

# Pengelompokan Provinsi di Indonesia Menggunakan *Gaussian Mixture Model* Berdasarkan Indikator Kemiskinan

Nyimas Nina Alyarahma<sup>1</sup>, Gusmi Kholijah<sup>2\*</sup>, Corry Sormin<sup>3</sup>

<sup>1</sup> Program Studi Matematika, Universitas Jambi, Jambi, Indonesia

<sup>2,3</sup> Program Studi Statistika, Universitas Jambi, Jambi, Indonesia

Corresponding Email\*: [gusmikholijah@unja.ac.id](mailto:gusmikholijah@unja.ac.id)

## Abstrak

Kemiskinan ialah sebuah gagasan multidimensional yang tidak hanya terlihat melalui segi ekonomi saja, namun juga dari segi sosial, budaya, serta politik. Kemiskinan umumnya dinyatakan sebagai konsep yang statis dengan jumlah kemiskinan sebagai gambaran kondisi kesejahteraan pada suatu waktu tertentu. Kenyataannya kemiskinan memiliki periode waktu yang tidak pernah terpotong dan memiliki pola yang kontinu seiring berjalannya waktu. Pemerintah Indonesia berkomitmen dalam upaya pengentasan kemiskinan agar dapat menekan angka kemiskinan Indonesia. Salah satu cara yang dapat dilakukan adalah alokasi bantuan untuk masyarakat. Pengelompokan yang dilakukan oleh analisis *cluster* menggunakan metode *Gaussian Mixture Model* dengan algoritma *Expectation-Maximization* (EM). Jumlah *cluster* optimal pada *Gaussian Mixture Model* menggunakan metode *Bayesian Information Criterion* (BIC). Berdasarkan penelitian yang dilakukan diperoleh hasil analisis dari pengelompokan *cluster* dengan metode *Gaussian Mixture Model* dengan nilai BIC terkecil yaitu -397,6876 dengan jumlah *cluster* sebanyak 2. Cluster 1 dicirikan dengan Garis Kemiskinan dan Rata-rata Lama Sekolah yang tinggi sedangkan Cluster 2 dicirikan dengan Tingkat Pengangguran Terbuka serta Rasio Gini yang tinggi.

**Kata Kunci:** Cluster, *Gaussian Mixture Model*, Kemiskinan, Algoritma EM

## Abstract

Poverty is a multidimensional concept that can not only be seen from an economic perspective, but can also be seen from a social, cultural and political perspective. Poverty is generally expressed as a static concept with the amount of poverty as a description of welfare conditions at a certain time. In reality, poverty has a time period that is never cut off and has a continuous pattern over time. The Indonesian government is committed to efforts to eradicate poverty in order to reduce Indonesia's poverty rate. One way that can be done is by allocating aid to the community. Grouping is carried out by cluster analysis using the *Gaussian Mixture Model* method with the *Expectation-Maximization* (EM) algorithm. The optimal number of clusters in the *Gaussian Mixture Model* uses the *Bayesian Information Criterion* (BIC) method. Based on the research carried out, the results of the analysis of cluster grouping using the *Gaussian Mixture Model* method were obtained with the smallest BIC value, namely -397.6876 with a total of 2 clusters.

**Keywords:** Cluster, *Gaussian Mixture Model*, Poverty, EM Algorithm

Received 11-07-2024, Revised :09-09-2024 Accepted :23-09-2024

## 1. Pendahuluan

Kemiskinan ialah sebuah gagasan multidimensional dimana bukan hanya terlihat dari segi ekonomi, tetapi terlihat juga dalam segi sosial, budaya, serta politik [1]. Kemiskinan umumnya dinyatakan sebagai konsep yang statis dengan jumlah kemiskinan sebagai refleksi Kualitas hidup pada periode tertentu [2]. Faktanya kemiskinan memiliki periode waktu yang tidak pernah terpotong dan selalu memiliki pola kontinu yang berkelanjutan seiring dengan berjalannya waktu [3]. Kondisi pola kontinu ini berkaitan dengan peningkatan jumlah orang miskin yang tinggal di sekitar ambang kemiskinan seiring dengan naiknya tingkat kemiskinan [4]. Akibat dari perubahan kondisi kemiskinan pada masyarakat yang cepat berubah, pemutakhiran data kemiskinan perlu terus diperbaharui. Pemutakhiran data kemiskinan yang dilakukan Badan Pusat Statistik dilakukan setiap tahun pada setiap semester yaitu bulan Maret dan September. Data tersebut dikumpulkan pada setiap semester untuk menilai efektivitas kebijakan

pemerintah terhadap kemiskinan, membandingkan tingkat kemiskinan dari waktu ke waktu dan antar daerah, serta menetapkan tujuan untuk mengurangi jumlah penduduk miskin dan meningkatkan peningkatan kehidupan mereka [5]. Data yang didapat dari survei, yaitu jumlah penduduk miskin pada Maret 2023 sebesar 25,90 juta orang. Jika diselisihkan, jumlah tersebut mengalami penurunan sebesar 0,46 juta orang pada bulan September 2022 yang bernilai 26,36 juta orang. Meskipun terjadi penurunan jumlah penduduk miskin pada Maret 2023, jumlah angka tersebut belum cukup besar jika dibandingkan dengan keseluruhan jumlah penduduk Indonesia yaitu sebesar 278,6 juta jiwa [6]. Akibat hal ini Indonesia tercatat berada pada urutan ke-73 dalam daftar 100 negara paling miskin di dunia [7].

Tingginya jumlah penduduk miskin di Indonesia mengakibatkan guncangan ekonomi yang mendorong masyarakat miskin berada di bawah garis kemiskinan [4]. Jika saat ini masyarakat dalam kondisi miskin, maka masyarakat tersebut akan terus terjebak dalam kemiskinan [4]. Pemerintah Indonesia berkomitmen dalam upaya pengentasan kemiskinan yang dijelaskan dalam Prioritas Rencana Kerja Pemerintah (RKP) 2023, yaitu percepatan penghapusan kemiskinan 2,5 hingga 3 juta penduduk. Salah satu program yang dapat menekan angka kemiskinan yaitu alokasi anggaran belanja pemerintah [8]. Menurut Sumargo, angka kemiskinan dapat ditekan dengan program alokasi dan besarnya nilai bantuan untuk masyarakat sesuai kebutuhan berdasarkan karakteristik yang dimiliki setiap provinsi [9]. Alokasi berdasarkan karakteristik tersebut dilakukan karena masing-masing provinsi di Indonesia memiliki perbedaan dalam hal ekonomi, sehingga program alokasi dan besar nilai bantuan dapat dianalisis berdasarkan masalah faktor yang memengaruhi karakteristik kemiskinan masing-masing provinsi. Permasalahan karakteristik kemiskinan dapat dilakukan dengan cara mengelompokkan provinsi-provisi yang memiliki kesamaan dan perbedaan karakteristik [10]. Salah satu metode pengelompokan pada bidang matematika yaitu analisis *cluster* [11]. Analisa *cluster* ialah pengelompokkan objek ke dalam berbagai kelompok yang mempunyai sifat sama dalam satu kelompok dan sifat berbeda antar kelompok [12]. Pengelompokan *clustering* tersebut dapat menunjukkan provinsi yang memiliki karakteristik sama antar kelompok dan yang memiliki karakteristik sangat berbeda antar kelompok *cluster*.

Pengelompokan dapat dilakukan jika karakteristik yang memengaruhi kemiskinan diketahui, salah satu cara yang dapat digunakan adalah distribusi. Distribusi dapat menggambarkan distribusi peluang suatu nilai untuk variabel yang berbeda, salah satu distribusi peluang adalah distribusi *Gaussian*. Distribusi peluang terbagi menjadi dua, salah satunya adalah distribusi peluang kontinu yang dimana salah satu dari bagian distribusi peluang kontinu adalah distribusi *gaussian* [13]. Distribusi *gaussian* adalah distribusi peluang berbentuk kurva lonceng yang simetris terhadap rata-rata [14]. Analisis *cluster* memiliki syarat yang perlu dipenuhi yaitu mengikuti distribusi *gaussian* [15]. Analisis distribusi kontinu memiliki keterkaitan dengan permasalahan kemiskinan, yaitu periode waktu kemiskinan di Indonesia tidak pernah terpotong dan selalu memiliki pola kontinu yang berkelanjutan seiring dengan berjalannya waktu [3], sehingga dapat diartikan bahwa data kemiskinan mengikuti distribusi *gaussian* [3] atau dengan kata lain masalah kemiskinan dapat menggunakan distribusi peluang. Suatu data kemiskinan juga dapat memiliki lebih dari satu karakteristik distribusi *gaussian*, sehingga data tersebut dapat di modelkan lebih dari satu distribusi. Data yang didistribusikan lebih dari satu distribusi disebut distribusi model campuran atau *mixture model* [16]. Salah satu model campuran yang dapat digunakan dalam distribusi *gaussian* ialah *Gaussian Mixture Model* (GMM).

GMM adalah model peluang menggunakan asumsi bahwa data mengikuti karakteristik dari beberapa distribusi *gaussian* dengan parameter yang tidak diketahui [17]. Permasalahan dalam penggunaan GMM adalah parameternya yang cukup banyak sehingga menyebabkan parameter GMM tidak dapat diestimasi secara langsung melalui metode *maximum likelihood estimation* (MLE). Cara mengatasi permasalahan pada MLE tersebut dapat menggunakan algoritma *Expectation Maximization* (EM) [18]. Algoritma EM merupakan iterasi menggunakan dua tahap, yaitu ekspektasi (*expectation*) dan maksimalisasi (*maximization*). Langkah ekspektasi adalah mengestimasi nilai ekspektasi variabel yang belum diketahui, lalu membentuk persamaan *log-likelihood* dari data lengkap. Langkah maksimalisasi adalah langkah untuk mencari estimasi parameter-parameter model dan memaksimalkan persamaan *log-likelihood* data lengkap yang diperoleh dari langkah ekspektasi [13]. Berdasarkan penjelasan sebelumnya, penelitian ini akan melakukan pengelompokan provinsi berdasarkan indikator kemiskinan dengan menggunakan analisis *Gaussian Mixture Model* (GMM) untuk provinsi-provisi yang ada di Indonesia.

## 2. Metode

### 2.1 Data

Tempat pada riset ini dilaksanakan di Badan Pusat Statistik (BPS) Provinsi Jambi beralamat di Jl. Ahmad Yani No.4, Telanaipura, Kota Jambi, Provinsi Jambi 36122. Waktu penelitian ini dilaksanakan bulan Januari – Maret 2024. Muasal data yang di gunakan dalam penelitian ini adalah data sekunder yang diperoleh dari laman *website* bps.go.id Badan Pusat Statistik (BPS) Indonesia. Metode analisis data untuk melakukan prediksi dikerjakan pada software R dengan tahapan sebagai berikut.

### 2.2 Clustering Gaussian Mixture Model (GMM)

1. Menentukan jumlah *cluster* terbaik dengan menggunakan BIC menggunakan persamaan (1)

$$BIC = -2 \log f(x) + s \log N \quad (1)$$

dengan,

$\log f(x)$  : *log-likelihood* distribusi *gaussian*

$s$  : jumlah parameter

$N$  : jumlah observasi

2. Menentukan model GMM pada masing-masing *cluster* yang dimana setiap *cluster* memiliki tiga parameter, yaitu rata-rata ( $\mu$ ), kovarians ( $\Sigma$ ) dan bobot ( $\lambda$ ). Dalam menentukan model GMM dirumuskan dengan menggunakan persamaan (2).

$$f(x) = \sum_{k=1}^K \lambda_k \left( \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp \left( -\frac{1}{2} (x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k) \right) \right) \quad (2)$$

dengan,

$f(x)$  : fungsi distribusi GMM

$p$  : variabel-variabel acak yang diinterpretasi sebagai objek observasi

$\lambda$  : bobot

$k$  : banyaknya cluster

Penentuan parameter rata-rata ( $\mu$ ), kovarians ( $\Sigma$ ) dan bobot ( $\lambda$ ) pada model GMM dilakukan dengan;

$$\begin{aligned} \mu_k &= \frac{\left( \sum_{n=1}^N \frac{\lambda_k \left( \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp \left( -\frac{1}{2} (x_N - \mu_k)^T \Sigma_k^{-1} (x_N - \mu_k) \right) \right)}{\sum_{k=1}^K \lambda_k \left( \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp \left( -\frac{1}{2} (x_N - \mu_k)^T \Sigma_k^{-1} (x_N - \mu_k) \right) \right)} \right) \times x_N}{\left( \sum_{n=1}^N \frac{\lambda_k \left( \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp \left( -\frac{1}{2} (x_N - \mu_k)^T \Sigma_k^{-1} (x_N - \mu_k) \right) \right)}{\sum_{k=1}^K \lambda_k \left( \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp \left( -\frac{1}{2} (x_N - \mu_k)^T \Sigma_k^{-1} (x_N - \mu_k) \right) \right)} \right)} \\ \sum_k &= \frac{\left( \sum_{n=1}^N \frac{\lambda_k \left( \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp \left( -\frac{1}{2} (x_N - \mu_k)^T \Sigma_k^{-1} (x_N - \mu_k) \right) \right)}{\sum_{k=1}^K \lambda_k \left( \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp \left( -\frac{1}{2} (x_N - \mu_k)^T \Sigma_k^{-1} (x_N - \mu_k) \right) \right)} \right) \times (x_N - \mu_k)(x_N - \mu_k)^T}{\left( \sum_{n=1}^N \frac{\lambda_k \left( \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp \left( -\frac{1}{2} (x_N - \mu_k)^T \Sigma_k^{-1} (x_N - \mu_k) \right) \right)}{\sum_{k=1}^K \lambda_k \left( \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp \left( -\frac{1}{2} (x_N - \mu_k)^T \Sigma_k^{-1} (x_N - \mu_k) \right) \right)} \right)} \\ \lambda_k &= \frac{\sum_{n=1}^N \frac{\lambda_k \left( \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp \left( -\frac{1}{2} (x_N - \mu_k)^T \Sigma_k^{-1} (x_N - \mu_k) \right) \right)}{\sum_{k=1}^K \lambda_k \left( \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp \left( -\frac{1}{2} (x_N - \mu_k)^T \Sigma_k^{-1} (x_N - \mu_k) \right) \right)}}{N} \end{aligned}$$

3. Inisialisasi ketiga parameter sebaran *Gaussian Mixture Model* yaitu rata-rata ( $\mu_k$ ), kovarians ( $\Sigma_k$ ) dan bobot ( $\lambda_k$ ) secara acak, lalu dilakukan evaluasi *log-likelihood*.

4. *Expectation* dilakukan penghitungan mengikuti iterasi yang ada pada peluang dengan menggunakan rumus persamaan (3)

$$\gamma_k(x) = \frac{\lambda_k f_k(x)}{\sum_{j=1}^K \lambda_j f_j(x)} \quad (3)$$

- $\gamma_k(x)$  : peluang distribusi pada saat  $k$   
 $\sum_{j=1}^K \lambda_j f_j(x)$  : total peluang distribusi GMM yang terjadi dari  $j$  hingga ke- $K$   
 $\lambda$  : bobot  
 $f(x)$  : fungsi distribusi *gaussian*

5. Melakukan *maximization*, pada langkah ini akan memperbarui nilai estimasi dan memaksimalkan nilai parameternya dengan menggunakan rumus persamaan (4) untuk nilai bobot ( $\lambda$ ), persamaan (5) untuk rata-rata ( $\mu$ ) dan persamaan (6) untuk nilai kovarians ( $\Sigma$ ) yang baru.

$$\lambda_k = \frac{\sum_{n=1}^N \gamma_k(x)}{N} \quad (4)$$

$$\mu_k = \frac{\sum_{n=1}^N \gamma_k(x) x_n}{\sum_{n=1}^N \gamma_k(x)} \quad (5)$$

$$\Sigma_k = \frac{\sum_{n=1}^N \gamma_k(x) (x_n - \mu_k)(x_n - \mu_k)^T}{\sum_{n=1}^N \gamma_k(x)} \quad (6)$$

6. Evaluasi *log-likelihood* menggunakan persamaan (7) dan periksa konvergensi parameternya. Jika belum konvergen maka kembali ke langkah 4.

$$\ln L(x) = \sum_{n=1}^N \ln \left\{ \sum_{k=1}^K \lambda_k \left( \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp \left( -\frac{1}{2} (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k) \right) \right) \right\} \quad (7)$$

7. Selanjutnya nilai parameter yang konvergen tersebut akan digunakan untuk menghitung peluang setiap data dalam dataset. Peluang ini kemudian digunakan untuk menentukan kluster data dengan cara memilih nilai peluang yang terbesar.  
 8. Menampilkan visualisasi pengelompokan dengan *Gaussian Mixture Model*.  
 9. Kesimpulan.

### 3. Hasil dan Pembahasan

Data yang digunakan dalam penelitian ini merupakan data indikator kemiskinan yaitu Persentase Penduduk Miskin ( $X_1$ ), Indeks Kedalaman Penduduk Miskin ( $X_2$ ), Indeks Keparahan Kemiskinan ( $X_3$ ), Garis Kemiskinan ( $X_4$ ), Rasio Gini ( $X_5$ ), Rata-rata Lama Sekolah ( $X_6$ ), Harapan Lama Sekolah ( $X_7$ ), Tingkat Pengangguran Terbuka ( $X_8$ ). Semua variabel-variabel tersebut merupakan data tahun 2023 sesuai laporan Profil Kemiskinan yang dikeluarkan oleh BPS.

#### 3.1 Analisis Deskriptif

Interpretasi berdasarkan masing-masing data indikator kemiskinan dengan rata-rata, nilai minimum, nilai maksimum dan rata-rata nasional yang dapat di lihat pada **Tabel 1** sebagai berikut:

**Tabel 1.** Data Indikator Kemiskinan 2023

Objek	Provinsi	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$
1	Aceh	14,45%	2,76%	0,78%	Rp.627535	0,30	9,89	14,38	6,03%
2	Sumatera Utara	8,15%	1,26%	0,32%	Rp.602999	0,31	10,07	13,48	5,89%
3	Sumatera Barat	5,95%	0,82%	0,18%	Rp.667925	0,28	9,59	14,11	5,94%
4	Riau	6,68%	1%	0,24%	Rp.658611	0,32	9,60	13,30	4,23%
5	Jambi	7,58%	1,2%	0,29%	Rp.599688	0,34	9,16	13,13	4,53%

6	Sumatera Selatan	11,78%	1,72%	0,39%	Rp.520754	0,34	8,90	12,63	4,11%
7	Bengkulu	14,04%	2,14%	0,52%	Rp.637142	0,33	9,35	13,74	3,42%
8	Lampung	11,11%	1,64%	0,36%	Rp.559011	0,32	8,72	12,77	4,23%
9	Kep. Bangka Belitung	4,52%	0,63%	0,13%	Rp.874204	0,25	8,66	12,31	4,56%
10	Kep. Riau	5,69%	0,82%	0,18%	Rp.742526	0,34	10,52	13,05	6,80%
11	DKI. Jakarta	4,44%	0,7%	0,18%	Rp.792515	0,43	11,42	13,33	6,53%
12	Jawa Barat	7,62	1,17%	0,27%	Rp.495229	0,43	9,16	12,68	7,44%
13	Jawa Tengah	10,77%	1,75%	0,42%	Rp.477580	0,37	8,44	12,85	5,13%
14	DI Yogyakarta	11,04%	1,72%	0,38%	Rp.573022	0,45	10,16	15,66	3,69%
15	Jawa Timur	10,35%	1,63%	0,37%	Rp.507286	0,39	8,53	13,38	4,88%
16	Banten	6,17%	1,2%	0,36%	Rp.618721	0,37	9,48	13,09	7,52%
17	Bali	4,25%	0,55%	0,1%	Rp.529643	0,36	9,74	13,58	2,69%
18	Nusa Tenggara Barat	13,85%	2,38%	0,59%	Rp.498996	0,38	8,39	13,97	2,80%
19	Nusa Tenggara Timur	19,96%	3,33%	0,8%	Rp.507203	0,33	8,31	13,22	3,14%
20	Kalimantan Barat	6,71%	1,03%	0,24%	Rp.563288	0,32	8,17	12,67	5,05%
21	Kalimantan Tengah	5,11%	0,66%	0,14%	Rp.596184	0,32	9,07	12,76	4,10%
22	Kalimantan Selatan	4,29%	0,61%	0,16%	Rp.604266	0,31	8,95	12,86	4,31%
23	Kalimantan Timur	6,11%	0,77%	0,14%	Rp.790186	0,32	10,17	14,02	5,31%
24	Kalimantan Utara	6,45%	0,64%	0,11%	Rp.817876	0,28	9,53	13,20	4,01%
25	Sulawesi Utara	7,38%	1,09%	0,26%	Rp.463432	0,37	9,94	12,96	6,10%
26	Sulawesi Tengah	12,41%	2,12%	0,54%	Rp.568248	0,30	9,22	13,33	2,95%
27	Sulawesi Selatan	8,7%	1,57%	0,41%	Rp.436025	0,38	9,12	13,54	4,33%
28	Sulawesi Tenggara	11,43%	1,96%	0,48%	Rp.443980	0,37	9,62	13,70	3,15%
29	Gorontalo	15,15%	2,92%	0,74%	Rp.442194	0,42	8,48	13,16	3,06%
30	Sulawesi Barat	11,49%	1,8%	0,42%	Rp.433131	0,35	8,48	12,88	2,27%
31	Maluku	16,42%	3,08%	0,86%	Rp.684020	0,29	10,38	14,08	6,31%
32	Maluku Utara	6,46%	1,08%	0,23%	Rp.564733	0,30	9,61	13,74	4,31%
33	Papua Barat	20,49%	4,94%	1,71%	Rp.728619	0,37	10,14	13,34	5,38%
34	Papua	26,03%	6,25%	2,15%	Rp.686469	0,39	7,34	11,15	2,67%
	Rata-rata	10,09%	1,73%	0,45%	Rp.597448	0,34	9,30	13,30	4,61%
	Minimum	4,25%	0,55%	0,1%	Rp.433131	0,24	7,34	11,15	2,27%
	Maksimum	26,03%	6,25%	2,15%	Rp.874204	0,45	11,42	15,66	7,52%
	Rata-rata Nasional	7,50%	1,53%	0,38%	Rp.550458	0,35	12	12	5%

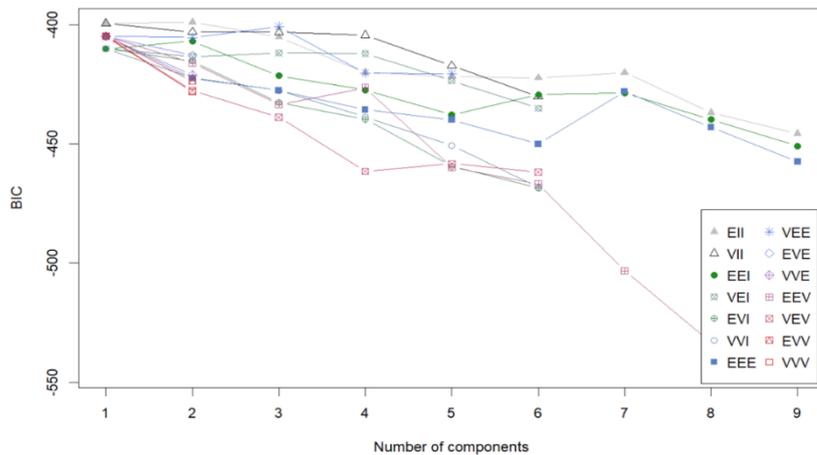
Data yang digunakan pada penelitian ini memiliki satuan yang berbeda antar variabel, sehingga dilakukan proses standarisasi agar hasil *cluster* yang diperoleh memiliki hasil yang optimal [19]. Selanjutnya dicek

asumsi normalitas menggunakan *Shapiro-Wilk* dan diperoleh variabel yang bedistribusi normal adalah variabel  $X_4, X_5, X_6$  dan  $X_8$ . Berikutnya analisis akan dilanjutkan dengan keempat variabel tersebut.

### 3.2 Clustering Gaussian Mixture Model

#### 3.2.1 Jumlah Cluster Optimal menggunakan Bayesian Information Criteration (BIC)

Pengelompokan yang akan dilakukan dengan menggunakan metode *Gaussian Mixture Model* perlu untuk menentukan jumlah *cluster* optimal berdasarkan nilai *Bayesian Information Criteration* (BIC) terkecil, maka semakin kecil nilai yang diperoleh semakin baik pula *cluster* yang terbentuk. Perhitungan BIC dilakukan dengan memanfaatkan nilai *log-likelihood*, formula dalam perhitungan BIC yaitu pada persamaan (1). Hasil perhitungan BIC menggunakan *software R* dengan menggunakan '*mclust*' package menghasilkan output jumlah *cluster* optimal dengan  $k = 2$ . Untuk menegaskan model parameter yang



**Gambar 1.** Visualisasi *Bayesian Information Criteration* (BIC)

optimal tampak dari visualisasi nilai BIC dalam **Gambar 1**. Tampak bahwa model parameter optimal adalah *Equal variance* (EII), dimana untuk *cluster* 1 beranggotakan 12 provinsi dan *cluster* 2 beranggotakan 22 provinsi.

#### 3.2.2 Menentukan model *Gaussian Mixture Model* dan iterasi *Expectation-Maximization*

Setelah menentukan jumlah *cluster* terbaik dengan menggunakan BIC, selanjutnya adalah menentukan model GMM pada masing-masing *cluster* yang dimana masing-masing *cluster* memiliki tiga parameter dengan menggunakan persamaan (2) lalu melakukan evaluasi *log-likelihood*. Setelah dilakukan evaluasi *log-likelihood*, tahap selanjutnya adalah melakukan iterasi *Expectation* pada peluang dengan menggunakan persamaan (3). Selanjutnya melakukan *maximization* dengan memperbarui nilai estimasi dan memaksimalkan nilai parameter, lalu evaluasi *log-likelihood* hingga konvergen. Setelah dilakukan tahap *Expectation-Maximization*, Diperoleh parameter distribusi *Gaussian Mixture Model* untuk setiap *cluster* antara lain:

Untuk *cluster* 1,

$$C_1 \sim \left( \mu_1 \begin{pmatrix} 0,9393576 \\ -0,4815911 \\ 0,7536465 \\ 0,7359485 \end{pmatrix}, \sigma_1^2 \begin{pmatrix} 0,6564251 \\ 0,6564251 \\ 0,6564251 \\ 0,6564251 \end{pmatrix} \right)$$

Untuk *cluster* 2,

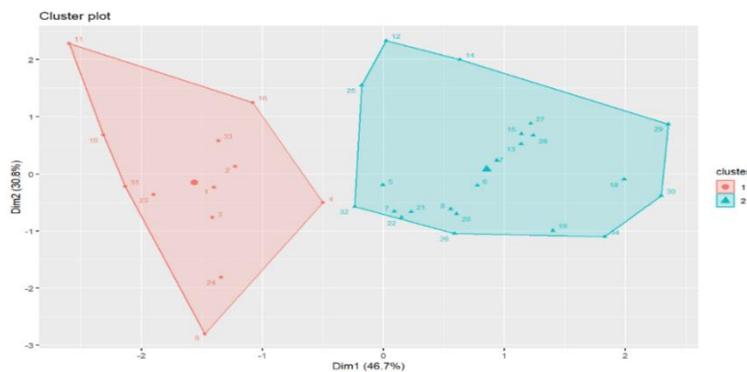
$$C_2 \sim \left( \mu_2 \begin{pmatrix} -0,5307938 \\ 0,2721281 \\ -0,4258558 \\ -0,4158554 \end{pmatrix}, \sigma_2^2 \begin{pmatrix} 0,6564251 \\ 0,6564251 \\ 0,6564251 \\ 0,6564251 \end{pmatrix} \right)$$

Nilai peluang setiap *cluster* untuk setiap data dapat diperoleh, sehingga *cluster* data dari setiap titik data dapat ditentukan dengan cara memilih nilai peluang tertinggi. Hasilnya antara lain:

**Tabel 2.** Peluang Objek pada Setiap Cluster

Objek	Cluster		Peluang Terbesar	Cluster
	1	2		
1	0,956359	0,043641	0,956359	1
2	0,925601	0,074399	0,925601	1
3	0,970192	0,029808	0,970192	1
4	0,533806	0,466194	0,533806	1
5	0,113895	0,886105	0,886105	2
6	0,010686	0,989314	0,989314	2
7	0,114571	0,885429	0,885429	2
8	0,024134	0,975866	0,975866	2
9	0,989457	0,010543	0,989457	1
10	0,99856	0,00144	0,99856	1
11	0,999001	0,000999	0,999001	1
12	0,078451	0,921549	0,921549	2
13	0,002837	0,997163	0,997163	2
14	0,017916	0,982084	0,982084	2
15	0,002882	0,997118	0,997118	2
16	0,890521	0,109479	0,890521	1
17	0,007702	0,992298	0,992298	2
18	0,000186	0,999814	0,999814	2
19	0,000947	0,999053	0,999053	2
20	0,023263	0,976737	0,976737	2
21	0,098844	0,901156	0,901156	2
22	0,123336	0,876664	0,876664	2
23	0,994962	0,005038	0,994962	1
24	0,980322	0,019678	0,980322	1
25	0,158534	0,841466	0,841466	2
26	0,028807	0,971193	0,971193	2
27	0,001743	0,998257	0,998257	2
28	0,001638	0,998362	0,998362	2
29	$3,77 \times 10^5$	0,999962	0,999962	2
30	$6,02 \times 10^5$	0,99994	0,99994	2
31	0,996934	0,003066	0,996934	1
32	0,278591	0,721409	0,721409	2
33	0,950384	0,049616	0,950384	1
34	0,000437	0,999563	0,999563	2

Dapat dilihat juga dengan plot dua dimensi hasil pengelompokan dengan metode *Gaussian Mixture Model* sebagai berikut:



**Gambar 2.** Visualisasi Hasil Cluster Menggunakan *Gaussian Mixture Model* (GMM)

Berdasarkan **Gambar 2**, dapat dilihat nilai observasi berwarna merah adalah *cluster* 1, dan berwarna biru adalah *cluster* 2. Pada pembagian anggota untuk masing-masing *cluster* dapat dilihat pada **Tabel 3**.

**Tabel 3.** Keanggotaan *Cluster*

<i>Cluster</i>	Jumlah Anggota	Anggota
1	12	Aceh, Sumatera Utara, Sumatera Barat, Riau, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, Banten, Kalimantan Timur, Kalimantan Utara, Maluku dan Papua Barat
2	22	Jambi, Sumatera Selatan, Bengkulu, Lampung, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Sulawesi Selatan, Maluku Utara dan Papua

Setiap anggota dari *cluster* yang sudah dibentuk akan dianalisis secara individual untuk mengidentifikasi ciri khas dari masing-masing *cluster* dengan menghitung nilai rata-rata dari setiap variabel dalam setiap *cluster*. Rata-rata masing-masing variabel dihitung agar dapat melihat kecenderungan untuk setiap *cluster* di setiap variabel, hal ini dijelaskan sebagai berikut:

- Variabel  $X_4$  yaitu Garis Kemiskinan menunjukkan semakin tinggi garis kemiskinan maka semakin tinggi nilai minimum rupiah yang dibutuhkan. Semakin tinggi garis kemiskinan, maka semakin rendah persentase penduduk miskin.
- Variabel  $X_5$  yaitu Rasio Gini menunjukkan rasio gini berpengaruh positif terhadap kemiskinan, sehingga ketidaksetaraan ekonomi dapat memperburuk kemiskinan. pemerataan pendapatan juga dapat lebih banyak mengakses sumber daya seperti pendidikan, kesehatan dan perumahan. Ketika pendapatan merata, maka kesempatan meningkatkan kualitas hidup juga lebih merata.
- Variabel  $X_6$  yaitu Rata-rata Lama Sekolah menunjukkan bertambah tinggi tingkat pendidikan individu, maka ilmu dan kemampuannya pula meninggi dan mendorong peningkatan kapasitas kerja dan sumber daya manusia akan semakin baik dan dapat menekan angka kemiskinan.
- Variabel  $X_8$  yaitu Tingkat Pengangguran Terbuka menunjukkan pengangguran yang sangat erat kaitannya oleh kemiskinan, untuk beberapa masyarakat yang tak mempunyai pekerjaan tetap ataupun paruh waktu, mereka cenderung berada dalam kelompok yang sangat miskin. Ketika tingkat pengangguran meningkat, maka kemiskinan di masyarakat juga akan bertambah.

**Tabel 4.** Rata-rata setiap Variabel

<i>Cluster</i>	$X_4$	$X_5$	$X_6$	$X_8$	Ket.
1	<b>714831,8</b>	<b>0,321</b>	<b>9,954</b>	5,709	Kemiskinan Rendah
2	536703,5	0,357	8,948	<b>4,016</b>	Kemiskinan Tinggi

Berdasarkan **Tabel 4** diberi tanda bahwa rata-rata nilai terbaik untuk masing-masing variabel bertulis tebal. Kecenderungan nilai untuk masing-masing variabel disimpulkan bahwa *cluster* 1 mengindikasikan kemiskinan rendah dan *cluster* 2 mengindikasikan kemiskinan tinggi, sehingga diberikan interpretasi sebagai berikut:

- Cluster* 1 dicirikan dengan Garis Kemiskinan ( $X_4$ ) dan Rata-rata Lama Sekolah ( $X_6$ ) yang tinggi. Hal ini mengindikasikan bahwa pengeluaran minimum penduduk yang dibutuhkan setiap

*Pengelompokan Provinsi di Indonesia Menggunakan Gaussian Mixture Model Berdasarkan Indikator Kemiskinan* bulannya cukup tinggi, sehingga perolehan atau pendapatan penduduk cukup tinggi pada *cluster* 1 serta kualitas pendidikan pada *cluster* 1 tinggi dalam mendukung pengurangan angka kemiskinan, sehingga pada *cluster* 1 memiliki indikator kemiskinan yang rendah.

- b. *Cluster* 2 dicirikan dengan Tingkat Pengangguran Terbuka ( $X_8$ ) yang tinggi serta Rasio Gini ( $X_5$ ) yang tinggi. Hal ini mengindikasikan masih tingginya tingkat pengangguran yang ada pada *cluster* 2 dan ketimpangan pendapatan pada *cluster* 2 sudah cukup tinggi, yang berarti ada ketimpangan pendapatan antar penduduk di *cluster* 2, sehingga pada *cluster* 2 memiliki indikator kemiskinan yang tinggi.

#### 4 Kesimpulan

Berdasarkan uraian hasil dan pembahasan diperoleh 2 *cluster* provinsi di Indonesia yakni pada garis kemiskinan yang tertinggi memiliki ciri tingginya rata-rata durasi pendidikan dan tingginya rasio gini hingga taraf pengangguran yang tinggi dalam indikator kemiskinan yang terbentuk dengan algoritma *Gaussian Mixture Model*. Adapun secara lengkap ke 2 klaster yang terbentuk berdasarkan indikator kemiskinan di 34 provinsi di Indonesia terdiri dari: (1) Klaster 1 (indikator kemiskinan rendah) yang ditempati oleh Aceh, Sumatera Utara, Sumatera Barat, Riau, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, Banten, Kalimantan Timur, Kalimantan Utara, Maluku dan Papua Barat, (2) *Cluster* 2 (indikator kemiskinan tinggi) yang ditempati oleh Jambi, Sumatera Selatan, Bengkulu, Lampung, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, Nusa Tenggara Barat, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku Utara dan Papua.

#### Referensi

- [1] S. W. Mustamin and Abdillah, *Kemiskinan Kota (Tinjauan Ekonomi Makro terhadap Kemiskinan di Kota Makassar)*. Makassar: Universitas Muhammadiyah Makassar, 2021.
- [2] A. Suryahadi and S. Sumarto, *The Chronic Poor, the Transient Poor, and the Vulnerable in Indonesia Before and After the Crisis*. Jakarta: The SMERU Research Institute, 2001.
- [3] R. Purwono, W. W. Wardana, T. Haryanto, and M. Khoerul Mubin, "Poverty dynamics in Indonesia: empirical evidence from three main approaches," *World Dev Perspect*, vol. 23, pp. 1–9, Sep. 2021, doi: 10.1016/j.wdp.2021.100346.
- [4] L. A. Pujiwati, "Masyarakat Indonesia masih rentan jatuh miskin, apa penyebabnya?," *The Conversation*, Feb. 06, 2023. Accessed: Jan. 04, 2024. [Online]. Available: <https://theconversation.com/masyarakat-indonesia-masih-rentan-jatuh-miskin-apa-penyebabnya-198679>
- [5] Badan Pusat Statistik, "Penjelasan Data Kemiskinan," Jakarta, Jan. 2011.
- [6] Badan Pusat Statistik, "Persentase Penduduk Miskin (P0) Menurut Provinsi dan Daerah (Persen) 2023," Bada Pusat Statistik. Accessed: Feb. 09, 2024. [Online]. Available: <https://www.bps.go.id/id/statistics-table/2/MTkyIzI=/persentase-penduduk-miskin--maret-2023.html>
- [7] Liputan6, "Angka Kemiskinan Turun di Mei 2023, Tapi Belum Pulih ke Level Sebelum Pandemi," Liputan6.com. Accessed: Dec. 05, 2023. [Online]. Available: <https://www.liputan6.com/bisnis/read/5346506/angka-kemiskinan-turun-di-mei-2023-tapi-belum-pulih-ke-level-sebelum-pandemi>
- [8] A. Purwanti, "Tantangan Mengentaskan Penduduk Miskin di Tengah Krisis," *kompas.id*, Jun. 23, 2021. Accessed: Jan. 04, 2024. [Online]. Available: <https://www.kompas.id/baca/riset/2021/06/23/tantangan-mengentaskan-penduduk-miskin-di-tengah-krisis/>
- [9] B. Sumargo, *Validitas dan Reliabilitas Pengukuran Kemiskinan*. Bogor: Institut Pertanian Bogor, 2002.
- [10] E. Widodo, P. Ermayani, L. N. Laila, and A. T. Madani, "Pengelompokan Provinsi di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Analisis Hierarchical Agglomerative Clustering,"

- Seminar Nasional Official Statistics*, vol. 2021, no. 1, pp. 557–566, Nov. 2021, doi: 10.34123/semnasoffstat.v2021i1.971.
- [11] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, “Multivariate Data Analysis Eighth Edition,” 2019. [Online]. Available: [www.cengage.com/highered](http://www.cengage.com/highered)
- [12] K. Backhaus, B. Erichson, S. Gensler, R. Weiber, and T. Weiber, *Multivariate Analysemethoden*, 17th ed. German: Springer Gabler, 2023. [Online]. Available: <https://ashcards.springernature.com/login>
- [13] C. M. Bishop, *Pattern Recognition and Machine Learning*. UK: Springer, 2006.
- [14] P. S. Mann, *Introductory Statistics (7th Ed)*. New York: John Wiley & Sons, 2010.
- [15] R. Arnold. Johnson and D. W. Wichern, *Applied multivariate statistical analysis*. Harlow: Pearson, 2014.
- [16] G. J. McLachlan and K. E. Basford, *Mixture Model*. New York: MARCEL DEKKER, 1988.
- [17] M. P. Deisenroth, A. , A. Faisal, and C. S. Ong, *Mathematics for machine learning*. UK: Cambridge University Press, 2020.
- [18] F. Novkaniza, Nico, and R. Al Kafi, “Pemodelan Jumlah Kasus Baru Harian Covid-19 Di Indonesia Menggunakan Gaussian Mixture Model,” *Jurnal Riset dan Aplikasi Matematika*, vol. 07, no. 02, pp. 116–127, Oct. 2023.
- [19] Supranto, J. *Analisis Multivariat Arti & Intepretasi*. Rineka Cipta. 2010.



© **The Author(s) 2024**. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License. Editorial of Journal of Mathematics: Theory and Applications, Department of Mathematics, Universitas Sulawesi Barat, Jalan Prof. Dr. Baharuddin Lopa, S.H., Talumung, Majene 91412, Sulawesi Barat.