

**PENGELOMPOKAN PROVINSI DI INDONESIA
BERDASARKAN KARAKTERISTIK KESEJAHTERAAN RAKYAT
MENGUNAKAN METODE *K-MEANS CLUSTER***

Fitra Ramdhani¹, Abdul Hoyyi², Moch. Abdul Mukid³

¹Mahasiswa Jurusan Statistika FSM UNDIP

^{2,3}Staf Pengajar Jurusan Statistika FSM UNDIP

ABSTRACT

Welfare have a relative explanation, dynamic, and quantitative. Quantitative formulation of welfare is never final because it will continue to evolve along with the development needs of human life. In 2011, the National Team for the Acceleration of Poverty Reduction (NTAPR) made priority sector that can serve as a benchmark the welfare in a region. From the priority sector will be made cluster or group which contains all 33 provinces based on the level of public welfare in the region uses data in 2012 were sourced from the Central Statistics Agency (CSA). The method that can be used to group the 33 provinces is K-Means Cluster method with number cluster as many as two, three, four, and five clusters. K-Means Cluster method is one of cluster analysis method who can partition the data into one or more clusters, so that the data with the same characteristics are grouped into the same cluster and data with different characteristics grouped into other clusters. To know the most optimal of the number of clusters we use Davies-Bouldin Index (DBI). We concluded that the optimal number of cluster is three with details the province in the first clusters have superiority in four sectors like net enrollment rate of primary school, net enrollment rate of junior high school, IMR (Infant Mortality Rate), and access to electricity. The province in the second clusters have superiority in one sector, that is open unemployment rate. The province in the third clusters have superiority in all sectors.

Keywords: Welfare, NTAPR Priority Sector, K-Means Cluster Method, Davies- Bouldin Index (DBI)

1. PENDAHULUAN

Kesejahteraan mengandung pengertian yang relatif, dinamis, dan kuantitatif. Rumusnya tidak pernah final karena akan terus berkembang seiring dengan perkembangan kebutuhan hidup manusia. Secara umum kesejahteraan dapat diartikan sebagai suatu keadaan dimana segenap warga negara selalu berada dalam kondisi serba kecukupan segala kebutuhannya, baik material maupun spiritual (Roestam, 1993).

Pada tahun 2011, Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K) membuat bidang prioritas kesejahteraan rakyat di suatu daerah yang berlaku di provinsi manapun di Indonesia. Bidang-bidang prioritas tersebut dibuat sebagai tolak ukur kesejahteraan rakyat secara keseluruhan disamping bidang lainnya. Enam bidang prioritas diantaranya adalah angka partisipasi murni SD atau MI, angka partisipasi murni SMP atau MTs, angka kematian bayi per 1.000 kelahiran hidup, akses terhadap air bersih, akses terhadap listrik, dan tingkat pengangguran terbuka.

Dari keenam bidang prioritas tersebut akan dibuat klaster yang berisi 33 provinsi berdasarkan kemiripan karakteristik kesejahteraan rakyat. Proses pengklasteran dilakukan dengan metode *K-Means Cluster* dengan menggunakan jumlah klaster sebanyak dua, tiga, empat, dan lima. Jumlah klaster yang optimal kemudian diperiksa menggunakan *Davies-Bouldin Index* (DBI). Jumlah klaster yang dipilih berdasarkan jumlah klaster yang memiliki nilai DBI terkecil (Permatadevi, *et al.*, 2013).

2. TINJAUAN PUSTAKA

2.1 Pengertian Analisis Klaster

Menurut Hair, *et al.* (2006), analisis klaster merupakan suatu metode statistik yang digunakan untuk mengelompokkan sekumpulan objek ke dalam kelompok-kelompok berdasarkan

karakteristik yang dimilikinya. Objek diklasifikasikan ke dalam satu atau lebih kluster sehingga objek-objek yang berada di dalam kluster mempunyai kemiripan atau kesamaan karakter. Solusi kluster secara keseluruhan bergantung pada variabel-variabel yang digunakan sebagai dasar untuk menilai kesamaan. Kluster sendiri didefinisikan sebagai sejumlah objek yang mirip yang dikelompokkan secara bersama.

Apabila terdapat n objek dan p variabel, maka observasi x_{ij} dengan $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, p$, dapat digambarkan sebagai berikut:

Tabel 1 Susunan Observasi Analisis Kluster

	Variabel 1	Variabel 2	Variabel 3	Variabel p
Objek 1	X_{11}	X_{12}	X_{13}	X_{1p}
Objek 2	X_{21}	X_{22}	X_{23}	X_{2p}
Objek 3	X_{31}	X_{32}	X_{33}	X_{3p}
.
.
.
Objek n	X_{n1}	X_{n2}	X_{n3}	X_{np}

Perbedaan analisis kluster dengan analisis faktor terletak pada karakteristik objek yang diringkas. Analisis faktor bertujuan untuk meringkas variabel atau faktor sedangkan analisis kluster bertujuan untuk meringkas atau membentuk setiap kelompok berdasarkan kesamaan karakteristik observasi atau kasus (Supranto, 2004).

2.2 Metode K-Means Cluster

Menurut Jhonson dan Wichern (2005), metode *K-Means* digunakan sebagai alternatif metode kluster untuk data dengan ukuran yang lebih besar. Hal ini dikarenakan metode ini memiliki kecepatan yang lebih tinggi jika dibandingkan dengan metode hirarki. Metode *K-Means* dapat digunakan untuk menjelaskan algoritma dalam penentuan suatu objek kedalam kluster tertentu berdasarkan rataan terdekat.

Dalam prosedur pembentukan *K-Means Cluster* terdapat langkah-langkah yang dapat dilakukan, antara lain:

1. Tentukan banyaknya kluster (k) yang akan dibentuk.
2. Bangkitkan k *centroid* awal (rata-rata setiap kluster).
3. Hitung jarak antara setiap objek dengan setiap *centroid* dan masukan objek tersebut ke dalam kluster yang sesuai berdasarkan jarak terdekat.
4. Tentukan *centroid* dari kluster yang baru.
5. Ulangi langkah 3 dan 4 sampai tidak ada lagi pemindahan objek antarkluster.

2.3 Pendeteksian Multikolinearitas

Menurut Hair, *et al.* (2006), multikolinearitas merupakan adanya hubungan yang linier di antara variabel penelitian. Apabila terdapat variabel-variabel yang mengalami multikolinearitas secara eksplisit dapat dipertimbangkan dengan lebih seksama apakah sebaiknya dibuang atau diganti dengan variabel lain.

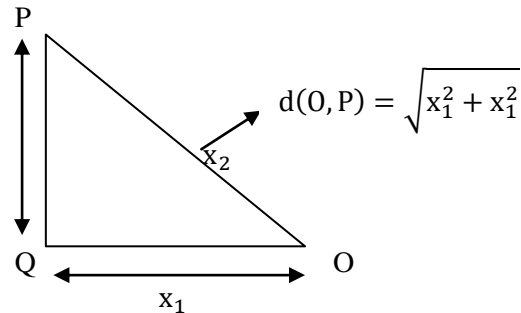
$$VIF_i = \frac{1}{1 - R_i^2} \quad (1)$$

Nilai VIF (*Variance Inflation Factor*) dapat digunakan untuk mengetahui ada atau tidaknya pengaruh multikolinearitas. Jika nilai VIF dari suatu variabel memiliki nilai lebih dari 10, maka variabel tersebut mengindikasikan terjadinya multikolinearitas. Nilai VIF dirumuskan pada persamaan (1) dengan R_i^2 adalah koefisien determinasi dari variabel bebas ke- i . Nilai VIF dari masing-masing variabel bebas dapat diperoleh dengan cara menjadikan satu variabel yang ingin

dicari nilai VIF-nya sebagai variabel tak bebas (*dependent*) dan menjadikan variabel sisanya sebagai variabel bebas (*independent*).

2.4 Memilih Ukuran Kemiripan

Untuk mengetahui seberapa mirip objek-objek yang diteliti dibutuhkan ukuran jarak (*distance-type measure*). Dengan memiliki sebuah ukuran kuantitatif, maka proses pengklasteran akan lebih mudah. Pada umumnya, jarak yang biasa digunakan adalah Jarak *Euclidean* (Supranto, 2004).



Gambar 1 Ilustrasi Jarak *Euclidean* dari Teorema Pitagoras

Jika pada suatu bidang dimisalkan titik $P = (x_1, x_2)$ dan $O = (0, 0)$, maka jarak adalah $d(O, P)$ berdasarkan teorema pythagoras maka,

$$d(O, P) = \sqrt{x_1^2 + x_2^2} \quad (2)$$

Akan tetapi, jika titik P memiliki koordinat sebanyak p sehingga $P = (x_1, x_2, \dots, x_p)$, maka jarak dari P ke titik asal O dengan $O = (0, 0, \dots, 0)$ adalah:

$$d(O, P) = \sqrt{x_1^2 + x_2^2 + \dots + x_p^2} \quad (3)$$

Jika jarak antara titik P dan Q dengan koordinat $P = (x_1, x_2, \dots, x_p)$ dan $Q = (y_1, y_2, \dots, y_p)$ maka,

$$\begin{aligned} d(P, Q) &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2} \\ d(x, y) &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2} \\ &= \sqrt{\sum_{i=1}^p (x_i - y_i)^2} \quad (4) \end{aligned}$$

Persamaan (4) jarak *Euclidean*. Meskipun pada awalnya tampak rumit, kebanyakan teknik multivariat didasarkan pada konsep jarak yang sederhana, salah satunya adalah jarak *Euclidean* (Jhonson dan Wichern, 2005).

2.5 Standardisasi Data

Jika jarak *Euclidean* semakin kecil, maka semakin mirip kasus atau objek tersebut. Akan tetapi, jarak *Euclidean* sangat sensitif terhadap ukuran sampel dan besarnya varian. Jika objek yang diteliti memiliki varian yang sangat berbeda, maka jarak *Euclidean* menjadi tidak akurat. Oleh sebab itu, perlu dilakukan standardisasi terhadap variabel penelitian sebelum dilakukan proses pengklasteran (Simamora, 2005).

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (5)$$

dengan:

- Z_{ij} = Data hasil standardisasi observasi ke variabel ke j
- x_{ij} = Observasi ke variabel ke j
- \bar{x}_j = Rata-rata variabel ke j
- s_j = Simpangan baku variabel ke j

2.6 Menentukan Jumlah Kluster

Menurut Permatadevi, *et al.* (2013), jika proses pengklasteran untuk masing-masing k selesai, maka untuk menentukan jumlah kluster yang paling optimal dapat dilakukan penilaian menggunakan *Davies-Bouldin Index* (DBI). Pengklasteran dengan jumlah kluster yang optimal adalah pengklasteran yang memiliki nilai DBI minimum. Nilai DBI dirumuskan pada persamaan (6).

$$DBI = \frac{1}{k} \sum_{i=1}^k \{R_i\} \quad (6)$$

dengan

$$R_i = \max_{j=1, \dots, k, i \neq j} R_{ij} \quad , \quad R_{ij} = \frac{s_i + s_j}{d_{ij}}$$

dan

$$s_i = \left[\frac{1}{n_i} \sum_{x \in n_i} d^2(x, v_i) \right]^{\frac{1}{2}}$$

dimana:

- k = Jumlah kluster
- R_{ij} = Ukuran kemiripan antara n_i dan n_j
- s_i = Ukuran dispersi kluster ke-i, $i = 1, 2, \dots, k$
- d_{ij} = Jarak antara *centroid* kluster ke-i dan *centroid* kluster ke-j ($d_{ij} = d_{ji}$)
- n_i = Banyaknya anggota kluster ke-i, $i = 1, 2, \dots, k$
- v_i = *Centroid* kluster dari n_i

3. METODOLOGI PENELITIAN

Data yang digunakan dalam penelitian ini adalah data sekunder pada tahun 2012 dan diperoleh dari survei yang dilakukan oleh BPS. Kantor pemerintah serta masyarakat setempat sebagai responden. Data diolah dengan menggunakan *Microsoft Excel*, *Minitab 14*, dan *SPSS 16*. Variabel yang digunakan merupakan enam bidang prioritas kesejahteraan rakyat yang dibuat oleh TNP2K. Adapun definisi dari variabel yang digunakan adalah sebagai berikut:

1. Angka Partisipasi Murni SD atau MI (APM SD atau MI)
Proporsi anak sekolah pada usia 7-12 tahun yang masih bersekolah pada jenjang SD (negeri dan swasta), MI, dan sederajat terhadap seluruh anak pada kelompok usia yang sama. Dinyatakan dalam satuan persen.
2. Angka Partisipasi Murni SMP atau MTs (APM SMP atau MTs)
Proporsi anak sekolah pada usia 13-15 tahun yang masih bersekolah pada jenjang SMP (negeri dan swasta), MTs, dan sederajat terhadap seluruh anak pada kelompok usia yang sama. Dinyatakan dalam satuan persen.
3. Angka Kematian Bayi (AKB)

Angka kematian bayi adalah angka yang menunjukkan banyaknya kematian bayi berusia nol tahun dari setiap 1.000 kelahiran hidup pada tahun tertentu. Dinyatakan dengan per seribu kelahiran hidup.

4. Akses Air Bersih

Air minum yang bersih adalah air minum yang terlindungi, meliputi air ledeng (keran), penampungan air hujan (PAH) atau mata air dan sumur terlindung, sumur bor atau sumur pompa, yang jaraknya minimal 10 m dari pembuangan kotoran, penampungan limbah dan sampah. Dinyatakan dalam satuan persen.

5. Akses Listrik

Listrik yang dipakai termasuk listrik yang bersumber dari PLN maupun non-PLN (dikelola oleh instansi atau pihak lain selain PLN). Dinyatakan dalam satuan persen.

6. Tingkat Pengangguran Terbuka (TPT)

TPT diperoleh persentase jumlah pengangguran terhadap jumlah angkatan kerja. Dinyatakan dalam satuan persen.

Tahapan-tahapan yang dilakukan dalam melakukan pengelompokan 33 provinsi berdasarkan enam bidang prioritas kesejahteraan rakyat adalah sebagai berikut:

1. Menentukan banyaknya klaster (k) yang akan dibuat, yaitu sebanyak dua, tiga, empat dan lima klaster.
2. Menentukan *centroid* awal.
3. Menghitung jarak antarobjek dengan setiap *centroid* dengan jarak *Euclidean*.
4. Menentukan klaster yang terbentuk berdasarkan nilai jarak terkecil.
5. Menentukan *centroid* baru untuk iterasi selanjutnya dengan cara menghitung rata-rata dari data yang ada pada setiap klaster.
6. Ulangi langkah 3, 4, dan 5 sampai tidak ada lagi perpindahan klaster di setiap objek. Jika sudah tidak terdapat perpindahan, maka *centroid* pada langkah e menjadi nilai rata-rata setiap klaster dan proses selesai.
7. Menginterpretasi dan membuat profil klaster, meliputi pengkajian mengenai *centroid* karena nilai *centroid* memungkinkan peneliti untuk menguraikan setiap klaster dengan cara memberi nama atau label.

4. HASIL DAN PEMBAHASAN

4.1 Pendeteksian Multikolinearitas

Untuk memeriksa data yang digunakan mengalami multikolinearitas atau tidak dapat diketahui dengan cara menghitung nilai VIF seperti pada persamaan (1). Berdasarkan *output* dari SPSS diperoleh R^2 untuk masing-masing variabel yang selanjutnya dapat digunakan untuk mencari nilai VIF. Tabel 2 menunjukkan nilai VIF dari setiap variabel.

Tabel 2 Nilai VIF dari Enam Variabel

Variabel	R^2	VIF
APM SD atau MI	0,609	2,558
APM SMP atau MTs	0,747	3,953
AKB	0,294	1,416
Akses Air	0,213	1,271
Akses Listrik	0,779	4,525
TPT	0,370	1,586

Pada Tabel 2 terlihat bahwa dari keenam variabel penelitian yang digunakan tidak memiliki nilai VIF lebih dari 10. Oleh karena itu, dapat disimpulkan tidak terdapat multikolinearitas antarvariabel bebas.

4.2 Proses Pengklasteran Metode *K-Means Cluster*

Seluruh data dibuat dalam bentuk persentase kecuali angka kematian bayi. Perbedaan satuan ini menyebabkan perlunya dilakukan standardisasi terhadap data. Proses standardisasi dapat

dilakukan dengan menggunakan persamaan (5). Jika data sudah diubah ke dalam bentuk standarnya, maka proses pengklasteran dapat dilakukan.

Tabel 3 Anggota Klaster untuk $k = 2, 3, 4, 5$

Provinsi	$k = 2$	$k = 3$	$k = 4$	$k = 5$
Aceh	1	1	1	1
Sumut	2	3	2	2
Sumbar	1	3	3	5
Riau	2	3	2	5
Kep. Riau	1	1	1	5
Jambi	2	3	2	2
Sumsel	2	3	2	2
Kep. Babel	2	3	2	5
Bengkulu	2	3	2	5
Lampung	2	3	2	5
DKI Jakarta	1	1	3	3
Jawa Barat	1	1	3	3
Banten	1	1	1	1
Jawa Tengah	2	3	2	2
DI Yogyakarta	2	3	2	2
Jawa Timur	2	3	2	2
Bali	2	3	2	2
NTB	2	3	2	2
NTT	2	2	4	4
Kalbar	2	3	2	2
Kalteng	2	2	2	2
Kalsel	2	3	2	2
Kaltim	1	1	3	3
Sulawesi Utara	2	3	2	3
Gorontalo	2	2	4	4
Sulawesi Tengah	2	2	4	2
Sulawesi Selatan	2	3	2	2
Sulawesi Barat	2	2	4	4
Sulawesi Tenggara	2	3	2	2
Maluku	2	3	2	2
Maluku Utara	2	2	4	2
Papua	2	2	4	4
Papua Barat	2	2	4	4

Banyaknya klaster yang akan dibuat adalah sebanyak dua, tiga, empat, dan lima klaster ($k = 2, 3, 4, 5$). Oleh karena itu, nilai *centroid* awal untuk masing-masing proses pengklasteran adalah k objek pertama. Hasil *output* dari Minitab 14 yang menunjukkan anggota klaster untuk setiap jumlah klaster dapat dilihat pada Tabel 3.

4.3 Penentuan jumlah Klaster

Jumlah klaster yang optimal dapat diketahui dengan menggunakan nilai DBI yang dirumuskan pada persamaan (8). Semakin kecil nilai DBI akan memberikan hasil yang baik. Berikut ini merupakan penghitungan nilai DBI untuk masing-masing jumlah klaster.

1. Untuk $k = 2$ diperoleh jarak antar *centroid* klaster (d_{ij}) sebagai berikut:

$$d_{12} = 2,8158$$

dan ukuran dispersi untuk klaster ke- i (s_i) adalah:

$$s_1 = 1,2379$$

$$s_2 = 2,30004$$

sehingga,

$$R_{12} = 1,2565$$

Jika dibuat dalam bentuk matriks, maka:

$$\mathbf{R} = \begin{bmatrix} - & 1,2565 \\ 1,2565 & - \end{bmatrix}$$

Sehingga, DBI untuk $k = 2$ adalah:

$$DBI = \frac{1}{2} \sum_{i=1}^2 \{R_i\} = \frac{1}{2} (1,2565 + 1,2565) = 1,2565$$

2. Untuk $k = 3$ diperoleh jarak antar *centroid* kluster (d_{ij}) sebagai berikut:

$$d_{12} = 4,4045$$

$$d_{13} = 2,5975$$

$$d_{23} = 2,9807$$

dan ukuran dispersi untuk kluster ke- i (s_i) adalah:

$$s_1 = 1,1821$$

$$s_2 = 2,5050$$

$$s_3 = 1,4683$$

sehingga,

$$R_{12} = 0,8371$$

$$R_{13} = 1,0204$$

$$R_{23} = 1,3330$$

Jika dibuat dalam bentuk matriks, maka:

$$\mathbf{R} = \begin{bmatrix} - & 0,8371 & 1,0204 \\ 0,8371 & - & 1,3330 \\ 1,0204 & 1,3330 & - \end{bmatrix}$$

Sehingga, DBI untuk $k = 3$ adalah:

$$DBI = \frac{1}{3} \sum_{i=1}^3 R_i = \frac{1}{3} (1,0204 + 1,3330 + 1,3330) = 1,2288$$

3. Untuk $k = 3$ diperoleh jarak antar *centroid* kluster (d_{ij}) sebagai berikut:

$$d_{12} = 2,8456$$

$$d_{13} = 1,3961$$

$$d_{14} = 4,5976$$

$$d_{23} = 2,3215$$

$$d_{24} = 3,0876$$

$$d_{13} = 4,4568$$

dan ukuran dispersi untuk kluster ke- i (s_i) adalah:

$$s_1 = 1,3034$$

$$s_2 = 2,2867$$

$$s_3 = 0,8688$$

$$s_4 = 6,4759$$

sehingga,

$$R_{12} = 0,9327$$

$$R_{13} = 1,4853$$

$$R_{14} = 0,8018$$

$$R_{23} = 1,0529$$

$$R_{24} = 1,3240$$

$$R_{34} = 0,7801$$

Jika dibuat dalam bentuk matriks, maka:

$$\mathbf{R} = \begin{bmatrix} - & 0,9327 & 1,4853 & 0,8018 \\ 0,9327 & - & 1,0529 & 1,3240 \\ 1,4853 & 1,0529 & - & 0,7801 \\ 0,8018 & 1,3240 & 0,7801 & - \end{bmatrix}$$

Sehingga, DBI untuk $k = 4$ adalah:

$$DBI = \frac{1}{4} \sum_{i=1}^4 R_i = \frac{1}{4} (1,4853 + 1,3240 + 1,4853 + 1,3240) = 1,3996$$

4. Untuk $k = 3$ diperoleh jarak antar *centroid* kluster (d_{ij}) sebagai berikut:

$$\begin{array}{ll} d_{12} = 3,3388 & d_{13} = 1,5785 \\ d_{14} = 5,1360 & d_{15} = 2,5058 \\ d_{23} = 2,6787 & d_{24} = 3,3816 \\ d_{25} = 1,8637 & d_{34} = 4,6323 \\ d_{35} = 2,0207 & d_{45} = 4,0728 \end{array}$$

dan ukuran dispersi untuk kluster ke- i (s_i) adalah:

$$\begin{array}{ll} s_1 = 0,7287 & s_2 = 1,4902 \\ s_3 = 1,1522 & s_4 = 2,7861 \\ s_5 = 4,2097 & \end{array}$$

sehingga,

$$\begin{array}{ll} R_{12} = 0,6646 & R_{13} = 1,1916 \\ R_{14} = 0,6843 & R_{15} = 1,9687 \\ R_{23} = 0,9865 & R_{24} = 1,2646 \\ R_{25} = 3,0586 & R_{34} = 0,8502 \\ R_{35} = 2,6533 & R_{45} = 1,7176 \end{array}$$

Jika dibuat dalam bentuk matriks, maka:

$$\mathbf{R} = \begin{bmatrix} - & 0,6646 & 1,1916 & 0,6843 & 1,9687 \\ 0,6646 & - & 0,9865 & 1,2646 & 3,0586 \\ 1,1916 & 0,9865 & - & 1,9687 & 2,6533 \\ 0,6843 & 1,2646 & 0,8502 & - & 1,7176 \\ 1,9687 & 3,0586 & 2,6533 & 1,7176 & - \end{bmatrix}$$

Sehingga, DBI untuk $k = 5$ adalah:

$$DBI = \frac{1}{5} \sum_{i=1}^5 R_i = \frac{1}{5} (1,9687 + 3,0586 + 2,6533 + 1,7176 + 3,0586) = 2,4914$$

Berdasarkan uraian perhitungan di atas, maka jika dibuat ke dalam tabel antara jumlah kluster dan nilai DBI-nya akan terlihat seperti pada Tabel 4.

Tabel 4 Nilai DBI untuk $k = 2, 3, 4,$ dan 5

k	DBI
2	1,2565
3	1,2288
4	1,3996
5	2,4914

Pada Tabel 4 terlihat bahwa nilai DBI terkecil adalah saat $k = 3$. Oleh karena itu, dapat dipastikan bahwa jumlah kluster yang optimal untuk pengklasteran ini adalah tiga kluster.

4.4 Interpretasi dan Profiling Hasil Kluster untuk $k = 3$

Nilai *cluster centroids* yang terdapat Tabel 5 dapat digunakan untuk mengetahui kondisi kesejahteraan rakyat dari 33 provinsi di Indonesia berdasarkan enam bidang prioritas TNP2K. Tanda positif ataupun negatif dikarenakan adanya pengaruh standardisasi data sebelum proses pengklasteran dilakukan.

Tabel 5 Rata-rata Setiap Kluster Setelah Proses Standardisasi

Variabel	Kluster 1	Kluster 2	Kluster 3
APM SD atau MI	0,3258	-0,6470	0,1695
APM SMP atau MTs	0,9451	-1,2243	0,2171
AKB	-0,5428	1,3898	-0,4138
Akses Air Bersih	-1,3070	-0,0676	0,4412
Akses Listrik	0,6360	-1,2263	0,3155
TPT	1,5426	-0,6957	-0,1942

Menurut Simamora (2005), penginterpretasian dari *Z-Score* atau nilai standardisasi yang bernilai positif menunjukkan ke arah yang kuat. Akan tetapi, cenderung ke arah yang lemah jika bernilai negatif. Dari Tabel 5 maka dapat diketahui bahwa kluster satu merupakan kluster dengan provinsi-provinsi yang memiliki keunggulan pada bidang APM SD atau MI, APM SMP atau MTs, AKB, dan akses listrik namun juga memiliki kelemahan pada akses air dan TPT. Kluster dua merupakan kluster dengan provinsi-provinsi yang memiliki keunggulan pada satu bidang saja, yaitu TPT, dan memiliki kelemahan pada ke lima bidang lainnya. Kluster tiga merupakan kluster dengan provinsi-provinsi yang memiliki keunggulan di segala bidang.

5. KESIMPULAN

Dari hasil analisis dan pembahasan yang telah diuraikan pada bab sebelumnya, maka diperoleh kesimpulan sebagai berikut:

1. Berdasarkan nilai yang diperoleh dari *Davies-Bouldin Index* untuk jumlah kluster dua, tiga, empat, dan lima diperoleh hasil bahwa jumlah kluster tiga adalah yang terbaik. Hal ini diketahui karena jumlah kluster tiga memiliki nilai DBI terkecil.
2. Berdasarkan nilai DBI, maka pengklasteran dibuat ke dalam tiga kluster. Hasil pengklasteran menunjukkan bahwa terdapat enam provinsi pada kluster satu, delapan provinsi pada kluster dua, dan 19 provinsi pada kluster tiga. Rincian dari ketiga kluster adalah sebagai berikut:
 - a. Kluster satu meliputi Provinsi Aceh, Provinsi Kepulauan Riau, Provinsi DKI Jakarta, Provinsi Jawa Barat, Provinsi Banten, dan Provinsi Kalimantan Timur. Kluster satu memiliki keunggulan pada empat bidang, yaitu APM SD atau MI, APM SMP atau MTs, AKB, dan akses listrik. Rendahnya jumlah rumah tangga yang mendapatkan akses air bersih dan tingginya jumlah pengangguran diharapkan dapat diatasi dengan program Penyediaan Air Minum Berbasis Masyarakat (Pamsimas) dan Pengembangan Infrastruktur Sosial Ekonomi Wilayah (PISEW).
 - b. Kluster dua meliputi Provinsi Nusa Tenggara Timur, Provinsi Kalimantan Tengah, Provinsi Gorontalo, Provinsi Sulawesi Tengah, Provinsi Sulawesi Barat, Provinsi Maluku Utara, Provinsi Papua, dan Provinsi Papua Barat. Kluster dua memiliki keunggulan hanya pada satu bidang saja, yaitu TPT. Rendahnya partisipasi siswa dalam wajib belajar 9 tahun diharapkan dapat diatasi dengan pengadaan program BOS dan BSM yang lebih intensif. Pemaksimalan peran Posyandu dan Puskesmas serta pemerataan sebaran dokter anak dan standardisasi terhadap rumah sakit juga harus ditingkatkan untuk mengatasi tingginya jumlah kematian bayi. Selain itu juga perlu dilakukan. Untuk mengatasi rendahnya rumah tangga yang memiliki akses air bersih maka program Pamsimas dapat dilakukan. Dalam mengatasi rendahnya akses listrik di beberapa provinsi di Indonesia pemerintah dapat mengkaji kembali terhambatnya perizinan pembangunan sejumlah pembangkit listrik. Selain itu, masalah pasokan batu bara sebagai salah satu bahan pendukung pemenuhan kebutuhan listrik nasional juga perlu segera diselesaikan
 - c. Kluster tiga meliputi Provinsi Sumatera Utara, Provinsi Sumatera Barat, Provinsi Riau, Provinsi Jambi, Provinsi Sumatera Selatan, Provinsi Kepulauan Bangka Belitung, Provinsi Bengkulu, Provinsi Lampung, Provinsi Jawa Tengah, Provinsi DI Yogyakarta, Provinsi Jawa Timur, Provinsi Bali, Provinsi Nusa Tenggara Barat, Provinsi Kalimantan Barat, Provinsi Kalimantan Selatan, Provinsi Sulawesi Utara, Provinsi

Sulawesi Selatan, Provinsi Sulawesi Tenggara, dan Provinsi Maluku. Klaster tiga merupakan kelompok provinsi yang memiliki keunggulan di segala bidang. Jumlah para siswa yang berpartisipasi dalam wajib belajar 9 tahun serta rumah tangga yang memiliki akses terhadap air bersih dan akses terhadap listrik menunjukkan angka di atas rata-rata. Selain itu, untuk tingkat kematian bayi dan persentase pengangguran di klaster tiga juga menunjukkan angka yang rendah atau di bawah rata-rata.

6. DAFTAR PUSTAKA

- Hair, J.F. Jr. *et al.* 2006. *Multivariate Data Analysis 6th Edition*. New Jersey: Pearson Prentice Hill, Inc.
- Haughton, J. dan Khandker, S.H. 2012. *Pedoman tentang Kemiskinan dan Ketimpangan*. Tanujaya, E, penerjemah. Jakarta: Salemba Empat. Terjemahan dari: *Handbook on Poverty & Inequality*.
- Jhonson, R.A. and Wichern, D.W. 2005. *Applied Multivariate Statistical Analysis 6th Edition*. New Jersey: Prentice Hill, Inc.
- Perrmatadevi, M. A., Hendrawan, R. A., dan Hafidz, I. 2013. *Karakteristik Pelanggan Telepon Kabel Menggunakan Clusteing SOM dan K-Means untuk Mengurangi Kesalahan Klasifikasi Pelanggan Perusahaan Telekomunikasi*. Jurnal Teknik Pomits Vol. 1, No. 1: Hal. 1-6. <http://digilib.its.ac.id/public/ITS-paper-34583-5209100025-Paper.pdf>. (28 Agustus 2015)
- Roestam, S. 1993. *Pembangunan Nasional untuk Kesejahteraan Rakyat*. Jakarta: Kantor Menteri Koordinator Bidang Kesejahteraan Rakyat Republik Indonesia.
- Simamora, B. 2005. *Analisis Multivariat Pemasaran*. Jakarta: PT. Gramedia Pustaka Utama.
- Supranto, J. 2004. *Analisis Multivariat: Arti dan Interpretasi*. Jakarta: PT. Rineka Cipta.
- [TNP2K] Tim Nasional Percepatan Penanggulangan Kemiskinan. 2011. *Indikator Kesejahteraan Daerah Provinsi DKI Jakarta*. http://data.tnp2k.go.id/file_data/Data/IKD/31_DKI.pdf. (1 Desember 2014)