariate Statistical Analysis*. Prentice Hall, New Jersey.
- Kader, G., & Perry, M. (2007). Variability for categorical variables. *Journal of Statistics*, 15(2): 1-16.
- Kementrian Kesehatan RI. 2020. *Standar Antropometri Anak*. Badan Litbangkes Depkes RI, Jakarta.

- Kementrian Kesehatan RI. 2018. *Profil Kesehatan Indonesia 2018*. Kemenkes RI, Jakarta.
- Kementrian Kesehatan RI. 2022. *Buku Saku Hasil Studi Status Gizi Indonesia (SSGI) Tahun 2022*. Badan Litbangkes Kemenkes RI, Jakarta.
- Pratiwi, R.H. 2015. Faktor-faktor yang berhubungan dengan Berat Kurang (*Underweight*) pada Balita di Perkotaan dan Perdesaan Indonesia berdasarkan Data Riskesdas tahun 2013. *Jurnal Kesehatan Masyarakat*. 3(4): 127-137.
- Okada, T. (1999). Sum of Squares Decomposition for Categorical Data. *Kwansei Gakuin Studies in Computer Science*, 14:1-6.
- Putri, W. 2007. Analisis Gerombol Menggunakan Metode *Two Step Cluster*. *Jurnal Statistika dan Komputasi*. 12 : 18-23.
- TNP2K. 2017. *100 Kabupaten/Kota Prioritas untuk Intervensi Anak Kerdil (Stunting)*. Sekretariat Wakil Presiden RI, Jakarta Pusat.
- UNICEF. 2019. *Stunting Pada Balita*. Pusat Promosi Kesehatan, Kementerian Kesehatan RI, Jakarta.
- WHO. 2017. *Reducing Stunting In Children: Equity Considerations for Achieving the Global Nutrition Targets 2025*. Geneva: WHO.
- Zhang, T, R. Ramakrishnon and M. Livny. 1996. BIRCH: An Efficient Data Clustering Method for Very Large Databases. *Proceeding of the ACM SIGMOD Conference on Management of Data*. Pp.103-114.
- Sumanto. (2014). *Statistika Deskriptif (Pertama ed.)*. Yogyakarta : Center of Academic Publishing Service .
- Hoppner, F., Klawonn, F., Kruse, R., & Runkler, T. (1999). *Fuzzy Cluster Analysis*. Wiley.
- Kusumadewi, S., & Hartati, S. (2006). *Neuro Fuzzy: Integrasi Sistem Fuzzy & Jaringan syaraf* . Yogyakarta: Graha ilmu.
- He, Z., Xu, X., & Deng, S. (2005). Clustering Mixed Numeric and Categorical Data: A Cluster Ensemble Approach. Harbin Institute of Technology, Department of Computer Science and Engineering.
- Yoon, H., Ahn, S., Lee, S., Cho, S., & Kim, J. (2006). Heterogeneous Clustering Ensemble Method for Combining Different Cluster Results. *BioDM 2006. Lecture Notes in Computer Science*, 3916.

Pengklastran Kabupaten/Kota di Indonesia Berdasarkan Masalah Gizi Balita Dengan Menggunakan Metode Two Step Cluster dan Ensemble K-Modes

- Mongi, C. E. 2015. Penggunaan Analisis Two Step Clustering Untuk Data Campuran. *Jurnal de Cartesian (JdC)*. 4(1):9-19.
- Andrew, D. M. (2018). Penggerombolan Desa/Kelurahan Berdasarkan Indikator Kemiskinan dengan menerapkan Algoritma Two Step Cluster dan K-Prototypes. *Indonesian Journal of Statistics and Its Applications*. 2(2): 63-76.
- Saputra, C. W. (2016). Pengklastran Akses Jeruk Persilangan Berdasarkan Karakter Kuantitatif dan Kualitatif Menggunakan Fuzzy C-Menas dan K-Modes. Skripsi, Institut Teknologi Sepuluh Nopember, Statistika.
- Larasati, D.P (2018). Penerapan Metode Data Campuran Ensemble K-Modes dan *Similarity Weight and Filter Method (SWFM)* pada Pengklastran Kabupaten/Kota di Jawa Timur berdasarkan Indikator Daerah Tertinggal. Skripsi, Institut Teknologi Sepuluh Nopember, Statistika.
- Madhulatha, T.S., 2012. An Overview On Clustering Methods. *IOSR Journal of Engineering*, II (4), pp.719-725
- Kodinariya, Trupti M. & Makwana, Prashant R., (2013). Review on determining number of cluster in K-Means Clustering. *International Journal of Advance Research in Computer Science and Management Studies*, I (6), pp. 90-95
- Bholowalia, Purnima & Kumar, Arvind, 2014. EBK-Means: A Clustering Techniques based on Elbow Method and K-Means in WSN. *International Journal of Computer Application (0975-8887)*, IX (105), pp. 17-24
- Irwanto, et. al (2012). Optimasi Kinerja Algoritma Klasterisasi K-Means untuk kuantisasi Warna Citra. *Jurnal Teknik ITS*, I (1), pp.197-202.
- Kusumadewi dan Purnomo .2004. Aplikasi Logika Fuzzy untuk Pendukung Keputusan. Yogyakarta : Graha Ilmu.
- Klawonn dand Höppner, (2001), "What is Fuzzy about Fuzzy Clustering? Understanding and Improving the Concept of the Fuzzier". *Science Journal*, pp.254-264
- Huang, Z., & Ng, M. K. (1999). A Fuzzy K-Modes Algorithm for Clustering Categorical Data. *IEEE Transactions On Fuzzy Systems*, 7(4).

Copyright holder:
Cichi Chelchillya Candra (2022)

First publication right:
Syntax Literate: Jurnal Ilmiah Indonesia

This article is licensed under:

