

# Pemodelan Indikator Kesehatan Anak Berdasarkan Provinsi Tahun 2022 Menggunakan Analisis Diskriminan dan Analisis *Cluster* dengan Algoritma K-Means

Adinda Prilly Cindana<sup>1\*</sup> dan Abraham M. P. Sumarsono<sup>2</sup>

<sup>1</sup>Author Affiliation: Departemen Statistika, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

<sup>2</sup>Author Affiliation: Departemen Statistika, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

\*Corresponding author: author@email.com

Received: 28 Juni 2024

Revised: 28 Juni 2024

Accepted: 28 Juni 2024

**ABSTRAK** – Status kesehatan anak adalah indikator kunci dalam mengevaluasi kemajuan pembangunan sebuah negara. Penelitian ini mengkaji indikator kesehatan anak menurut provinsi yang bersumber dari Profil Kesehatan Ibu dan Anak Tahun 2022 menggunakan analisis kluster K-Means dan analisis diskriminan. Penelitian ini bertujuan untuk mengetahui kelompok-kelompok provinsi di Indonesia berdasarkan indikator kesehatan anak menggunakan metode K-Means dan mengetahui indikator kesehatan anak yang berpengaruh signifikan terhadap pengelompokan provinsi di Indonesia. Hasil analisis kluster didapat 3 kluster provinsi yang memiliki nilai indikator kesehatan dominan bagus, sedang, dan kurang bagus. Analisis diskriminan dilakukan dengan menjadikan hasil analisis kluster sebagai variabel dependen dan menghasilkan 2 fungsi diskriminan yang berisi tiga dari 7 indikator kesehatan anak. Fungsi diskriminan memiliki ketepatan klasifikasi sebesar 94.1%.

**Keywords**– Analisis Diskriminan, Analisis Kluster, Indikator Kesehatan Anak

## I. PENDAHULUAN

### A. Latar Belakang

Segala upaya dalam meningkatkan tingkat kesehatan ibu dan anak merupakan investasi bagi suatu negara. Anak dengan status kesehatan yang baik akan memberikan *rate of return* yang lebih tinggi dalam jangka waktu pendek maupun panjang. Contoh dari *rate of return* ini adalah kemampuan belajar yang baik biasanya berasal dari anak yang sehat (Beattie, Yates, & Noble, 2016). Sementara itu, dalam jangka panjang, anak akan lebih produktif dan sehat saat dewasa sehingga mampu mendorong pertumbuhan ekonomi. Anak yang sehat akan menekan tingkat kematian sehingga proporsi populasi usia kerja di masa depan dapat meningkat,

Tidak hanya itu, pertumbuhan ekonomi yang terjadi di Asia disebabkan oleh perubahan demografi dan kesehatan yang diakibatkan karena penurunan kematian bayi dan anak, tingkat kelahiran, dan peningkatan dalam kesehatan reproduksi (Bloom dan Williamson, 1998).

Status kesehatan anak adalah indikator kunci dalam mengevaluasi kemajuan pembangunan sebuah negara. Salah satu indikator yang dapat menggambarkan kondisi kesehatan masyarakat di suatu wilayah merupakan keluhan kesehatan. Keluhan kesehatan merupakan keadaan seseorang yang mengalami gangguan kesehatan atau kejiwaan, baik karena gangguan/penyakit yang sering dialami, seperti panas, batuk, pilek, diare, sakit kepala, maupun kena penyakit akut, penyakit kronis, karena kecelakaan, kriminalitas, atau keluhan kesehatan lainnya.

Dalam mengatasi keluhan kesehatan, terdapat beberapa upaya kesehatan yang terdiri dari mengobati sendiri, rawat jalan, dan rawat inap. Di tahun 2022, sekitar 82 dari 100 anak yang memiliki keluhan kesehatan pernah mengobati diri sendiri dalam sebulan terakhir. Sementara itu, persentase anak yang memiliki keluhan kesehatan dan pernah rawat jalan dalam sebulan terakhir menunjukkan peningkatan pada tahun 2021 – 2022. Ketika keluhan kesehatan semakin parah, biasanya dilakukan rawat inap. Dari tahun 2020 – 2022, terjadi tren penurunan untuk persentase anak yang pernah rawat inap dalam setahun terakhir.

Beberapa faktor yang memengaruhi keluhan kesehatan yang dialami oleh anak antara lain, ketersediaan jaminan kesehatan anak untuk akses kesehatan yang lebih terjangkau, kebiasaan merokok yang dimulai pada usia dini, pemberian imunisasi yang tepat waktu, serta pemberian ASI eksklusif dan makanan bergizi untuk memenuhi nutrisi yang dibutuhkan anak dalam masa perkembangan otaknya. Penelitian ini diarahkan untuk mengidentifikasi faktor-faktor yang secara signifikan memengaruhi jumlah keluhan kesehatan anak di Indonesia. Dengan demikian, diharapkan hasil penelitian ini dapat memberikan dasar bagi pengembangan strategi yang lebih terfokus dan memberikan kontribusi positif dalam mengurangi angka keluhan kesehatan anak. Keseluruhan penelitian ini diharapkan dapat memberikan langkah-langkah preventif yang konkret serta memberikan masukan dan saran berharga kepada pemerintah dan masyarakat dalam upaya bersama menjaga kesehatan generasi penerus.

### B. Rumusan Masalah

Berdasarkan latar belakang pada bagian sebelumnya, didapatkan beberapa rumusan masalah sebagai berikut.

1. Bagaimana indikator kesehatan anak di Indonesia dapat dikelompokkan dalam beberapa kategori menggunakan metode K-Means?
2. Apa saja indikator kesehatan anak yang memengaruhi pengelompokan provinsi di Indonesia?

### C. Tujuan

Berdasarkan rumusan masalah tersebut, didapatkan beberapa tujuan sebagai berikut.

1. Mengetahui kelompok-kelompok provinsi di Indonesia berdasarkan indikator kesehatan anak menggunakan metode K-Means.
2. Mengetahui faktor-faktor kesehatan anak yang berpengaruh signifikan terhadap pengelompokan provinsi di Indonesia.

## II. METODOLOGI

### A. Dasar Teori

#### 1) Statistika Deskriptif

Statistika deskriptif merupakan teknik yang melibatkan pengumpulan dan penyajian data untuk menghasilkan wawasan yang berharga. Statistik deskriptif berupaya menjelaskan aspek atau atribut suatu situasi dan dapat memberikan gambaran yang metodis dan tepat mengenai fakta dan kualitas peristiwa yang diamati. Statistik deskriptif hanya bertujuan untuk menyediakan informasi terkait data, tanpa terlibat dalam pengambilan keputusan atau menarik kesimpulan.

#### 2) Analisis Cluster

Analisis *cluster* merupakan salah satu analisis multivariat yang berfungsi untuk mengelompokkan objek-objek atau beberapa variabel berdasarkan karakteristik yang dimiliki. Selain itu, analisis *cluster* juga bertujuan dalam memaksimalkan kesamaan objek dalam *cluster* sementara itu juga memaksimalkan perbedaan antar *cluster* (Hair et al, 2009). Pada proses pengclusteran, dilakukan satu perhitungan terhadap pengukuran yang menghasilkan kesamaan atau kedekatan antar objek untuk mengetahui mirip atau berbedanya suatu data yang kompleks, yaitu ukuran jarak atau kemiripan. Secara umum, jarak yang digunakan adalah jarak *euclidean* atau *euclidean distance* (Rachmatin, 2014). Perhitungan untuk jarak *euclidean* dilakukan menggunakan persamaan (1).

$$d_{ij} = \sqrt{\sum_{k=1}^p (y_{ik} - y_{jk})^2}; i, j = 1, 2, \dots, n \quad (1)$$

Analisis *cluster* bermanfaat dalam menyimpulkan suatu data yang kompleks dengan cara mengelompokkan objek-objek yang memiliki kemiripan karakteristik (Santoso, 2015). Adapun *cluster* yang baik mempunyai ciri-ciri sebagai berikut.

- a. Memiliki kesamaan (homogenitas) yang tinggi antar objek dalam satu *cluster* (*within cluster*)
- b. Memiliki perbedaan (heterogenitas) yang tinggi antar *cluster* yang satu dengan *cluster* lainnya (*between cluster*)

#### 3) Algoritma K-Means

Algoritma K-means adalah salah satu metode *cluster* yang paling populer dan banyak digunakan dalam analisis data. Algoritma ini merupakan algoritma *unsupervised learning* yang digunakan untuk mengelompokkan data menjadi *cluster* sebanyak k berdasarkan fitur-fiturnya. Algoritma K-Means digunakan untuk tipe data numerik. Berikut merupakan tahapan-tahapan dalam algoritma ini.

- a. Menentukan jumlah *cluster*
- b. Menginisialisasi pusat *cluster* sebanyak k. Inisialisasi ini dilakukan secara acak dengan k titik dimana setiap titik akan menjadi pusat (*centroid*) dari setiap kelompok.
- c. Menghitung jarak seluruh *data point* ke *centroid* menggunakan *euclidean distance*.
- d. Menentukan *centroid* baru dari *data point* berdasarkan tiap klaster.

$$\mu_j(t+1) = \frac{1}{N_{sj}} \sum_{j \in s_j} x_j \quad (2)$$

- e. Mengulang langkah (b) hingga (d) hingga tidak ada perubahan *data points* yang berpindah *cluster*.

#### 4) Analisis Diskriminan

Analisis diskriminan merupakan analisis statistika yang digunakan untuk mengklasifikasikan kasus-kasus pada variabel bebas ke dalam grup atau kategori pada variabel terikat. Tujuan analisis diskriminan ingin membedakan variabel bebas yang secara nyata dapat memengaruhi variabel terikat (Rizkiana & Hendikawati, 2015).

#### 5) Kesehatan Anak

Kesehatan merupakan keadaan sempurna secara fisik, mental dan sosial, tidak hanya bebas dari penyakit dan kecacatan. Kesehatan secara sosial diartikan sebagai kemampuan seseorang dalam hidup bersama di dalam suatu

kelompok masyarakat. Dengan demikian kesehatan dapat disimpulkan sebagai keadaan yang sempurna baik secara fisik, mental, sosial dan spiritual, tidak hanya bebas dari penyakit, kelemahan dan kecatatan yang memungkinkan seseorang untuk hidup produktif secara sosial dan ekonomi (Nies & McEwen, 2001).

Berikut merupakan ciri-ciri anak sehat dari segi fisik, psikis, dan sosial (Santoso & Ranti, 2004).

- Segi fisik, anak memiliki badan yang sehat dengan pertumbuhan jasmani yang normal.
- Segi psikis, anak dengan jiwa sehat berkembang secara wajar disertai pikiran yang bertambah cerdas dan perasaan yang bertambah peka.
- Segi sosialisasi, anak tampak aktif, gesit, dan gembira serta mudah menyesuaikan diri dengan lingkungan.

## B. Sumber Data

Data yang digunakan pada penelitian ini merupakan data sekunder yang diperoleh dari publikasi Badan Pusat Statistik Indonesia yang berjudul "Profil Kesehatan Ibu dan Anak 2022", khususnya pada bab Kesehatan Anak Investasi untuk Masa Depan Bangsa. Data yang digunakan dalam penelitian adalah sebanyak 34 data sesuai dengan jumlah provinsi di Indonesia pada tahun 2022. Adapun struktur data yang digunakan pada penelitian ini dapat dilihat sebagaimana pada Tabel 1.

**Tabel 1** Struktur Data

Provinsi	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$
Aceh	$x_{1;1}$	$x_{1;1}$	$x_{1;1}$	$x_{1;1}$	$x_{1;1}$	$x_{1;1}$	$x_{1;1}$
Sumatera Barat	$x_{1;2}$	$x_{1;2}$	$x_{1;2}$	$x_{1;2}$	$x_{1;2}$	$x_{1;2}$	$x_{1;2}$
...	...	...	...	...	...	...	...
Papua	$x_{1;34}$	$x_{1;34}$	$x_{1;34}$	$x_{1;34}$	$x_{1;34}$	$x_{1;34}$	$x_{1;34}$

## C. Variabel Penelitian

Penelitian ini merupakan analisis unsupervised dimana variabel yang digunakan merupakan variabel X dan tanpa target variabel. Variabel yang digunakan merupakan beberapa determinan kesehatan anak menurut publikasi Profil Kesehatan Ibu dan Anak 2022. Determinan yang dipilih sebagai variabel pada penelitian ini disajikan pada Tabel 2.

**Tabel 2** Variabel Penelitian

Simbol	Variabel	Skala
$X_1$	Persentase Anak yang Mempunyai Keluhan Kesehatan dalam Sebulan Terakhir	Rasio
$X_2$	Persentase Anak yang Memiliki Jaminan Kesehatan	Rasio
$X_3$	Persentase Anak Umur 5-17 yang Merokok selama Sebulan Terakhir	Rasio
$X_4$	Persentase Anak Umur 12-23 Bulan yang Menerima Imunisasi Dasar Lengkap	Rasio
$X_5$	Persentase Anak Umur 0-23 Bulan yang Pernah dan Masih Diberi ASI	Rasio
$X_6$	Persentase Anak Umur 0-5 Bulan yang Menerima ASI Eksklusif	Rasio
$X_7$	Persentase Anak Umur 6-23 Bulan yang Mengonsumsi Setidaknya Lima dari Delapan Kelompok Makanan dan Minuman Selama Sehari-hari Kemarin	Rasio

## D. Langkah Analisis

Langkah analisis pada penelitian ini adalah sebagai berikut.

- 1) Mendefinisikan latar belakang penelitian, rumusan masalah, dan tujuan penelitian.
- 2) Mengumpulkan data yang akan digunakan dalam analisis.
- 3) Mempersiapkan library yang digunakan untuk analisis pada Google Colaboratory dengan bahasa pemrograman Python.
- 4) Melakukan pre-processing pada data (mengecek *missing value*, data duplikat, dll.)
- 5) Melakukan analisis statistika deskriptif seluruh variabel.
- 6) Melakukan *Exploratory Data Analysis* (EDA) untuk mengetahui karakteristik data.
- 7) Mencari tahu jumlah *cluster* optimum menggunakan *elbow method*.
- 8) Melakukan analisis *clustering* dengan metode K-Means berdasarkan kluster optimum yang didapatkan.
- 9) Mendapatkan *cluster membership* dan menginterpretasi karakteristik tiap kluster.

- 10) Melakukan uji normalitas multivariat
- 11) Melakukan analisis diskriminan.
- 12) Menyusun kesimpulan dan saran

### III. ANALISIS DAN PEMBAHASAN

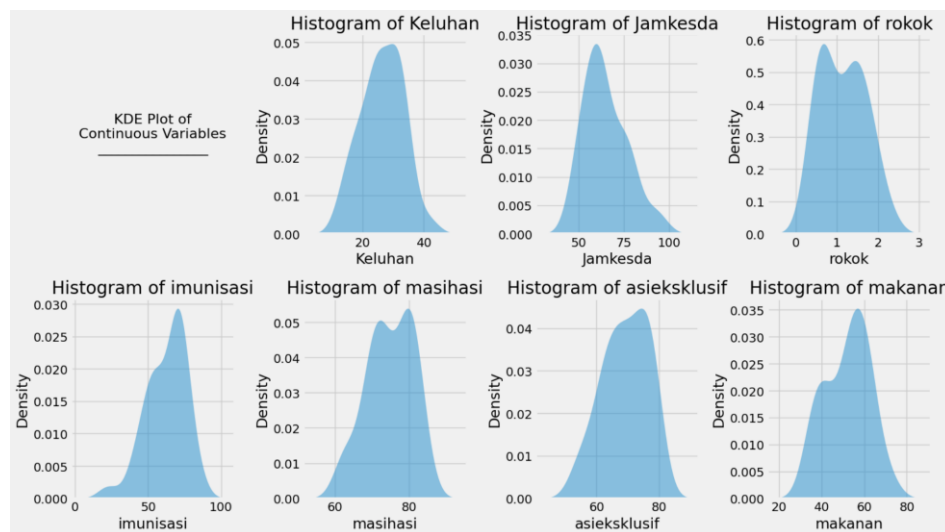
#### A. Statistika Deskriptif

Statistika deskriptif merupakan Teknik yang melibatkan pengumpulan dan penyajian data untuk menghasilkan wawasan yang berharga. Adapun statistika deskriptif pada penelitian ini disajikan dalam Tabel 3.

**Tabel 3** Statistika Deskriptif

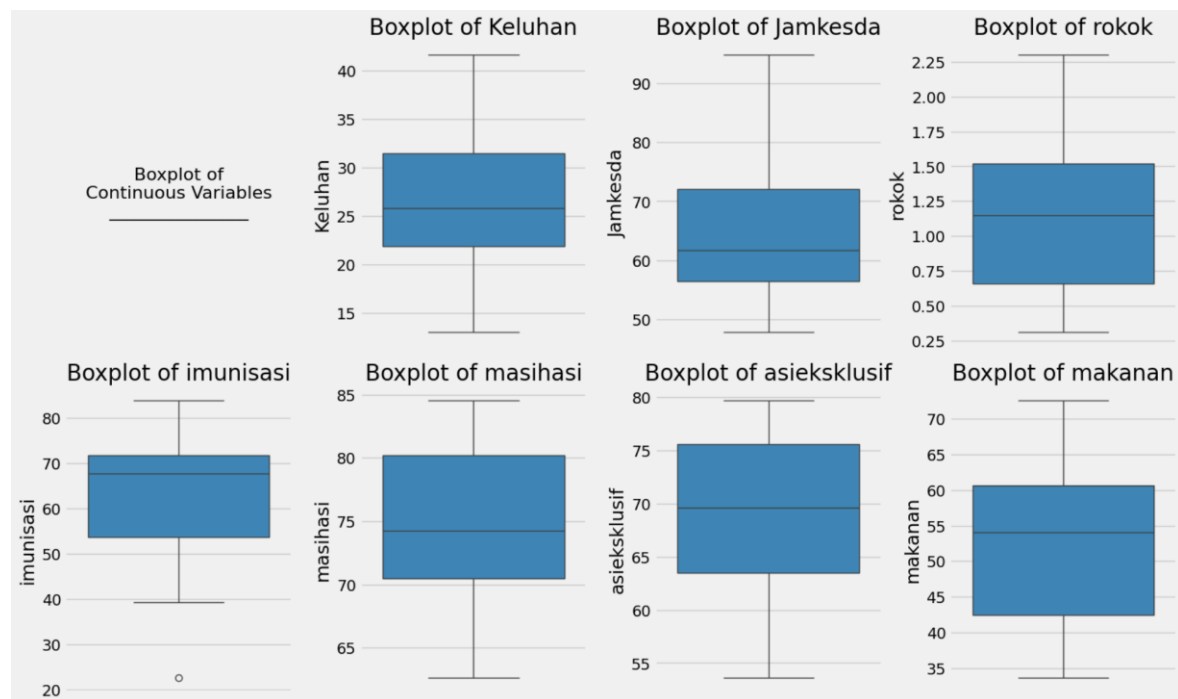
Provinsi	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$
Mean	26.36	64.71	1.14	62.18	74.81	69.21	52.02
StDev	6.77	11.79	0.55	13.53	6.11	7.26	10.35
Minimum	13.03	47.78	0.31	22.52	62.62	53.60	33.63
Median	25.81	61.73	1.15	67.73	74.28	69.62	54.06
Maksimum	41.65	94.87	2.30	83.89	84.52	79.69	72.54

Selanjutnya, persebaran data di masing-masing variabel dapat dilihat pada Gambar 1.

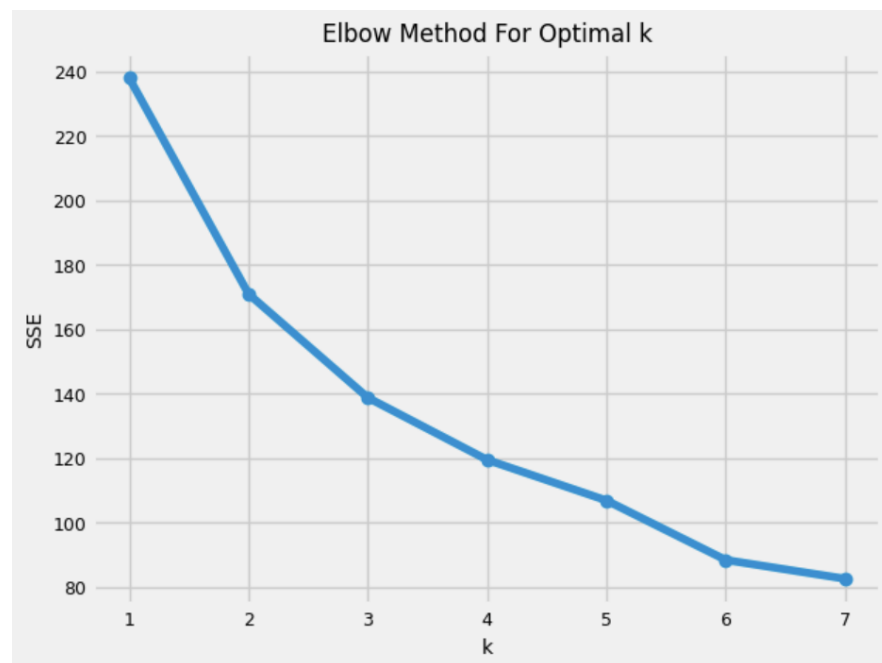


**Gambar 1** Density Plot Indikator Kesehatan Anak Tahun 2022

Selain itu, EDA dengan boxplot juga dilakukan untuk melihat persebaran data dan identifikasi outlier. Berdasarkan Gambar 2, terlihat adanya 1 outlier pada variabel Persentase Anak Umur 12-23 Bulan yang Menerima Imunisasi Dasar Lengkap. Provinsi yang merupakan outlier dari variabel tersebut adalah provinsi Aceh.



Gambar 2 Box Plot Indikator Kesehatan Anak Tahun 2022



Gambar 3 Hasil Elbow Method

Setelah dilakukan EDA, dilanjutkan analisis kluster. Gambar 3 menunjukkan hasil perhitungan menggunakan Elbow Method untuk mendapatkan kluster optimum yang akan digunakan untuk analisis *clustering*. Setelah itu, untuk menentukan lokasi siku, digunakan fungsi `KneeLocator()` dan didapatkan bahwa kluster optimum adalah  $k = 3$ . Dengan begitu, selanjutnya dilakukan analisis kluster K-Means dengan jumlah kluster sebanyak 3. Hasil pengelompokan provinsi berdasarkan determinan kesehatan anak Indonesia Tahun 2022 disajikan pada Tabel 4.

Tabel 4 Hasil Pengelompokan Provinsi Berdasarkan Indikator Kesehatan Anak di Indonesia Tahun 2022

Klaster	Jumlah	Provinsi
0	15	Sumatera Barat, Jambi, Sumatera Selatan, Bangkulu, Lampung, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Selatan, Sulawesi Selatan
1	9	Aceh, Riau, Kep. Riau, DKI Jakarta, Bali, Kalimantan Timur, Kalimantan Utara, Sulawesi Barat, Papua

2	10	Sumatera Utara, Kep. Bangka Belitung, Kalimantan Tengah, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Maluku, Maluku Utara, Papua Barat
---	----	---

Hasil analisis kluster k-means didapatkan 3 kluster yang terbentuk yaitu kluster 0 terdiri dari 15 provinsi, kluster 1 terdiri dari 10 provinsi, dan kluster 2 terdiri dari 9 provinsi. Setelah itu, untuk mengidentifikasi karakteristik setiap kluster yang terbentuk, dilihat rata-rata dari setiap indikator Kesehatan anak seluruh anggota kluster yang selanjutnya dibandingkan dengan kluster lain.

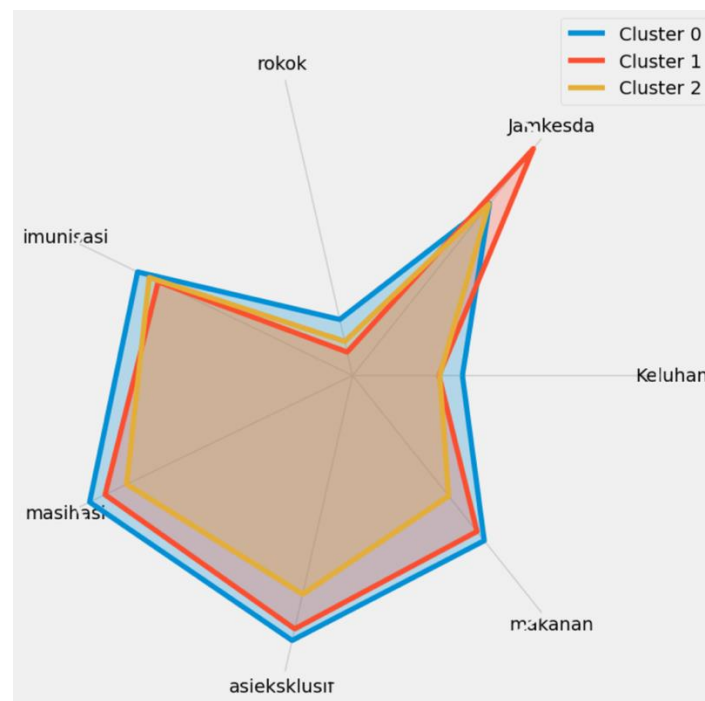
**Tabel 5** Profiling Hasil Analisis Berdasarkan Kluster yang Terbentuk

Variabel	Rata-rata		
	Kluster 0	Kluster 1	Kluster 2
Persentase Anak yang Mempunyai Keluhan Kesehatan dalam Sebulan Terakhir	29.85	23.49	23.69
Persentase Anak yang Memiliki Jaminan Kesehatan	59.70	78.86	59.49
Persentase Anak Umur 5-17 yang Merokok selama Sebulan Terakhir	1.56	0.65	0.95
Persentase Anak Umur 12-23 Bulan yang Menerima Imunisasi Dasar Lengkap	64.86	58.64	61.36
Persentase Anak Umur 0-23 Bulan yang Pernah dan Masih Diberi ASI	79.32	74.67	68.16
Persentase Anak Umur 0-5 Bulan yang Menerima ASI Eksklusif	73.85	70.63	60.97
Persentase Anak Umur 6-23 Bulan yang Mengonsumsi Setidaknya Lima dari Delapan Kelompok Makanan dan Minuman Selama Sehari Kemarin	57.42	54.24	41.91

Provinsi yang termasuk dalam kluster 0 memiliki karakteristik Kesehatan anak yang bernilai dominan bagus dibandingkan kluster lainnya. Hal ini dibuktikan dengan provinsi yang termasuk dalam kluster 0 ini memiliki nilai rata-rata anak umur 6-23 bulan yang mengonsumsi setidaknya lima dari delapan kelompok manan dan minuman, rata-rata anak umur 0-5 bulan penerima eksklusif, rata-rata anak umur 0-23 yang pernah dan masih diberi ASI, rata-rata anak umur 12-23 bulan yang menerima imunisasi dasar lengkap yang lebih tinggi daripada kluster lainnya dan menunjukkan kriteria yang bagus. Disisi lain, provinsi yang tercakup dalam kluster 0 ini memiliki rata-rata anak umur 5-17 tahun yang merokok dan anak yang mempunyai keluhan Kesehatan yang lebih tinggi dibanding kluster lain.

Provinsi yang termasuk dalam kluster 1 memiliki karakteristik Kesehatan anak yang bernilai dominan sedang jika dibandingkan dengan kluster yang lain. Hal ini dibuktikan dengan nilai rata-rata anak umur 6-23 bulan yang mengonsumsi setidaknya lima dari delapan kelompok manan dan minuman, rata-rata anak umur 0-5 bulan penerima eksklusif, rata-rata anak umur 0-23 yang pernah dan masih diberi ASI yang sedang dibandingkan kluster lainnya. Di sisi lain, provinsi yang tercakup dalam kluster 1 ini memiliki nilai rata-rata anak umur 12-23 bulan yang menerima imunisasi dasar lengkap, rata-rata anak umur 5-17 tahun yang merokok, dan anak yang mempunyai keluhan kesehatan yang lebih rendah dibanding kluster lain. Selain itu, provinsi yang termasuk dalam kluster 1 ini memiliki nilai rata-rata anak yang memiliki jaminan Kesehatan paling tinggi dibanding kluster lainnya.

Provinsi yang termasuk dalam kluster 2 memiliki karakteristik Kesehatan anak yang bernilai dominan kurang bagus jika dibandingkan dengan kluster yang lain. Hal ini dibuktikan dengan nilai rata-rata anak umur 6-23 bulan yang mengonsumsi setidaknya lima dari delapan kelompok manan dan minuman, rata-rata anak umur 0-5 bulan penerima eksklusif, rata-rata anak umur 0-23 yang pernah dan masih diberi ASI, dan rata-rata anak yang memiliki jaminan kesehatan yang rendah dibandingkan kluster lainnya. Di sisi lain, provinsi yang tercakup dalam kluster 2 ini memiliki nilai rata-rata anak umur 12-23 bulan yang menerima imunisasi dasar lengkap, rata-rata anak umur 5-17 tahun yang merokok, dan anak yang mempunyai keluhan kesehatan yang sedang dibanding kluster lain. Gambaran karakteristik berdasarkan nilai rata-rata tiap kluster dapat dilihat pada Gambar 4.



**Gambar 5** Radial Axis Plot Analisis Berdasarkan Kluster yang Terbentuk

Selanjutnya, ingin diketahui indikator kesehatan anak apa yang secara signifikan mampu membedakan antara kluster kelompok-kelompok provinsi yang terbentuk. Hal ini dilakukan menggunakan analisis diskriminan dimana hasil analisis kluster dijadikan variabel dependen. Sebelum melakukan analisis diskriminan, perlu dilakukan pengujian normal multivariat menggunakan *Mardia's Test*.

**Tabel 6** Hasil Uji Multivariat Normal

Metrik	p-value
Skewness	0.03
Kurtosis	0.43

Dengan taraf signifikansi 0.01, keputusan Gagal Tolak  $H_0$  yang berarti tidak cukup bukti untuk menyimpulkan bahwa ketujuh variabel tidak mengikuti distribusi normal multivariat. Dengan begitu, dilanjutkan analisis diskriminan.

Analisis diskriminan pada penelitian ini dilakukan menggunakan software SPSS. Ada beberapa hal yang perlu dibahas dari output SPSS. Pertama, identifikasi variabel diskriminan dapat dilihat pada Tabel 7.

**Tabel 7** Uji Signifikansi Variabel Bebas

Variabel	p-value
Persentase Anak yang Mempunyai Keluhan Kesehatan dalam Sebulan Terakhir	0.022
Persentase Anak yang Memiliki Jaminan Kesehatan	0.000
Persentase Anak Umur 5-17 yang Merokok selama Sebulan Terakhir	0.000
Persentase Anak Umur 12-23 Bulan yang Menerima Imunisasi Dasar Lengkap	0.552
Persentase Anak Umur 0-23 Bulan yang Pernah dan Masih Diberi ASI	0.000
Persentase Anak Umur 0-5 Bulan yang Menerima ASI Eksklusif	0.000
Persentase Anak Umur 6-23 Bulan yang Mengonsumsi Setidaknya Lima dari Delapan Kelompok Makanan dan Minuman Selama Sehari-hari Kemarin	0.000

Identifikasi variabel bertujuan untuk mengetahui apakah variabel-variabel independen pada data dapat digunakan untuk menguji perbedaan antar kelompok. Tabel 7 menunjukkan bahwa secara univariat, semua variabel independent



kecuali persentase anak yang mempunyai keluhan Kesehatan dan persentase anak yang menerima imunisasi dasar lengkap mampu membedakan antar kategori yang dilihat dari nilai  $p\text{-value} < 0.01$ .

Langkah selanjutnya adalah proses penentuan variabel. Variabel yang terbentuk merupakan variabel yang signifikan dan dapat dimasukkan dalam persamaan diskriminan menggunakan pendekatan stepwise.

**Tabel 8** Penentuan Variabel Bebas dalam Fungsi Diskriminan

Tahap	Variabel	F	p-value
1	Persentase Anak Umur 0-23 Bulan yang Pernah dan Masih Diberi ASI	0.395	0.000
2	Persentase Anak yang Memiliki Jaminan Kesehatan	0.171	0.000
3	Persentase Anak Umur 5-17 yang Merokok selama Sebulan Terakhir	0.106	0.000

Tabel 8 menunjukkan hanya 3 variabel yang dapat dimasukkan dalam persamaan diskriminan. Hal ini menunjukkan bahwa persentase anak umur 0-23 bulan yang pernah dan masih diberi ASI, anak yang memiliki jaminan kesehatan, dan anak umur 5-17 tahun yang merokok sangat berpengaruh terhadap pengelompokan indikator kesehatan anak.

Setelah itu, nilai Canonical Correlation diinterpretasikan untuk menjelaskan seberapa besar variabel memberikan perbedaan pada tiap klaster. Nilai Canonical Correlation berada pada nilai 0-1 yang mana semakin tinggi nilai tersebut, semakin baik fungsi menjelaskan variabel yang diamati.

**Tabel 9** Keeratan Hubungan Nilai Diskriminan dengan Kategori Kepuasan

Fungsi	Nilai Eigen	% Variansi	Kumulatif % Variansi	Canonical Correlation
1	2.23	53.8	53.8	0.831
2	1.91	46.2	100.0	0.810

Nilai Canonical Correlation 0.831 menunjukkan keeratan hubungan antara skor diskriminan dengan klaster provinsi cukup tinggi. Selanjutnya, dilakukan uji untuk mengetahui perbedaan rata-rata variabel diskriminan 3 faktor untuk menentukan signifikansi fungsi diskriminan yang dapat diketahui melalui nilai Wilks' Lambda.

**Tabel 10** Nilai Wilks' Lambda

Fungsi	Wilks' Lambda	Chi-square	df	p-value
1 dari 2	0.106	67.210	6	0.000
2	0.343	32.079	2	0.000

Berdasarkan Tabel 10, nilai  $p\text{-value}$  pada baris pertama dan kedua  $< 0.01$  sehingga  $H_0$  ditolak. Hal ini berarti terdapat perbedaan rata-rata tiga klaster dari kedua fungsi diskriminan.

**Tabel 11** Struktur Matriks Fungsi Diskriminan

	Fungsi	
	1	2
Persentase Anak Umur 5-17 yang Merokok selama Sebulan Terakhir	0.697*	-0.042
Persentase Anak Umur 0-5 Bulan yang Menerima ASI Eksklusif <sup>b</sup>	0.169*	-0.053
Persentase Anak Umur 6-23 Bulan yang Mengonsumsi Setidaknya Lima dari Delapan Kelompok Makanan dan Minuman Selama Sehari-hari Kemarin <sup>b</sup>	0.045*	0.025
Persentase Anak Umur 0-23 Bulan yang Pernah dan Masih Diberi ASI	0.591	0.629*
Persentase Anak yang Memiliki Jaminan Kesehatan	-0.508	0.547*
Persentase Anak yang Mempunyai Keluhan Kesehatan dalam Sebulan Terakhir <sup>b</sup>	0.225	0.290*
Persentase Anak Umur 12-23 Bulan yang Menerima Imunisasi Dasar Lengkap <sup>b</sup>	-0.068	-0.179*



Tabel 11 menunjukkan besar korelasi antara variabel bebas dengan dua fungsi diskriminan yang terbentuk dan variabel-variabel dengan tanda huruf b dikanan atas nama variabel melambangkan variabel tersebut tidak dapat masuk ke dalam fungsi diskriminan.

Fungsi diskriminan digunakan untuk mengetahui sebuah case masuk pada kluster satu atau tergolong pada grup lainnya. Pembentukan fungsi diskriminan dapat dilihat pada Tabel 12.

**Tabel 12** Koefisien Fungsi Diskriminan

Variabel	Fungsi	
	1	2
Persentase Anak yang Memiliki Jaminan Kesehatan	-0.054	0.098
Persentase Anak Umur 5-17 yang Merokok selama Sebulan Terakhir	1.865	-0.365
Persentase Anak Umur 0-23 Bulan yang Pernah dan Masih Diberi ASI	0.112	0.220
Konstanta	-6.947	-22.369

Fungsi diskriminan yang terbentuk berdasarkan Tabel 12 distuliskan sebagai berikut.

$$zScore = \alpha + b_2x_2 + b_3x_3 + b_5x_5$$

Dengan fungsi diskriminan pada fungsi 1:

$$zScore_1 = -6.947 - 0.054x_2 + 1.865x_3 + 0.112x_5$$

Dengan fungsi diskriminan pada fungsi 2:

$$zScore = -22.369 + 0.098x_2 - 0.365x_3 + 0.220x_5$$

Setelah fungsi diskriminan diperoleh, kemudian dilakukan pengujian klasifikasi fungsi diskriminan, selanjutnya dilihat seberapa besar fungsi diskriminan yang terbentuk dapat dengan tepat mengklasifikasikan kategori atau dalam penelitian ini kluster. Hal ini dapat dilihat pada Tabel 13.

**Tabel 13** Ketepatan Hasil Klasifikasi

Sumber	Klaster	Prediksi Keanggotaan Klaster			Total	
		0	1	2		
Data Asal	Jumlah	0	14	1	0	15
		1	1	8	0	9
		2	0	0	10	10
	%	0	93.3	6.7	0.0	100.0
		1	11.1	88.9	0.0	100.0
		2	0.0	0.0	100.0	100.0
Hasil Validasi Silang	Jumlah	0	13	2	0	15
		1	1	8	0	9
		2	0	0	10	10
	%	0	86.7	13.3	0.0	100.0
		1	11.1	88.9	0.0	100.0
		2	0.0	0.0	100.0	100.0

Berdasarkan perhitungan tingkat akurasi, diketahui bahwa model yang terbentuk memiliki tingkat akurasi dalam memprediksi skor diskriminan sebesar 94.1%.

## IV. KESIMPULAN

Hasil pengelompokan provinsi di Indonesia berdasarkan indikator Kesehatan anak dengan analisis kluster k-means didaatkan 3 kluster. Kluster dengan indikator Kesehatan dominan bagus terdiri dari 15 provinsi, yaitu Sumatera Barat, Jambi, Sumatera Selatan, Bangkulu, Lampung, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Selatan, Sulawesi Selatan. Kluster dengan indikator Kesehatan dominan sedang terdiri dari 9 provinsi, yaitu Aceh, Riau, Kep. Riau, DKI Jakarta, Bali, Kalimantan Timur, Kalimantan Utara, Sulawesi Barat, Papua. Kluster dengan indikator Kesehatan anak dominan kurang bagus yang terdiri dari 10 provinsi, yaitu Sumatera Utara, Kep. Bangka Belitung, Kalimantan Tengah, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Maluku, Maluku Utara, Papua Barat.

Dari hasil analisis kluster, kluster dijadikan variabel dependen untuk analisis diskriminan untuk mengetahui indikator Kesehatan mana yang berpengaruh signifikan dalam pengelompokan kluster. Dari 7 indikator Kesehatan anak, hanya tiga variabel yang masuk dalam fungsi diskriminan, yaitu persentase anak umur 0-23 bulan yang pernah dan masih diberi ASI, anak yang memiliki jaminan kesehatan, dan anak umur 5-17 tahun yang merokok. Model diskriminan yang terbentuk adalah  $zScore_1 = -6.947 - 0.054x_2 + 1.865x_3 + 0.112x_5$  dan  $zScore = -22.369 + 0.098x_2 - 0.365x_3 + 0.220x_5$ . Model diskriminan memiliki ketepatan pengelompokan sebesar 94.1%.

Berdasarkan karakteristik yang dimiliki dalam tiap kluster, prioritas penanganan masalah kesehatan anak dapat dilakukan dari kluster dengan nilai indikator kesehatan anak yang kurang bagus, disusul kluster dengan nilai indikator kesehatan dominan sedang dan prioritas terendah yaitu pada kluster dengan nilai indikator kesehatan dominan bag

## REFERENCES

- [1] Beattie, A., Yates, R., & Noble, D. J. (2016). Accelerating progress towards universal health coverage in Asia and Pacific: Improving the future for women and children. <https://doi.org/10.1136/bmjgh-2016-000190>
- [2] Bloom D.E, & Williamson, J. G. (1998). Demographic Transitions and Economic Miracles in Emerging Asia. *World Bank Economic Review*, 12: 419-45.
- [3] Hair, Joseph F., Black, W. C., Babin, B. J., dkk. 2009. *Multivariat Data Analysis (7th ed)*. Upper Saddle River: Prentice-Hall Internasional, Inc
- [4] Nies, M. A., & McEwen, M. (2015). *Community/public health nursing: promoting the health of populations*. 6th edition. St. Louis, Elsevier/Saunders.
- [5] Rachmatin, Dewi. 2014. Aplikasi Metode-Metode Agglomerative dalam Analisis Kluster pada Data Tingkat Polusi Udara. *Jurnal Ilmiah Studi Matematika STKIP Siliwangi Bandung*, Vol 3, No 2
- [6] Santoso, Singgih. 2015. *Menguasai Statistika Multivariat Konsep Dasar dan Aplikasi dengan SPSS*. Jakarta: PT Elex Media Komputindo.
- [7] Santoso, Soegeng dan Ranti, Anne Lies. 2004. *Kesehatan & Gizi*. Jakarta: Rineka Cipta
- [8] Rizkiana, A. & Hendikawati, P. 2015. Klasifikasi Tingkat Kepuasan Pelanggan dengan Analisis Diskriminan. *Jurnal MIPA*, Vol. 38, No. 1.