

Penerapan *Gaussian Mixture Model* pada Klasterisasi Desa di Kalimantan Barat berdasarkan Indeks Desa Membangun Tahun 2021

Rizkia Narulita*, Abdul Kudus

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

* rizkianarulita74@gmail.com, abdul.kudus@unisba.ac.id

Abstract. West Kalimantan is a province that actively supports the 2021 developing village index (IDM). Based on the 2021 West Kalimantan IDM data, the constituent indexes have various values and there are certain patterns. Therefore, clustering method is needed to determine groups/clusters consisting of several villages. By considering the Gaussian Mixture Model method using the expectation maximization algorithm. The purpose of this study was to determine the optimal number of clusters, and to find out the results of clustering on IDM West Kalimantan 2021 data by applying the Gaussian Mixture Model. The results of the study obtained that the optimal number of clusters was $K=7$, with a BIC value of 9979.939 and p -value BLRT is 0.498. Meanwhile, the results of the clustering are: (1) cluster 1 there were 436 villages with advanced social security status, very lagging economic resilience, independent environmental resilience; (2) cluster 2 there are 854 villages with advanced social security status, lagging economic resilience, developing environmental resilience; (3) cluster 3 there are 455 villages with independent social and environmental resilience status, advanced economic resilience; (4) cluster 4 contains 66 villages with independent social and economic resilience status, developing environmental resilience; (5) cluster 5 there are 106 villages with developing social and environmental resilience status, economic resilience is very lagging; (6) cluster 6 there are 64 villages with advanced social security status, lagging economic resilience, very lagging environmental resilience; (7) cluster 7 there are 50 villages with developing social security status, very lagging economic resilience, advanced environmental resilience.

Keywords: *IDM, Gaussian Mixture Model.*

Abstrak. Kalimantan Barat merupakan provinsi yang aktif mendukung indeks desa membangun (IDM) 2021. Berdasarkan data IDM Kalimantan Barat 2021, indeks penyusun memiliki nilai yang beragam dan terdapat pola tertentu. Sehingga diperlukan metode klasterisasi untuk menentukan kelompok/klaster yang terdiri dari beberapa desa. Dengan pertimbangan metode *Gaussian Mixture Model* menggunakan penduga algoritme *expectation maximization*. Tujuan penelitian ini adalah untuk mengetahui jumlah klaster optimal, dan mengetahui hasil klasterisasi pada data IDM Kalimantan Barat 2021 dengan menerapkan *Gaussian Mixture Model*. Adapun hasil penelitian diperoleh jumlah klaster optimal $K=7$, dengan nilai BIC sebesar 9979,939 dan p -value BLRT sebesar 0,498. Sedangkan untuk hasil klasterisasinya, yaitu: (1) klaster 1 terdapat 436 desa dengan status ketahanan sosial maju, ketahanan ekonomi sangat tertinggal, ketahanan lingkungan mandiri; (2) klaster 2 terdapat 854 desa dengan status ketahanan sosial maju, ketahanan ekonomi tertinggal, ketahanan lingkungan berkembang; (3) klaster 3 terdapat 455 desa dengan status ketahanan sosial dan lingkungan mandiri, ketahanan ekonomi maju; (4) klaster 4 terdapat 66 desa dengan status ketahanan sosial dan ekonomi mandiri, ketahanan lingkungan berkembang; (5) klaster 5 terdapat 106 desa dengan status ketahanan sosial dan lingkungan berkembang, ketahanan ekonomi sangat tertinggal; (6) klaster 6 terdapat 64 desa dengan status ketahanan sosial maju, ketahanan ekonomi tertinggal, ketahanan lingkungan sangat tertinggal; (7) klaster 7 terdapat 50 desa dengan status ketahanan sosial berkembang, ketahanan ekonomi sangat tertinggal, ketahanan lingkungan maju.

Kata Kunci: *IDM, Gaussian Mixture Model.*

A. Pendahuluan

Kalimantan Barat merupakan salah satu provinsi di Indonesia yang berperan aktif dalam mendukung pengukuran indeks desa membangun. Salah satunya dengan menjadi provinsi tercepat secara nasional dalam menyampaikan laporan pemutakhiran data indeks desa membangun (IDM) tahun 2021. Selain itu, Kalimantan Barat juga memberikan tanggapan baik atas program pengukuran indeks desa membangun (IDM) tahun 2021. Hal tersebut dikarenakan Kalimantan Barat berhasil menaikkan posisi dari peringkat 15 (IDM tahun 2020) menjadi peringkat 6 (IDM tahun 2021) dari 33 provinsi di Indonesia. Kalimantan Barat juga menaikkan status IDM di tahun 2021 menjadi “maju”, yang mana di tahun sebelumnya status IDM provinsi ini adalah “berkembang”. Sebagaimana dikemukakan oleh Gubernur Kalimantan Barat bahwa penyampaian laporan mutakhir data indeks desa membangun (IDM) tahun 2021 merupakan capaian keberhasilan dalam meningkatkan desa mandiri di Indonesia (1).

Dalam hal ini, data indeks desa membangun (IDM) 2021 pada provinsi Kalimantan Barat memberikan hasil yang signifikan naik. Hal tersebut, tidak terlepas dari indeks penyusun dari data indeks desa membangun (IDM). Apabila dilihat dari data indeks desa membangun (IDM) 2021 diketahui bahwa, ketiga indeks penyusun memiliki nilai yang relatif beragam dan memungkinkan adanya pola tertentu dari data. Salah satu metode data mining yang dapat digunakan adalah klusterisasi (clustering), dimana metode tersebut memungkinkan pembentukan beberapa kelompok/klaster yang terdiri atas beberapa desa di Kalimantan Barat. Terdapat beberapa metode klusterisasi dalam kepustakaan. Di antaranya adalah metode *k-means* dan *Gaussian Mixture Model*.

Kebanyakan penelitian metode klusterisasi umumnya sering menggunakan metode *k-means* yang berbasis jarak (*distance-based model*) karena relatif lebih terukur, efisien untuk pengolahan objek dalam jumlah besar dan tidak terpengaruh terhadap urutan objek. Akan tetapi, *k-means* mempunyai kelemahan jika diterapkan pada data yang pengelompokannya mengikuti pola sirkular. Oleh karena itu, perlu dipertimbangkan untuk menggunakan metode alternatif yang berbasis fungsi distribusi (*distribution-based model*), yakni *Gaussian Mixture Model* dengan penduga algoritme *expectation maximization*.

Berdasarkan uraian di atas, maka tujuan dalam penelitian ini adalah:

1. Mengetahui jumlah klaster optimal pada *Gaussian Mixture Model* klusterisasi desa di Kalimantan Barat berdasarkan data indeks desa membangun 2021.
2. Mengetahui hasil klusterisasi desa di Kalimantan Barat pada penerapan *Gaussian Mixture Model* berdasarkan data indeks desa membangun 2021.

B. Metodologi Penelitian

Peneliti menggunakan metode *Gaussian Mixture Model* dengan penduga algoritme *expectation maximization* dan menerapkan data indeks desa membangun 2021 di Kalimantan Barat. Populasi penelitian sebanyak 2031 desa di Kalimantan Barat. Data penelitian diperoleh dari Dinas Komunikasi dan Informatika Provinsi Kalimantan Barat (2).

Gaussian Mixture Model - Multivariat

Gaussian Mixture Model pertama kali dikenalkan oleh Stauffer dan Grimson (3). *Gaussian Mixture Model*, diartikan sebagai suatu himpunan komponen model fungsi kepadatan/densitas (density) yang terdiri dari k buah komponen dengan $k \in \{1, \dots, K\}$. Artinya, fungsi densitas untuk *Gaussian Mixture Model* merupakan fungsi densitas gaussian yang memiliki $k \geq 2$. Jika jumlah variabelnya banyak, maka akan merupakan fungsi densitas gaussian multivariat. Berdasarkan Bishop dalam bukunya (4), fungsi densitas *Gaussian Mixture Model* - multivariat dirumuskan sebagai berikut:

$$p(x_i | \pi, \mu, \Sigma) = \sum_{k=1}^K \pi_k \mathcal{N}(x_i | \mu_k, \Sigma_k), \quad i = 1, 2, \dots, n \quad (1)$$

Keterangan:

x_i : Nilai titik objek dalam dataset

μ_k : Vektor rata-rata dari komponen mixture ke- k

Σ_k : Matriks kovarians dari komponen mixture ke- k

π_k : Koefisien mixture ke- k

Metode Penaksir Maksimum Likelihood

Bain dan Engelhardt (5) mendefinisikan metode maksimum likelihood sebagai salah satu cara untuk menaksir parameter yang tidak diketahui (*unknown parameter*). Dari persamaan rumus (2.1) ditaksir menggunakan maksimum likelihood, sehingga menghasilkan persamaan rumus sebagai berikut:

$$L(\Psi) = \prod_{i=1}^n \sum_{k=1}^K \pi_k \mathcal{N}(x_i | \mu_k, \Sigma_k) \quad (2)$$

$$\ln L(\psi) = \sum_{i=1}^n \ln \sum_{k=1}^K \pi_k \mathcal{N}(x_i | \mu_k, \Sigma_k) \quad (3)$$

Persamaan di atas diturunkan terhadap ψ . Akan tetapi, persamaan ini tidak mudah diturunkan baik secara analitik maupun secara pendekatan numerik dikarenakan persamaan fungsi likelihood pada persamaan di atas tidak dalam bentuk *closed-form*. Dengan demikian, salah satu cara untuk mengatasi masalah, yakni menggunakan algoritme *expectation maximization*.

Algoritme Expectation Maximization untuk menduga GMM

Dalam bukunya Ng, Xiang dan Yau (6) Algoritme Expectation Maximization untuk menduga GMM dijelaskan sebagai berikut.

Pada iterasi ($k+1$) algoritme EM, *E-Step* menghitung fungsi Q , yang merupakan ekspektasi dari data lengkap log likelihood yang dikondisikan pada x yang diamati dengan menggunakan kecocokan (*fit*) saat ini untuk Ψ . Hal tersebut dapat dituliskan dalam persamaan sebagai berikut:

$$Q(\Psi, \Psi^{(k)}) = \sum_{k=1}^K \sum_{i=1}^n \tau_k(x_i, \Psi^{(k)}) \{ \ln \pi_k + \ln \mathcal{N}(x_i | \mu_k, \Sigma_k) \} \quad (4)$$

Dimana $E_{\Psi^{(k)}}$ menunjukkan harapan menggunakan vektor parameter saat ini ($\Psi^{(k)}$). Adapun $\tau_k(x_i, \Psi^{(k)})$ dituliskan dalam persamaan sebagai berikut:

$$\tau_k(x_i, \Psi^{(k)}) = \frac{\pi_k^{(k)} \ln \mathcal{N}(x_i | \mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j^{(k)} \ln \mathcal{N}(x_i | \mu_j, \Sigma_j)} \quad (5)$$

$\tau_k(x_i, \Psi^{(k)})$ merupakan estimasi probabilitas posterior bahwa pengamatan ke- i (x_i) termasuk ke dalam komponen ke- k dari model campuran (*mixture model*) berdasarkan $\Psi^{(k)}$ ($k = 1, \dots, K; i = 1, \dots, n$). Untuk campuran dengan densitas komponen normal, secara komputasi menguntungkan untuk bekerja dalam *E-Step* dalam hal statistik yang memadai (Ng dan McLachlan, 2003b) yang memberikan persamaan sebagai berikut:

$$T_{k1}^{(k)} = \sum_{i=1}^n \tau_k(x_i | \Psi^{(k)}) \quad (6)$$

$$T_{k2}^{(k)} = \sum_{i=1}^n \tau_k(x_i | \Psi^{(k)}) x_i$$

$$T_{k3}^{(k)} = \sum_{i=1}^n \tau_k(x_i | \Psi^{(k)}) x_i x_i^T$$

Selanjutnya *M-Step* akan memperbaharui estimasi dari Ψ menjadi $\Psi^{(k+1)}$ yang memaksimalkan fungsi Q terhadap Ψ di atas parameter *space*. Untuk campuran dengan densitas komponen normal, *M-Step* terdapat dalam *closed-form*. Oleh karena dasar dari statistik cukup pada persamaan (2.30), maka pembaruan parameter dalam *M-Step* memberikan persamaan sebagai berikut:

$$\pi_k^{(k+1)} = T_{k1}^{(k)} / n \tag{7}$$

$$\mu_k^{(k+1)} = T_{k2}^{(k)} / T_{k1}^{(k)}$$

$$\Sigma_k^{(k+1)} = \{T_{k3}^{(k)} - T_{k1}^{(k)-1} T_{k2}^{(k)} T_{k2}^{(k)T}\} / T_{k1}^{(k)}$$

Pemilihan Jumlah Kluster Optimal

Mengingat banyaknya komponen *K* dalam GMM perlu dicari berapa besarnya, maka kita dapat menggunakan *Bayesian Information Criteria* (BIC). Nilai BIC dirumuskan sebagai berikut:

$$BIC = q \ln(n) - \ln L_C(\Psi) \tag{8}$$

Diketahui *q* adalah banyaknya parameter dalam model, *n* adalah banyaknya data dan $L_C(\Psi)$ adalah fungsi likelihood dari penduga maksimum likelihood Ψ . Pemilihan jumlah kluster optimal ditentukan dari nilai BIC terbesar dari suatu kluster di antara nilai BIC kluster lainnya.

Selain menggunakan BIC, pemilihan jumlah kluster optimal juga dapat dilakukan dengan *bootstrap likelihood ratio test*. *Bootstrap likelihood ratio test* dirumuskan pada persamaan (2.9).

$$p - value = \frac{\{1 + \#LRT_b^a \geq LRTS_{obs}\}}{B + 1} \tag{9}$$

Diketahui *B* adalah jumlah bootstrap sampel, $LRTS_{obs}$ adalah LRTS yang dihitung dari data pengamatan, dan LRT_b^a adalah LRTS yang dihitung dari bootstrap sampel ke-*b*. Adapun jumlah kluster optimal apabila nilai dari *p*-value suatu kluster lebih besar dari nilai *alpha*. Nilai *p*-value menggunakan statistik rasio kemungkinan bootstrap untuk *k* versus *k*+1.

C. Hasil Penelitian dan Pembahasan

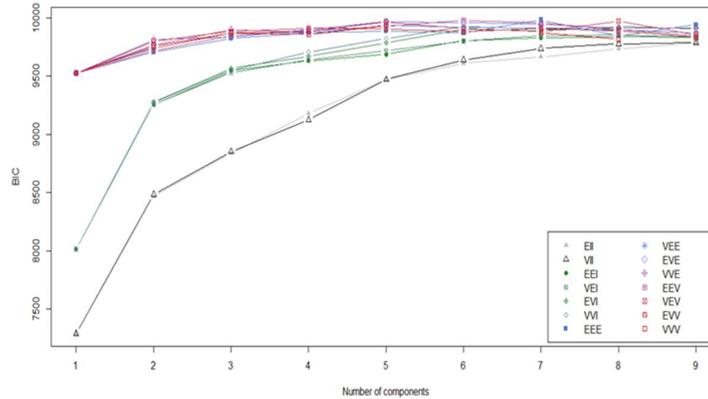
Pemilihan Jumlah Kluster Optimal

Sebelum melakukan klasterisasi berdasarkan data indeks desa membangun 2021 pada desa di Kalimantan Barat dengan menerapkan *Gaussian Mixture Model*, perlu ditetapkan jumlah kluster optimal yang akan digunakan. Menentukan jumlah kluster yang optimal, diketahui dengan mendapatkan nilai BIC serta *p*-value dari *bootstrap likelihood ratio test*. Berikut pada gambar 1 dan gambar 2 disajikan hasil output dari nilai BIC.

```
> dataBIC
Bayesian Information Criterion (BIC):
      EII      VII      EEI      VEI      EVI
1 7292.108 7292.108 8015.831 8015.831 8015.831
2 8464.589 8485.837 9252.395 9252.890 9275.096
3 8835.335 8851.515 9543.666 9528.479 9565.207
4 9177.197 9124.090 9629.562 9639.660 9670.095
5 9465.678 9470.050 9685.221 9716.808 9784.398
6 9609.589 9637.495 9803.293 9798.008 9901.470
7 9663.589 9735.101 9825.513 9843.751 9887.095
8 9731.165 9775.289 9840.431 9852.408 9921.453
9 9784.974 9786.175 9828.668 9836.330 9906.960
      VVI      EEE      VEE      EVE      VVE
1 8015.831 9523.975 9523.975 9523.975 9523.975
2 9271.470 9706.351 9740.469 9811.449 9766.519
3 9549.637 9824.680 9867.488 9830.692 9897.925
4 9704.176 9870.046 9869.411 9898.180 9878.734
5 9820.918 9886.904 9934.970 9964.834 9970.620
6 9921.898 9871.340 9917.980 9959.891      NA
7 9903.490 9979.939 9914.060 9943.714 9940.670
8 9893.235 9843.761 9897.740      NA 9911.377
9 9870.838 9939.358 9869.765      NA 9909.351
      EEV      VEV      EVV      VVV
1 9523.975 9523.975 9523.975 9523.975
2 9715.901 9739.980 9799.282 9758.115
3 9846.340 9871.523 9880.008 9868.147
4 9889.082 9885.733 9910.375 9857.140
5 9924.871 9963.014 9904.298 9935.235
6 9978.636 9905.228 9881.147      NA
7 9954.407 9884.253 9905.587 9871.946
8 9895.537 9971.327 9885.997 9814.177
9 9848.568 9852.693 9831.472      NA

Top 3 models based on the BIC criterion:
      EEE,7      EEV,6      VEV,8
9979.939 9978.636 9971.327
```

Gambar 1. Nilai BIC



Gambar 2. Plot Nilai BIC

Dari gambar 1 dan gambar 2, dapat diketahui bahwa model terbaik berdasarkan kriteria BIC untuk data penelitian adalah $K=7$ dengan model EEE. Hal tersebut dikarenakan nilai BIC pada $K=7$ dan model EEE terbesar di antara kluster dan model kluster lainnya, yakni sebesar 9979,939. Selanjutnya, berikut disajikan gambar 3 hasil output dari *bootstrap likelihood ratio test*.

```
> mclustBootstrapLRT(data, modelName = "EEE", nboot = 999, lev
e1 = 0.05, maxG = NULL,
+ verbose = interactive())
bootstrapping LRTS ...
|=====| 100%
-----
Bootstrap sequential LRT for the number of mixture components
-----
Model = EEE
Replications = 999
LRTS bootstrap p-value
1 vs 2 212.935665 0.001
2 vs 3 148.636334 0.001
3 vs 4 75.322276 0.001
4 vs 5 47.981821 0.001
5 vs 6 124.414254 0.001
6 vs 7 29.405509 0.001
7 vs 8 1.148664 0.498
```

Gambar 3. p-value Bootstrap Likelihood Ratio Test

Berdasarkan hasil output pada gambar 3, diketahui bahwa p-value *bootstrap likelihood ratio test* $> 0,05$ (*alpha*) ketika jumlah kluster adalah 7. Sehingga jumlah kluster optimal untuk data penelitian ini ditetapkan sebanyak 7 atau $K \in \{1,2,3,\dots,7\}$.

Klasterisasi menggunakan GMM

Dari 2031 desa pada data indeks desa membangun (IDM) tahun 2021 di Kalimantan Barat akan dilakukan klasterisasi. Jumlah kluster optimal ditetapkan sebanyak 7 atau $K \in \{1,2,3,\dots,7\}$. Berikut disajikan tabel 1 hasil klasterisasi data indeks desa membangun (IDM) 2021 pada desa di Kalimantan Barat dengan menggunakan *Gaussian Mixture Model*.

Tabel 1. Hasil Klusterisasi Data IDM Kalbar 2021 menggunakan GMM

No	NAMA. KABUPATEN	NAMA. KECAMATAN	NAMA.DESA	Klaster
1	KABUPATEN SAMBAS	SAMBAS	DALAM KAUM	3
2	KABUPATEN SAMBAS	SAMBAS	LUBUK DAGANG	2
3	KABUPATEN SAMBAS	SAMBAS	TANJUNG BUGIS	4
4	KABUPATEN SAMBAS	SAMBAS	PENDAWAN	3
5	KABUPATEN SAMBAS	SAMBAS	PASAR MELAYU	3
6	KABUPATEN SAMBAS	SAMBAS	DURIAN	4
7	KABUPATEN SAMBAS	SAMBAS	LORONG	1
8	KABUPATEN SAMBAS	SAMBAS	JAGUR	3
9	KABUPATEN SAMBAS	SAMBAS	TUMUK MANGGIS	3
10	KABUPATEN SAMBAS	SAMBAS	TANJUNG MEKAR	2
⋮	⋮	⋮	⋮	⋮
2031.	KABUPATEN KUBU RAYA	SUNGAI KAKAP	PUNGGUR KAPUAS	1

Berdasarkan hasil klusterisasi di tabel 1 dengan menggunakan *Gaussian Mixture Model* membentuk 7 klaster dari 2031 desa, yakni klaster 1 terdapat 436 desa, klaster 2 terdapat 854 desa, klaster 3 terdapat 455 desa, klaster 4 terdapat 66 desa, klaster 5 terdapat 106 desa, klaster 6 terdapat 64 desa, dan klaster 7 terdapat 50 desa.

Interpretasi Hasil Klusterisasi menggunakan GMM

Dalam menginterpretasi isi dari masing-masing klaster menggunakan penerapan *Gaussian Mixture Model*, diperlukan bantuan metode k-means pada inisialisasi parameter, algoritme *expectation maximization* sebagai metode iteratif mencari taksiran maksimum likelihood, dan metode maksimum log likelihood. Adapun penelitian ini ditetapkan jumlah klaster $K \in \{1,2,3,\dots,7\}$. Sehingga, dengan bantuan *software* R didapatkan hasil output nilai estimasi parameter π_k , μ_k dan Σ_k yang merupakan parameter sebaran Gaussian untuk masing-masing klaster. Estimasi parameternya disajikan sebagai berikut:

Untuk klaster 1,

$$K_1 \sim N \left(\begin{matrix} \pi_1 = (0,21060116), \mu_1 = \begin{pmatrix} 0,7545661 \\ 0,4971082 \\ 0,8479567 \end{pmatrix}, \\ \Sigma_1 = \begin{pmatrix} 0,0033501039 & 0,0031777277 & -0,0004257580 \\ 0,003177728 & 0,0128658727 & -0,0007309489 \\ -0,000425758 & -0,0007309489 & 0,0037023886 \end{pmatrix} \end{matrix} \right)$$

Dan seterusnya sampai dengan klaster 7.

Selain parameter π_k , μ_k dan Σ_k , didapatkan pula output τ_k yang memberikan nilai estimasi probabilitas posterior. Adapun suatu pengamatan ke-i (x_i) termasuk ke dalam klaster ke-k, jika nilai estimasi probabilitas posterior di klaster tersebut memiliki nilai terbesar di antara nilai estimasi probabilitas posterior di klaster lainnya. Berikut di bawah ini disajikan tabel 2 nilai estimasi probabilitas posterior setiap klaster.

Tabel 2. Nilai Estimasi Parameter τ_k

No	<i>Posterior Probability</i>						
	τ_k						
	1	2	3	4	5	6	7
1	6.057307 e-04	1.067795 e-08	9.991762 e-01	2.180307 e-04	1.223240 e-13	4.742590 e-19	8.967331 e-09
2	6.308533 e-03	7.414009 e-01	8.976133 e-04	2.431969 e-01	7.603039 e-03	1.052705 e-04	4.877499 e-04
3	2.248059 e-03	2.194173 e-01	2.573033 e-03	7.753762 e-01	3.399315 e-04	1.938008 e-05	2.610438 e-05
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
203 1	7.585161 e-01	4.431306 e-02	1.694215 e-01	2.118249 e-02	1.005859 e-04	4.868386 e-09	6.466271 e-03

*No menunjukkan nomor urutan desa

Berdasarkan dari tabel di atas, dapat diketahui bahwa pada desa urutan ke 1 (Desa Dalam Kaum) berada pada kluster 3. Hal ini dikarenakan nilai estimasi probabilitas posterior di kluster 3 merupakan nilai terbesar di antara nilai estimasi probabilitas posterior kluster lainnya. Adapun proses klusterisasi ini berlaku untuk desa urutan ke 2 (Desa Lubuk Kaum) sampai desa urutan ke 2031 (Desa Punggur Kapuas).

Setelah 2031 desa di Kalimantan Barat diketahui termasuk anggota kluster mana, maka selanjutnya dilihat karakteristik dari masing-masing keanggotaan klasternya yang sudah terbentuk. Adapun salah satu caranya dengan melihat nilai rata-rata masing-masing variable (indeks penyusun IDM tahun 2021). Adapun selengkapannya disajikan pada tabel 3.

Tabel 3. Karakteristik Setiap Kluster

Nomor_Kluster	Rata-rata IKS	Rata-rata IKE	Rata-rata IKL
1	0,751	0,489	0,854
2	0,766	0,528	0,638
3	0,866	0,751	0,901
4	0,903	0,854	0,691
5	0,619	0,453	0,608
6	0,765	0,564	0,432
7	0,611	0,408	0,797

Berdasarkan tabel 3 karakteristik pada kluster 1 sampai 7 dapat diinterpretasikan klasifikasi statusnya dari ketiga indeks penyusun data IDM di masing-masing kluster, sebagai berikut:

1. Kluster 1 memiliki status ketahanan sosial maju, ketahanan ekonomi sangat tertinggal, ketahanan lingkungan mandiri.
2. Kluster 2 memiliki status ketahanan sosial maju, ketahanan ekonomi tertinggal, ketahanan lingkungan berkembang.
3. Kluster 3 memiliki status ketahanan sosial dan lingkungan mandiri, ketahanan ekonomi maju.
4. Kluster 4 memiliki status ketahanan sosial dan ekonomi mandiri, ketahanan lingkungan berkembang
5. Kluster 5 memiliki status ketahanan sosial dan lingkungan berkembang, ketahanan ekonomi sangat tertinggal
6. Kluster 6 memiliki status ketahanan sosial maju, ketahanan ekonomi tertinggal, ketahanan lingkungan sangat tertinggal
7. Kluster 7 memiliki status ketahanan sosial berkembang, ketahanan ekonomi sangat tertinggal, ketahanan lingkungan maju.

D. Kesimpulan

Berdasarkan pembahasan dalam penelitian ini, peneliti menyimpulkan beberapa hasil penelitian sebagai berikut:

1. Pada pemilihan jumlah kluster optimal berdasarkan indeks desa membangun (IDM) 2021 pada desa di Kalimantan Barat menggunakan *Gaussian Mixture Model*, terpilih jumlah kluster optimalnya yaitu sebanyak kluster maksimal $K=7$. Adapun pemilihan jumlah kluster optimal berdasarkan nilai BIC sebesar 9979,939 dan p-value bootstrap likelihood ratio test sebesar 0,498.
2. Hasil klusterisasi IDM 2021 pada desa di Kalimantan Barat menggunakan *Gaussian Mixture Model*, yaitu pada kluster 1 terdapat 436 desa dengan status ketahanan sosial maju, ketahanan ekonomi sangat tertinggal, ketahanan lingkungan mandiri, kluster 2 terdapat 854 desa dengan status ketahanan sosial maju, ketahanan ekonomi tertinggal, ketahanan lingkungan berkembang, kluster 3 terdapat 455 desa dengan status ketahanan sosial dan lingkungan mandiri, ketahanan ekonomi maju, kluster 4 terdapat 66 desa dengan status ketahanan sosial dan ekonomi mandiri, ketahanan lingkungan berkembang, kluster 5 terdapat 106 desa dengan status ketahanan sosial dan lingkungan berkembang, ketahanan ekonomi sangat tertinggal, kluster 6 terdapat 64 desa dengan status ketahanan sosial maju, ketahanan ekonomi tertinggal, ketahanan lingkungan sangat tertinggal, dan kluster 7 terdapat 50 desa dengan status ketahanan sosial berkembang, ketahanan ekonomi sangat tertinggal, ketahanan lingkungan maju.

Acknowledge

Penulis mengucapkan terima kasih kepada Bapak Abdul Kudus, S.Si., M.Si. Ph.D. selaku dosen pembimbing yang telah memberikan bimbingan, masukan dan wawasan baru yang sangat membantu dalam menyelesaikan penelitian ini. Penulis juga mengucapkan terima kasih kepada dosen-dosen pengampu mata kuliah pada jurusan statistika untuk segala ilmu yang diberikan untuk turut membantu penyusunan penelitian ini.

Daftar Pustaka

- [1] Sutiana, W. (2021). *Penyampaian Laporan Data IDM Desa Mandiri Tahun 2021, Kalbar Menjadi Provinsi Tercepat Nasional* (Online). (<https://kalbarprov.go.id/berita/penyampaian-laporan-data-idm-desa-mandiri-tahun-2021-kalbar-menjadi-provinsi-tercepat-nasional.html>, diakses 1 Juli 2022)
- [2] Dinas Komunikasi dan Informatika Provinsi Kalimantan Barat. (2021). *Data Indeks Desa Membangun Provinsi Kalimantan Barat Tahun 2021* (Online). (<http://data.kalbarprov.go.id/dataset/data-indeks-desa-membangun-provinsi-kalimantan-barat-tahun-2021>, diakses 29 Juni 2022)
- [3] Murinto, et al. (2001). *Deteksi Jenis Warna Kulit Wajah untuk Klasifikasi Ras Manusia menggunakan Transformasi Warna*. Yogyakarta: Universitas Ahmad Dahlan.
- [4] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Singapore: Springer.
- [5] Bain, L. J. & Engelhardt, M. (1992). *Introduction to Probability and Mathematical Statistics*. California: Wadsworth Publishing Company.
- [6] Ng, S.K, L. Xiang and K.K.W Yau. (2019). *Mixture Modelling for Medical and Health Sciences*. CRC Press, NY, USA.
- [7] Khoeriyah, Risti Yulianti. (2021). *Regresi Terboboti Geografis Semiparametrik (RTG-S) untuk Pemodelan Indeks Pembangunan Kesehatan Masyarakat Kabupaten/Kota di Sumatera Utara*, Jurnal Riset Statistika, 1(1), 43-50.