

METODE K-HARMONIC MEANS CLUSTERING DENGAN VALIDASI SILHOUETTE COEFFICIENT

(Studi Kasus : Empat Faktor Utama Penyebab *Stunting* 34 Provinsi di Indonesia Tahun 2018)

Silvy 'Aina Salsabila¹, Tatik Widiharih^{2*}, Sudarno³

^{1,2,3} Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

*e-mail: widiharih@gmail.com

ABSTRACT

The k-harmonic means method is a method of using the cluster center point value, which is to determine each cluster from its center point based on the calculation of the harmonic average. The k-harmonic means determines the existence of each data point based on the membership function and weighting function by using a distance measure. In the clustering, which aims to increase the importance of data that is far from each central point. This causes the k-harmonic means to be insensitive in initialization in determining the cluster center point and significantly improves the quality of clustering compared to k-means. In determining the level of similarity, the determination of the level of similarity uses the distance measure and the distance measure used is the Euclidean distance measure. The distance measure used in cluster analysis can affect the cluster results obtained. Thus, to determine the quality of the results of the cluster analysis, validation tests were carried out using an internal criteria approach, namely silhouette coefficient. In this study, the k-harmonic means used to classify provinces in Indonesia based on the causes of stunting in 2018. The stunting in children under five in Indonesia has exceeded the limit set by WHO. In 2016-2017 there was an increase in the prevalence of stunting by 27.5% to 29.6%. The k-harmonic means method is used so that the four main factors causing stunting in every province in Indonesia can be seen and the prevention and cure of stunting can run optimally. This method is also used because the data on the four factors that cause stunting show a significant rate of change and as a measure of central tendency in 34 provincial objects in Indonesia. Four factors that cause stunting are used, namely the percentage of households that do not have access to clean drinking water, the percentage of exclusive breastfeeding, the percentage of Low Birth Weight Babies (LBW) 2,500-grams born safely and the percentage of households that do not have proper sanitation facilities. The results obtained by the cluster which is optimal at $k=3$ using the Euclidean, where the silhouette coefficient = $0,3040722675 \approx 0,3$. Based on the results of the cluster analysis, it is known that in cluster one, the main factor that stands out the most is the percentage of exclusive breastfeeding. In cluster two, the main factor that stands out the most is the percentage of Low Birth Weight Babies (LBW) 2,500-grams born safely. In cluster three, the most prominent main factors are the percentage of Low Birth Weight Babies (LBW) 2,500-grams born safely and the percentage of households that do not have proper sanitation facilities with the highest average centroid among other clusters.

Keywords: Clustering, K-Harmonic Means, Euclidean distance, Silhouette Coefficient, Stunting

1. Pendahuluan

Upaya penurunan prevalensi balita pendek (*stunting*) menjadi salah satu prioritas pembangunan nasional yang tercantum di dalam sasaran pokok Rencana Pembangunan Jangka Menengah (RPJM) tahun 2015-2019 dengan target penurunan prevalensi *stunting* (pendek dan sangat pendek) pada balita adalah sebesar 28%. Menurut Pusat Data dan Informasi Kementerian Kesehatan RI (2017), masalah balita pendek dipengaruhi oleh empat faktor utama yaitu rumah tangga yang tidak memiliki akses air minum bersih, pemberian Air Susu Ibu (ASI) eksklusif, bayi Berat Badan Lahir Rendah (BBLR) ≤ 2.500 -gram lahir dengan selamat dan rumah tangga yang tidak memiliki fasilitas sanitasi yang layak. Anak yang menderita *stunting* akan lebih rentan terhadap penyakit dan ketika dewasa berisiko untuk mengidap penyakit degeneratif. Dampak *stunting* tidak hanya ada pada segi kesehatan tetapi juga mempengaruhi tingkat kecerdasan anak. Berdasarkan data Pemantauan Status Gizi (PSG) selama tiga tahun terakhir (2015-2017), *stunting* (pendek) di Indonesia memiliki

prevalensi tertinggi dibandingkan dengan masalah status gizi lainnya seperti gizi kurang, gizi kronis, obesitas dan lainnya. Prevalensi balita pendek tersebut mengalami peningkatan dari tahun 2016 sebesar 27.5% menjadi 29.6%. Terkait kasus tersebut menentukan lokasi prioritas untuk mengetahui faktor utama kasus *stunting* pada setiap Provinsi di Indonesia dengan mengelompokkan provinsi-provinsi mana saja yang memiliki kesamaan faktor utama penyebab *stunting* akan sangat mempermudah tugas pemerintah dalam memberikan pertolongan, bantuan serta kegiatan pencegahan terhadap kasus *stunting* di setiap Provinsi di Indonesia. Pengelompokan daerah ini dapat dilakukan dengan menggunakan analisis kelompok atau klaster.

Dalam mengatasi masalah yang terjadi pada inialisasi *centroid* atau nilai titik pusat klaster, Zhang, Hsu dan Dayal mengusulkan sebuah metode baru yang diberi nama *k-harmonic means* yang kemudian dimodifikasi oleh Hammerly dan Elkan (Hammerly, 2002). Pengembangan yang diterapkan di dalam *k-harmonic means* adalah penggunaan nilai titik pusat klaster yaitu, menentukan setiap klaster dari titik pusatnya berdasarkan perhitungan rata-rata harmonik. Metode *k-harmonic means* menentukan keberadaan tiap titik datanya berdasarkan dari fungsi keanggotaan, kemudian menggunakan fungsi pembobotan dengan menggunakan ukuran jarak *euclidean* pada proses *clustering* yang bertujuan untuk meningkatkan kepentingan data-data yang jauh dari tiap-tiap titik pusat. Hal ini menyebabkan *k-harmonic means* bersifat tidak sensitif dalam inialisasi awal titik pusat klaster dan secara signifikan meningkatkan kualitas klasterisasi dibandingkan dengan *k-means* (Hamerly & Elkan, 2002).

Analisis klaster adalah teknik pengklasteran untuk mengelompokkan objek berdasarkan karakteristik yang dimiliki dari objek tersebut. Salah satu metode analisis klaster yang sering digunakan adalah metode *non hierarki partitioning*. Metode tersebut objek-objek pada data dikelompokkan ke dalam *k* klaster, dengan banyak klaster ditentukan oleh peneliti. Metode *partitioning* yang sering digunakan yaitu metode *k-means*. Menurut Han dan Kamber (2006), algoritma *k-means* sensitif terhadap pencilan, karena menggunakan nilai rata-rata (*mean*) sebagai pusat kelompoknya.

Pada analisis klaster, objek-objek dikelompokkan berdasarkan kemiripannya, untuk mengukur tingkat kemiripan dapat digunakan ukuran jarak. Semakin besar nilai jarak yang diperoleh, maka semakin jauh letak objek dengan pusat klaster yang terbentuk. Pada penelitian ini akan digunakan ukuran jarak *euclidean*. Hal yang perlu juga diperhatikan dalam analisis klaster yaitu validasi hasil pengklasteran. Validasi hasil pengklasteran dilakukan untuk memperoleh partisi yang paling sesuai dengan data. Jika tidak divalidasi, maka akan berpengaruh pada hasil analisis. Pada penelitian ini digunakan validasi dengan pendekatan kriteria internal yaitu validasi *silhouette coefficient*.

Provinsi-provinsi di Indonesia akan dikelompokkan menjadi 3 klaster berdasarkan empat faktor utama penyebab *stunting* untuk tahun 2018, sehingga dapat diketahui faktor utama apa yang paling menonjol pada masing-masing Provinsi di Indonesia. Pengelompokan tersebut dapat dilakukan dengan menggunakan analisis klaster *K-Harmonic Means*.

2. Tinjauan Pustaka

2.1. Analisis Klaster

Analisis klaster adalah teknik pengklasteran untuk mengelompokkan objek berdasarkan karakteristik yang dimiliki oleh objek tersebut. Objek diklasifikasikan ke dalam satu atau lebih klaster sehingga objek-objek yang berada di dalam klaster akan mempunyai kemiripan atau kesamaan karakter (Hair *et al.*, 2010).

Menurut Hair *et al.*, (2010) terdapat dua asumsi dalam analisis kluster yaitu sampel yang representatif dan tidak terjadi multikolinieritas.

a. Sampel Representatif

Sampel representatif adalah sampel yang diambil dapat mewakili populasi yang ada. Pengujian sampel yang mewakili populasi dapat dilakukan dengan melihat syarat kecukupan suatu sampel menggunakan uji KMO (*Kaiser Mayer Olkin*).

Berikut uji hipotesis untuk melihat apakah sampel dapat mewakili populasi:

Menurut Widarjono (2010), rumus KMO adalah sebagai berikut :

$$KMO = \frac{\sum_{j=1}^p \sum_{l=1, l \neq j}^p r_{x_j x_l}^2}{\sum_{j=1}^p \sum_{l=1, l \neq j}^p r_{x_j x_l}^2 + \sum_{j=1}^p \sum_{l=1, l \neq j}^p \rho_{x_j x_l x_m}^2}$$

dengan, p : banyaknya variabel, n : banyaknya objek, x_i : objek pada pengamatan ke- i , $r_{x_j x_l}$: korelasi antara variabel x_j dan x_l , \bar{x}_j : rata-rata variabel x_j , \bar{x}_l : rata-rata variabel x_l , dan $\rho_{x_j x_l x_m}$: korelasi parsial antara variabel x_j dan x_l dengan menjaga agar x_m konstan. Sampel dikatakan dapat mewakili populasi atau sampel representatif apabila diperoleh nilai KMO berkisar antara 0,5 sampai dengan 1 ($0,5 \leq KMO \leq 1$).

b. Tidak Terjadi Multikolinieritas

Multikolinieritas adalah suatu peristiwa dimana terjadi korelasi yang kuat antara dua atau lebih variabel. Multikolinieritas merupakan masalah yang perlu diperhatikan dalam analisis kluster, karena dapat mempersulit dalam menentukan pengaruh/efek dari masing-masing variabel dan mempengaruhi hasil pengklasteran akhir. Pada analisis kluster sebaiknya variabel-variabel tidak terindikasi multikolinieritas (Hair *et al.*, 2010). Berikut uji hipotesis untuk melihat multikolinieritas data:

Menurut Gujarati (2009), salah satu cara indentifikasi adanya multikolinieritas adalah menghitung nilai *Variance Inflation Factor* (VIF) yang dirumuskan sebagai berikut:

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

dengan, R_i^2 : koefisien determinasi yang diperoleh bila nilai variabel ke- i diregresikan dengan variabel lainnya, i : variabel, dengan $i = 1, 2, \dots, p$ dan VIF_i : nilai *Variance Inflation Factor* variabel ke- i

Jika nilai VIF lebih besar dari 10, maka menunjukkan adanya multikolinieritas (hubungan linier antar variabel)

Pada penelitian ini digunakan ukuran jarak yaitu, jarak *euclidean* (*euclidean distance*)

Jarak *euclidean* merupakan akar dari jumlah kuadrat selisih antar objek yang dikuadratkan (Supranto, 2004). Adapun persamaan untuk menghitung jarak *euclidean* adalah sebagai berikut:

$$d_{euc}(x_i, c_k) = \sqrt{\sum_{j=1}^p (x_{i,j} - c_{k,j})^2}$$

$i = 1, 2, 3, \dots, n$ dan $k = 1, 2, 3, \dots, l$ serta $j = 1, 2, 3, \dots, p$

dengan, $d_{euc}(x_i, c_k)$: jarak *Euclidean* antara pengamatan ke- i , variabel ke- j ke pusat kluster ke- k pada variabel ke- j

x_{ij} : Objek pada pengamatan ke- i pada variabel ke- j

c_{kj} : Pusat kelompok ke- k pada variabel ke- j

j : Banyaknya variabel yang diamati

k : Banyaknya kluster

Berikut tahapan-tahapan algoritma *k-medoids*, yaitu:

1. Menentukan k sebagai banyaknya kluster yang ingin dibentuk
2. Membangkitkan inisialisasi posisi titik pusat kluster awal dimana $C_{(k)}$ *centroid* k , sebanyak k *centroid* secara acak dari data yang ada.
3. Menghitung jarak setiap *data point* x_i menggunakan rumus perhitungan jarak terhadap masing-masing $C_{(k)}$ *centroid*
4. Mencari jarak terdekat $d_{i,min}$ setiap data point x_i dengan setiap titik pusat kluster $C_{(k)}$ *centroid* berdasarkan hasil perhitungan jarak
5. Memasukan setiap *data point* x_i kedalam *cluster* sesuai dengan $C_{(k)}$ *centroid* terdekat
6. Menghitung kembali titik pusat $C_{(k)}$ *centroid*, sebanyak k untuk posisi titik pusat kluster dari semua data berdasarkan persamaan *K-Harmonic Means* berikut:

$$m_{k,j} = \frac{\sum_{i=1}^N \frac{1}{d_{i,k}^3 \left(\sum_{k=1}^p \frac{1}{d_{euc}(x_{ij}, c_{kj})^2} \right)^2} \cdot x_i}{\sum_{i=1}^N \frac{1}{d_{i,k}^3 \left(\sum_{k=1}^p \frac{1}{d_{euc}(x_{ij}, c_{kj})^2} \right)^2}}$$

dengan, N : Jumlah data yang akan dikluster
 $d_{i,k} = d_{i,min}$: Jarak terkecil setiap data dalam x_i k kluster
 $d_{euc}(x_{ij}, c_{kj})$: Jarak Euclidean antara pengamatan ke- i , variabel ke- j ke pusat kluster ke- k pada variabel ke- j
 x_i : Objek pada pengamatan ke- i
 k : Banyaknya kluster yang diamati
 $m_{k,j}$: Titik pusat terbaru pada kluster ke- k , variabel ke- j
 c_{kj} : Titik pusat sebelumnya, pada kluster ke- k , variabel ke-

7. Ulangi Langkah (3) sampai (6) hingga mendapatkan posisi anggota *cluster* yang tidak terdapat perubahan atau tidak terdapat perubahan yang signifikan
8. Tetapkan keanggotaan data x_i pada suatu kluster dengan titik pusat kluster $C_{(k)}$ *centroid* dengan nilai keanggotaan x_i terhadap $C_{(k)}$ *centroid*

x_i merupakan anggota kluster dengan titik pusat kluster $C_{(k)}$ *centroid* apabila nilai keanggotaan $m_k(C_{(k)}/x_i)$ adalah yang terbesar dibandingkan dengan nilai keanggotaannya ke titik pusat kluster lain.

Validasi hasil analisis kluster dilakukan untuk memperoleh partisi yang paling sesuai dengan data. Jika kluster tidak divalidasi, maka akan berpengaruh pada hasil analisis. Pada penelitian ini digunakan validasi untuk memilih jarak dan validasi terbaik dalam pengklasteran *K-Harmonic Means* yaitu kriteria internal *Silhouette Coefficient*.

Metode validasi *Silhouette Coefficient* merupakan salah satu ukuran validasi yang berbasis kriteria internal. *Silhouette Coefficient* akan mengevaluasi penempatan setiap objek dalam setiap kluster dengan membandingkan jarak rata-rata objek dalam satu kluster dan

jarak antara objek dengan kluster yang berbeda (Aini *et al.*, 2014). Menghitung koefisien *silhouette* yang didefinisikan sebagai rata-rata $s(i)$ yaitu,

$$SC = \frac{1}{n} \sum_{i=1}^n s(i)$$

dengan, $s(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))}$, $b(i) = \min d(x_i, E)$, dan $a(i) = \frac{1}{n(A)-1} \sum_{i \in A, i \neq h} d(x_i, x_h)$

$b(i)$: nilai minimum dari jarak rata-rata objek i dengan semua objek pada kluster lain E lainnya

$a(i)$: rata-rata jarak objek ke- i dengan semua objek yang berada di dalam satu kluster

Hasil perhitungan nilai koefisien *silhouette* berada pada *range* -1 sampai 1.

Semakin besar nilai koefisien *silhouette* akan semakin baik kualitas suatu kelompok

3. Metodologi Penelitian

Data yang digunakan merupakan data sekunder yang diperoleh dari Badan Pusat Statistik dan Kementerian Kesehatan Tahun 2018. Data tersebut merupakan data persentase 4 faktor utama *stunting* dengan objek 36 Provinsi di Indonesia pada tahun 2018.

Variabel yang digunakan dalam penelitian ini adalah persentase rumah tangga yang tidak memiliki akses air minum bersih (X_1), persentase pemberian Air Susu Ibu (ASI) eksklusif (X_2), persentase Bayi Berat Lahir Rendah (BBLR) ≤ 2.500 -gram lahir dengan selamat (X_3), persentase rumah tangga yang tidak memiliki fasilitas sanitasi layak (X_4).

Tahapan analisis data dengan *k-harmonic means* adalah sebagai berikut:

1. Uji asumsi dalam analisis kluster yaitu
 - a. Melakukan uji asumsi sampel mewakili populasi (representatif), dengan menggunakan uji *Kaiser Mayer Olkin* (KMO)
 - b. Melakukan uji asumsi non-multikolinieritas, dengan menggunakan nilai *Variance Inflation Factor* (VIF). Apabila terjadi multikolinieritas pada salah satu variabel maka dilakukan analisis komponen utama, jika tidak terjadi multikolinieritas dapat dilanjutkan ke tahap berikutnya
2. Penentuan banyaknya kluster yang akan dibuat (k), nilai k yang digunakan dalam penelitian ini yaitu $k = 3, 4$ dan 5
3. Melakukan analisis kluster dengan algoritma *k-harmonic means*:
 - a. Membangkitkan inisialisasi posisi titik pusat kluster $C_{(k)}$ *centroid* awal sebanyak k *centroid* secara acak dari data yang ada
 - b. Menghitung jarak setiap *data point* x_i dengan $C_{(k)}$ *centroid* iterasi ke- t pada tiap kluster dengan perhitungan jarak menggunakan jarak *euclidean*.

Rumus ukuran jarak yang digunakan yaitu :

- Jarak *euclidean* : $d_{euc}(x_i, c_k) = \sqrt{\sum_{j=1}^p (x_{i,j} - c_{k,j})^2}$

- c. Mencari jarak terdekat $d_{i,min}$ setiap *data point* x_i dengan setiap titik pusat kluster $C_{(k)}$ *centroid* berdasarkan hasil perhitungan jarak
- d. Memasukan setiap *data point* x_i kedalam kluster dengan kluster sesuai dengan $C_{(k)}$ *centroid* terdekat
- e. Menghitung kembali titik pusat $C_{(k)}$ *centroid*, sebanyak k untuk posisi titik pusat kluster untuk posisi titik pusat kluster dari semua data berdasarkan persamaan *k-harmonic means* berikut:

$$m_{k,j} = \frac{\sum_{i=1}^N \frac{1}{d_{i,k}^3 \left(\sum_{k=1}^p \frac{1}{d_{euc}(x_{ij}, c_{kj})^2} \right)^2} \cdot x_i}{\sum_{i=1}^N \frac{1}{d_{i,k}^3 \left(\sum_{k=1}^p \frac{1}{d_{euc}(x_{ij}, c_{kj})^2} \right)^2}}$$

- f. Ulangi Langkah (3) sampai (4) hingga mendapatkan posisi anggota kluster yang tidak terdapat perubahan atau tidak terdapat perubahan yang signifikan
 - g. Tetapkan keanggotaan data x_i pada suatu kluster dengan titik pusat kluster $C_{(k)}$ centroid sesuai dengan nilai keanggotaan x_i terhadap $C_{(k)}$ centroid
4. Validasi hasil pengklasteran
 - a. Hitung nilai *silhouette coefficient* pada masing-masing kluster yang terbentuk
 - b. Bandingkan nilai *silhouette coefficient* dari $k = 3, 4$ dan 5 , pengklasteran dikatakan terbaik jika nilai validasi untuk *silhouette coefficient* mendekati angka 1
 5. Interpretasi dan profilisasi karakteristik wilayah dari hasil pengklasteran terbaik.

4. Hasil dan Pembahasan

4.1. Analisis Kluster

a. Representatif

Berdasarkan output pengujian KMO, diperoleh nilai KMO sebesar 0,557098 dimana nilai tersebut berkisar diantara 0,5 sampai 1. Jadi, dapat disimpulkan bahwa sampel mewakili populasi atau sampel representatif terpenuhi.

b. Tidak Terjadi Multikolinieritas

Berdasarkan hasil pengujian VIF dalam studi kasus ini, diperoleh hasil *output* pada Tabel 1.

Tabel 1. Nilai VIF dari 4 Variabel

Variabel	Nilai VIF
X ₁	1,849258
X ₂	1,026993
X ₃	1,624741
X ₄	2,589005

Menurut **Tabel 1** di atas, diketahui 4 variabel dengan nilai $VIF \leq 10$, maka variabel tersebut tidak terjadi multikolinieritas.

Berdasarkan hasil analisis pengklasteran dengan metode *k-harmonic means* dan uji validasi hasil kluster dengan *silhouette coefficient* untuk $k = 3, 4$, dan 5 dengan jarak *euclidean*. Maka dilakukan perbandingan nilai validasi untuk mendapatkan pengklasteran terbaik, dapat dilihat pada tabel 3.

Tabel 3. Perbandingan Pengklasteran *K-Harmonic Means*

Banyak Kluster	Jarak	<i>Silhouette Coefficient</i>
3	<i>Euclidean</i>	0,3040722675 \approx 0,3
4	<i>Euclidean</i>	0,258290813 \approx 0,26
5	<i>Euclidean</i>	0,255938758 \approx 0,26
<i>Silhouette Coefficient</i> (Maksimum)		0,3040722675 \approx 0,3

Berdasarkan tabel di atas, dapat diketahui bahwa pada pengelompokan 34 Provinsi di Indonesia dengan metode *k-harmonic means* dengan validasi *silhouette coefficient*, diperoleh pengklasteran terbaik yaitu, untuk validasi *silhouette coefficient*, diperoleh nilai validasi maksimum pada $k = 3$ menggunakan jarak *euclidean* yaitu sebesar $0,3040722675 \approx 0,3$

Oleh karena itu, berdasarkan pengklasteran dengan validasi *silhouette coefficient* dapat disimpulkan bahwa hasil pengklasteran terbaik dengan validasi tersebut berdasarkan data persentase 4 faktor utama penyebab *stunting* 34 Provinsi di Indonesia tahun 2018 dengan metode *k-harmonic means* yaitu pengklasteran pada $k= 3$ dengan pengukuran jarak *euclidean*, karena memiliki nilai indeks validasi yang paling optimum.

Profilisasi hasil analisis kluster dilakukan pada hasil pengklasteran terbaik, yaitu pengklasteran data 4 faktor utama penyebab *stunting* 34 Provinsi di Indonesia tahun 2018 ke dalam 3 kluster dengan metode *k-harmonic means* dengan pengukuran jarak *euclidean*. Pada tahap profilisasi akan dilihat karakteristik dari tiap kluster yang terbentuk, sehingga dapat dilihat kecenderungan dari tiap kluster. Pada metode *k-harmonic means*, karakteristik dari kluster-kluster yang terbentuk dapat direpresentasikan dengan nilai rata-rata dari setiap variabel. Variabel yang digunakan adalah variabel awal. Hasil perhitungan nilai *centroids*/rata-rata tiap variabel pada setiap kluster ditunjukkan pada **Tabel 4**.

Tabel 4. Nilai *Centroid*/Rata-rata Setiap Variabel pada Setiap Kluster

Variabel	Kluster 1	Kluster 2	Kluster 3
X1	36,976	35,36286	29,57583
X2	63,46067	33,74571	64,65417
X3	62,64733	69,28857	78,91167
X4	57,46933	62,43	78,04333

Berdasarkan **Tabel 4** di atas diperoleh informasi sebagai berikut:

a. Kluster 1

Anggota kluster satu terdiri dari Prov. Kepulauan Bangka Belitung, Prov. Kepulauan Riau, Prov. DKI Jakarta, Prov. Jawa Tengah, Prov. DI Yogyakarta, Prov. Jawa Timur, Prov. Bali, Prov. Nusa Tenggara Barat, Prov. Kalimantan Utara, Prov. Kalimantan Timur, Prov. Sulawesi Selatan dan Prov. Sulawesi Tenggara. Kluster ini memiliki variabel atau faktor penyebab *stunting* dengan *centroid* atau rata-rata paling tinggi adalah persentase Pemberian Air Susu Ibu (ASI) eksklusif dan persentase Bayi Berat Lahir Rendah (BBLR) ≤ 2.500 -gram lahir dengan selamat. Akan tetapi, dalam kluster ini faktor persentase rumah tangga yang tidak memiliki akses air minum bersih memiliki *centroid* atau rata-rata paling rendah di dalam kluster 1.

b. Kluster 2

Anggota kluster satu terdiri dari Prov. Aceh, Prov. Sumatera Barat, Prov. Riau, Prov. Jambi, Prov. Sumatera Selatan, Prov. Bengkulu, Prov. Lampung, Prov. Jawa Barat, Prov. Nusa Tenggara Timur, Prov. Kalimantan Barat, Prov. Kalimantan Tengah, Prov. Kalimantan Selatan, Prov. Sulawesi Tengah, Prov. Sulawesi Barat dan Prov. Maluku Utara. Kluster ini memiliki variabel atau faktor penyebab *stunting* dengan *centroid* atau rata-rata paling tinggi persentase Bayi Berat Lahir Rendah (BBLR) ≤ 2.500 -gram lahir dengan selamat dan persentase rumah tangga yang tidak memiliki fasilitas sanitasi layak. Akan tetapi, dalam kluster ini faktor persentase rumah tangga yang tidak memiliki akses air minum bersih dan persentase pemberian Air Susu Ibu (ASI) eksklusif memiliki *centroid* atau rata-rata paling rendah di dalam kluster 2.

c. Kluster 3

Anggota kluster satu terdiri Prov. Sumatera Utara, Prov. Banten, Prov. Sulawesi Utara, Prov. Gorontalo, Prov. Maluku, Prov. Papua Barat dan Prov. Papua. Kluster ini faktor penyebab *stunting* dengan *centroid* atau rata-rata paling tinggi di antara kluster lain adalah faktor persentase Bayi Berat Lahir Rendah (BBLR) ≤ 2.500 -gram lahir dengan selamat dan persentase rumah tangga yang tidak memiliki fasilitas sanitasi layak.

5. Penutup

5.1. Kesimpulan

1. Pengklasteran menggunakan metode *k-harmonic means* dengan jarak *euclidean*, diperoleh kluster yang optimal pada $k = 3$ dengan jarak *euclidean* dimana nilai $SC = 0,3040722675 \approx 0,3$.
2. Berdasarkan profilisasi hasil analisis kluster, diketahui bahwa pada kluster 1 faktor utama yang paling menonjol yaitu persentase pemberian Air Susu Ibu (ASI) eksklusif. Untuk kluster 2 faktor utama yang paling menonjol yaitu, persentase Bayi Berat Lahir Rendah (BBLR) ≤ 2.500 -gram lahir dengan selamat, sedangkan untuk kluster 3 faktor utama yang paling menonjol juga persentase Bayi Berat Lahir Rendah (BBLR) ≤ 2.500 -gram lahir dengan selamat dan persentase rumah tangga yang tidak memiliki fasilitas sanitasi yang layak dengan *centroid*/rata-rata paling tinggi diantara kluster lain.

5.2. Saran

1. Penelitian selanjutnya dapat dilakukan pengklasteran dengan pengklasteran dengan perbandingan metode *k-harmonic means* dan *k-means* untuk mengetahui pengklasteran mana yang lebih *robust* (kuat/sensitif). Metode *k-harmonic means* menggunakan *harmonic means* sebagai pusat klasternya dan *k-means* menggunakan Mean sebagai pusat klasternya.
2. Karena masih terdapat daerah-daerah dengan faktor-faktor yang cukup tinggi seperti persentase Bayi Berat Lahir Rendah (BBLR) ≤ 2.500 -gram lahir dengan selamat dan persentase rumah tangga yang tidak memiliki fasilitas sanitasi layak disusul dengan faktor persentase pemberian Air Susu Ibu (ASI) eksklusif, maka dari itu harus adanya kerjasama antara pemerintah pusat, pemerintah daerah dan masyarakat dalam meminimalisir empat faktor utama yang mejadi peyebab terjadi kasus *stunting* di Indonesia.

DAFTAR PUSTAKA

Aini, F.N., Palgunadi. S., dan Anggrainingsih. R. 2014. *Clustering Business Rocess Model Petri Net dengan Complete Linkage*. Jurnal ITSMART Vol 3. No. 2: Hal. 47-51.

Badan Pusat Statistik. 2010. Statistik Kesejahteraan Rakyat Tahun 2017. Jakarta.

Departemen Kesehatan RI. 2008. Riset Kesehatan Dasar Tahun 2007. Jakarta.

Folzmisner, P. 2005. *Identification of Multivariate Outliers: A Performance Study*. *Australian Journal of Statistics* Vol 34. No. 2: Hal. 127-138.

Gujarati, D. 2009. Dasar-dasar Ekonometrika Jilid 2. Jakarta: Erlangga.

- Hair, J.F., Anderson, R.E., Thatham, R.L., dan Black, W.C. 2010. *Multivariate Data Analysis Seventh Edition*. New Jersey: Pearson Education. Inc.
- Hamerly, G., & Elkan, C. 2002. *Alternatives to the k-means algorithm that find better Clusterings*. In Proceedings of the Eleventh International Conference on Information and Knowledge Management, 600-607.
- Han, J. dan Kamber, M. 2006. *Data Mining Concepts and Techniques Second Edition*. San Francisco: Elsever.
- Johnson, R.A. dan Wichern, D.W. 2007. *Applied Multivariate Statistical Analysis Six Edition*. New Jersey: Prentice Hill. Inc.
- Kementerian Kesehatan RI. 2010. Keputusan Menteri Kesehatan Nomor 1995/Menkes/SK/XII/2010 tentang Standar Antropometri Penilaian Status Gizi Anak. Jakarta.
- Kementerian Kesehatan RI. 2011. Riset Kesehatan Dasar Tahun 2010. Jakarta.
- Kementerian Kesehatan RI. 2013. Peraturan Menteri Kesehatan Nomor 75 Tahun 2013 tentang Angka Kecukupan Gizi yang Dianjurkan bagi Bangsa Indonesia. Jakarta.
- Kementerian Kesehatan RI. 2014. Peraturan Menteri Kesehatan Nomor 23 Tahun 2014 tentang Upaya Perbaikan Gizi. Jakarta.
- Kementerian Kesehatan RI. 2014. Riset Kesehatan Dasar Tahun 2013. Jakarta.
- Kementerian Kesehatan RI. 2016. Buku Saku Pemantauan Status Gizi Tahun 2015. Jakarta.
- Kementerian Kesehatan RI. 2016. Peraturan Menteri Kesehatan Nomor 39 Tahun 2016 tentang Pendoman Penyelenggaraan Program Indonesia Sehat dengan Pendekatan Keluarga. Jakarta.
- Kementerian Kesehatan RI. 2016. Peraturan Menteri Kesehatan Nomor 51 Tahun 2016 tentang Standar Produk Suplementasi Gizi. Jakarta.
- Kementerian Kesehatan RI. 2017. Buku Saku Pemantauan Status Gizi Tahun 2016. Jakarta.
- Kementerian Kesehatan RI. 2018. Buku Saku Pemantauan Status Gizi Tahun 2017. Jakarta.
- Kementerian Perencanaan Pembangunan Nasional (Bappenas) dan United Nations Children's Fund. 2017. Laporan Baseline SDG tentang Anak-Anak Indonesia. Jakarta: BAPPENAS dan UNICEF.
- Pen, J.M., Lozano, J.A., dan Larranaga, P. 1999. *An Empirical Comparison of Four Initialization Methods for The K-Means Algorithm*. Pattern Recognition Letters, Vol.20, hal. 1027-1040.
- Simamora. 2005. Analisis Multivariat Pemasaran Edisi Pertama. Jakarta: PT. Gramedia Pustaka Utama.

- Struyf, A., Hubert, M., dan Rousseeuw, P.J. 1997. *Integrating Robust Clustering Techniques in S-PLUS*. *Jurnal of Computation Statistics and Data Analysis* Vol 26. No. 1: Hal. 17-37.
- Supranto, J. 2004. *Analisis Multivariat : Arti dan Interpretasi*. Jakarta: PT. Rineka Cipta.
- WHO. 2014. *WHO Global Nutrition Target: Stunting Policy Brief*. Geneva.
- WHO. 2017. *Stunted Growth and Development*. Geneva.
- Zhang, B., Hsu, M., & Dayal, U. 1999. *Clustering Algorithm K-Harmonic Means-A Data Clustering Algorithm*. Hewlett-Packard Labs Technical Report HPL-1999-124.