

PENERAPAN METODE *TWO STEP CLUSTER* UNTUK PENGELOMPOKAN POTENSI DESA

(Studi Kasus Data Potensi Desa Kabupaten Konawe Tahun 2016)

Agung Hendra Setiawan¹, Noviana Pratiwi^{2*}

^{1, 2}. Jurusan Statistika, FST, Institut Sains & Teknologi AKPRID Yogyakarta

*Corresponden author : novianapratiwi@akprind.ac.id

Abstract. *Konawe is one of 17 regencies / cities in Southeast Sulawesi Province. Village, potential is one way to develop this district. By knowing the classification of village potential, the government can determine what potential can be developed so that development is right on target. Clustering is the process of grouping objects into clusters that have similarities. Some problems that are often encountered in cluster analysis are the measurement scale of variables is not the same and the number of large objects and the number of cluster is unknown. One approach to dealing with this problem without transforming these variables is to use the Two Step Cluster method. This study aims to apply the Two Step Cluster method by classifying villages in Konawe Regency. The results are cluster 2 better than cluster 1 so the government must pay more attention to the 13 variables in cluster 1.*

Keywords : *two step cluster, village potential data, konawe regency*

Abstrak. Kabupaten Konawe merupakan satu dari 17 kabupaten/kota yang menjadi bagian wilayah dari Provinsi Sulawesi Tenggara. Usalah satu cara untuk mengembangkan kabupaten Konawe adalah dengan meningkatkan potensi desa. Dengan mengetahui klasifikasi potensi desa maka pemerintah bisa menentukan potensi apa yang bisa dikembangkan agar pembangunan tepat sasaran. Pengklasteran adalah proses pengelompokan objek ke dalam klaster-klaster yang memiliki kemiripan. Beberapa masalah yang sering dijumpai dalam analisis klaster yaitu skala pengukuran variabel tidak sama dan jumlah objek besar serta jumlah klaster tidak diketahui. Salah satu pendekatan untuk menangani masalah ini tanpa mentransformasi variabel-variabel tersebut adalah dengan menggunakan metode *Two Step Cluster*. Penelitian ini bertujuan menerapkan metode *Two Step Cluster* dengan mengklasterkan kelurahan/desa yang berada di Kabupaten Konawe. Hasil pengklasteran dengan metode *Two Step Cluster* adalah klaster 2 lebih baik daripada klaster 1 sehingga pemerintah harus lebih memperhatikan 13 variabel-variabel pada claster 1.

Kata Kunci : *Metode Two Step Cluster, data potensi desa, kabupaten konawe*

1. Pendahuluan

Kabupaten Konawe merupakan satu dari 17 kabupaten/kota yang menjadi bagian wilayah dari Provinsi Sulawesi Tenggara, Ibukota Kabupaten Konawe terletak di Unaaha. Potensi perekonomian terus berkembang dengan banyaknya industri-industri rumah tangga yang tentu memberikan peran untuk penyerapan tenaga kerja bagi warga setempat dan berkembang kios-kios setiap desa. Potensi pendidikan dan kesehatan terus berkembang itu terlihat dari hampir setiap kelurahan/desa di Kabupaten Konawe memiliki fasilitas pendidikan dari TK sampai Perguruan Tinggi, dan fasilitas serta tenaga kesehatan yang hampir tersedia diseluruh kelurahan/desa di Kabubaten Konawe. Potensi perekonomian, potensi pendidikan dan kesehatan ini tersebar di wilayah kelurahan/desa Kabupaten Konawe. Untuk mengetahui kelurahan/desa yang memiki karakteristik potensi kelurahan/desa yang sama, dapat dilakukan pengelompokan dengan mengelompokkan kelurahan-kelurahan atau desa-desa berdasarkan sarana, prasarana, sosial, ekonomi, dan penduduk. Sehingga pemerintah Kabupaten Konawe dapat mengembangkan atau membangun kelurahan/desa secara merata sehingga dapat mensejahterakan seluruh masyarakat Kabupaten Konawe.

Dalam statistika, salah satu metode yang digunakan dalam mengelompokkan variabel atau objek adalah metode klaster. Metode klaster adalah proses mengelompokkan objek ke dalam kelompok-kelompok yang memiliki kemiripan. Dalam analisis klaster dijumpai skala

pengukuran variabel tidak sama dan jumlah objek besar serta jumlah kluster tidak diketahui. Misalnya Pada data potensi desa memiliki skala pengukuran variabel yang berbeda dan mempunyai jumlah data yang besar, maka untuk menangani data potensi desa seperti itu, diperlukan metode Two Step Cluster.

Penelitian ini akan menggunakan metode Two Step Cluster pada data potensi desa Kabupaten Konawe Tahun 2016.

2. Metode Klaster

Metode klaster merupakan salah satu dari teknik variabel ganda (multivariate) yang tujuan utama adalah mengelompokkan objek berdasarkan atas kriteria yang dimiliki. Metode klaster mengklasifikasi objek, sehingga antara satu objek dengan objek lainnya yang terletak dalam satu klaster akan memiliki kesamaan tinggi yang sesuai dengan kriteria pemilihan yang ditentukan. Hasil dari pengklasteran harus memperlihatkan keragaman yang homogen di dalam klaster dan keragaman yang heterogen antar klaster yang terbentuk.

Ukuran Jarak

Ukuran kemiripan dan ketakmiripan yang digunakan dalam Metode klaster adalah jarak objek dan jarak antar klaster. Fungsi jarak yang sering digunakan antara lain adalah :

Jarak Manhattan

Ukuran ini merupakan bentuk umum dari jarak *Euclidean* [4]. Fungsi jaraknya didefinisikan :

$$d(i, j) = \sum_{k=1}^p |x_{ik} - x_{jk}| \quad (1)$$

Keterangan :

$d(i, j)$ = Jarak antara objek i ke objek j

x_{ik} = Nilai objek i pada variabel ke- k

x_{jk} = Nilai objek j pada variabel ke- k

p = Jumlah variabel yang diamati

Jarak Euclidean

Jarak *Euclidean* adalah jarak yang paling umum dan paling sering digunakan dalam Metode klaster. Jarak *Euclidean* antara dua titik dapat terdefinisikan dengan jelas. Jarak ini digunakan apabila semua peubah yang digunakan adalah peubah kontinu [4].

Jarak *Euclidean* antara klaster ke- i dan klaster ke- j dari p peubah didefinisikan:

$$d(i, j) = \sqrt{\sum_{k=1}^p |x_{ik} - x_{jk}|^2} \quad (2)$$

Keterangan :

$d(i, j)$ = Jarak antara objek i ke objek j

x_{ik} = Nilai objek i pada variabel ke- k

x_{jk} = Nilai objek j pada variabel ke- k

p = Jumlah variabel yang diamati

Jarak Log-Likelihood

Jarak *Log-Likelihood* dapat digunakan untuk peubah bertipe kontinu (numerik) maupun kategorik. Asumsi yang ada pada jarak ini adalah peubah kontinu menyebar normal, peubah kategorik menyebar multinomial dan antar peubahnya saling bebas. Metode *Two Step Cluster* cukup tegar (*robust*) terhadap pelanggaran asumsi tersebut sehingga metode ini masih dapat digunakan ketika terjadi pelanggaran asumsi [6].

Jarak antara kluster j dan s didefinisikan sebagai berikut:

$$d(j, s) = \xi_j + \xi_s - \xi_{(j,s)} \quad (3)$$

Keterangan:

$$\begin{aligned} \xi_j &= -N_j \left(\sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 + \hat{\sigma}_{jk}^2) + \sum_{k=1}^{K^B} \hat{E}_{jk} \right) \\ \xi_s &= -N_s \left(\sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 + \hat{\sigma}_{sk}^2) + \sum_{k=1}^{K^B} \hat{E}_{sk} \right) \\ \xi_{(j,s)} &= -N \left(\sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 + \hat{\sigma}_{jsk}^2) + \sum_{k=1}^{K^B} \hat{E}_{jsk} \right) \\ \hat{E}_{jk} &= - \sum_{l=1}^{L_k} \frac{N_{jkl}}{N_j} \log \frac{N_{jkl}}{N_j} \end{aligned}$$

Keterangan :

N = Jumlah data

N_j = Jumlah data di dalam kluster j

N_{jkl} = Jumlah data di kluster j untuk variabel kategorik ke- k dengan ke- l

$\hat{\sigma}_k^2$ = Ragam dugaan untuk variabel kontinu ke- k untuk keseluruhan data

$\hat{\sigma}_{jk}^2$ = Ragam dugaan untuk variabel kontinu ke- k dalam kluster j

K_A = Jumlah variabel kontinu

K_B = Jumlah variabel kategorik

L_K = Jumlah kategori untuk variabel kategorik ke- k

$d(j, s)$ = Jarak antara kluster j dan s

(j, s) = Indeks kombinasi kluster j dan s

Metode Kluster Berhierarchy

Metode kluster berhierarchy digunakan jika banyak kluster yang akan dibentuk belum diketahui sebelum. Metode kluster berhierarchy ini dibedakan menjadi dua yaitu metode penggabungan (*agglomerative*) dan metode pemisahan (*divisive*). Metode *agglomerative* dimulai dengan n buah kluster yang masing-masing beranggotakan satu objek. Kemudian dua kluster yang paling dekat digabungkan dan ditentukan kembali kedekatan antar kluster yang baru. Proses ini berlanjut sampai didapatkan satu kluster yang anggotanya adalah seluruh objek. Metode *divisive* dimulai dengan satu kluster yang anggotanya adalah seluruh objek, kemudian objek-objek yang paling jauh dipisahkan dan membentuk kluster lain. Proses ini berlanjut sampai semua objek masing-masing membentuk satu kluster.

Dalam metode berhierarchy terdapat beberapa cara penggabungan kluster, antara lain jarak minimum atau pautan tunggal (*single linkage*), Jarak maksimum atau pautan lengkap (*complete linkage*), rata-rata dari semua jarak atau pautan rata-rata (*average linkage*). Jenis variabel yang dapat diklusterkan dengan metode ini adalah variabel kontinu (rasio dan interval) dan fungsi jarak yang sering digunakan dalam metode berhierarchy ini adalah jarak Eulidian atau jarak mahalanobis.

Metode Kluster Non Hierarchy

Metode kluster non hierarki digunakan jika banyak kluster yang akan dibentuk sudah diketahui sebelum. Contoh dari metode non hierarki adalah K-means. Langkah pertama dalam

k-means yaitu menentukan besarnya k, yaitu banyaknya kluster. Pemilihan k dapat ditentukan secara subyektif berdasarkan latar belakang bidang masing-masing. Fungsi jarak yang sering digunakan adalah jarak Euclidian. Jenis variabel yang dapat diklusterkan dengan metode ini adalah variabel kontinu (rasio dan interval).

Metode Kluster Dua Langkah (Two Step Cluster Method)

Metode *Two Step Cluster* adalah metode kluster yang dirancang untuk dapat mengatasi masalah skala pengukuran yang tidak sama, dalam hal ini bertipe kontinu dan kategorik, serta memiliki jumlah objek amatan relatif besar. Jarak yang digunakan dalam metode *Two Step Cluster* adalah jarak *Log Likelihood* untuk skala data tidak sama yaitu numerik dan kategorik sedangkan jarak *Euclidean* hanya untuk skala data numerik.

Ukuran jarak *Log-Likelihood* didasarkan pada tiga asumsi, yaitu antar variabel saling bebas, variabel kategorik diasumsikan berdistribusi multinomial, dan variabel numerik diasumsikan berdistribusi normal. Metode pengklusteran dua tahap cukup kekar (*robust*) terhadap asumsi kebebasan dan asumsi distribusi tersebut [6].

Prosedur pengklusteran objek dalam metode *Two Step Cluster* dilakukan melalui dua tahapan [1], yaitu tahap pembentukan kluster awal (*pre-clustering*) dan tahap pembentukan kluster optimal.

Tahap 1 : Pembentukan Kluster awal

Tahap pertama dari metode *Two Step Cluster* adalah pembentukan kluster awal (*pre-clustering*) yang menggunakan pendekatan kluster secara sekuensial, yaitu setiap pasang objek diamati berdasarkan ukuran jarak dan memutuskan objek tersebut akan bergabung dengan kluster yang telah terbentuk sebelumnya atau memulai kluster baru [5]. Pendekatan ini digambarkan dengan membentuk Pohon Ciri Kluster (*Cluster Feature Tree/CF Tree*) [7].

CF Tree terdiri dari tingkatan cabang (*node*) dan masing-masing cabang berisikan sejumlah data dientrikan. Apabila dimisalkan sebuah pohon, maka tingkatan cabang tersebut dari batang pohon, dahan dan daun. Dalam *CF Tree* tingkatan daun dikenal dengan nama daun entri (entrian pada cabang daun) dimana pada tingkatan ini merepresentasikan hasil sub-kluster yang diinginkan.

Jika tidak ada lagi tempat dalam cabang daun untuk menciptakan daun entri baru (node telah melewati batas maksimum), maka cabang daun akan terbagi menjadi dua. Jika dimisalkan pada sebuah pohon, dari satu dahan membelah menjadi dua dahan. Entrian pada cabang dua grup (dahan) dengan menggunakan pasangan terjauh sebagai penempatan dan membagi-bagikan sisa entrian berdasarkan kriteria kedekatan.

Jika tidak tersedia tempat dalam cabang dahan untuk menciptakan daun entri baru, maka cabang dahan akan terbagi menjadi dua. Apabila dimisalkan pada sebuah pohon, dari satu pohon kemudian membelah menjadi dua pohon. Entrian pada cabang dahan yang asli akan dibagi kedalam dua grup (pohon) dengan menggunakan pasangan dahan terjauh sebagai penempatan, dan membagi-bagikan kembali sisa entrian berdasarkan atas kriteria kedekatan yang telah ditetapkan. Proses ini berlanjut sampai semua data selesai dimasukkan.

Tahap 2 : Pembentukan kluster optimal

Pembentukan kluster akhir ditandai dengan terbentuknya kluster yang optimal. Suatu kluster dikatakan optimal apabila memiliki jarak antar kluster paling jauh dan jarak antar objek dalam kluster tersebut paling dekat. Semakin dekat jarak antar objek maka semakin besar kemiripan antar objek dalam satu kluster. Pada tahapan ini, sub-kluster (*leaf entry*) hasil dari tahap pertama digunakan sebagai input yang kemudian dikelompokkan menjadi sejumlah kluster yang diinginkan. Karena jumlah sub-kluster lebih sedikit daripada jumlah/ukuran data awal, maka metode penggabungan sederhana akan efektif untuk digunakan. Metode *Two Step Cluster* menggunakan metode kluster berhirarki dengan metode penggabungan (*agglomerative*) karena memiliki prosedur pengklusteran otomatis yang dapat dengan baik membentuk kluster [8]. Tiap-tiap sub-kluster yang terbentuk pada tahap pertama akan

digabungkan satu persatu sesuai ukuran jarak yang telah ditetapkan. Prosedur ini berakhir sampai seluruh sub-klaster menjadi satu klaster.

Penentuan Jumlah Klaster Optimal

Dalam menentukan jumlah klaster optimal dapat ditentukan secara otomatis, terdapat dua tahap yang harus dilakukan. Tahap pertama menduga banyaknya klaster maksimal dengan menghitung kriteria *BIC* (*Bayesian Information Criterion*) untuk tiap-tiap klaster. Rumus *BIC* (*Bayesian Information Criterion*) untuk jumlah klaster sebanyak *J* [1], adalah sebagai berikut :

$$BIC(J) = -2 \sum_{j=i}^J \xi_j + m_j \log(N) \tag{4}$$

Keterangan

$$\xi_j = -N_j \left(\sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 + \hat{\sigma}_{jk}^2) + \sum_{k=1}^{K^B} \hat{E}_{jk} \right)$$

$$\hat{E}_{jk} = - \sum_{l=1}^{L_k} \frac{N_{jkl}}{N_j} \log \frac{N_{jkl}}{N_j}$$

$$m_j = J \left\{ 2K^A + \sum_{k=i}^{K^B} (L_k - 1) \right\}$$

Keterangan :

N = jumlah data

N_j = Jumlah data di dalam klaster *j*

N_{jkl} = Jumlah data di klaster *j* untuk variabel kategorik ke-*k* dengan ke-*l*

$\hat{\sigma}_k^2$ = Ragam dugaan untuk variabel kontinu ke-*k* untuk keseluruhan data

$\hat{\sigma}_{jk}^2$ = Ragam dugaan untuk variabel kontinu ke-*k* dalam klaster *j*

K_A = jumlah variabel kontinu

K_B = jumlah variabel kategorik

L_k = jumlah kategori untuk variabel kategorik ke-*k*

Kemudian hasil perhitungan tersebut digunakan untuk menduga jumlah klaster. Langkah yang kedua yaitu mencari peningkatan jarak terbesar antara dua klaster terdekat pada masing-masing tahapan pengklasteran. Solusi klaster yang terbaik memiliki *BIC* terkecil, tetapi ada beberapa kasus dalam pengklasteran dimana *BIC* akan terus menurun nilainya bila jumlah klaster semakin meningkat. Maka dalam situasi tersebut, *ratio BIC Changes* (rasio perubahan *BIC*) dan *ratio of Distance Measure Changes* (rasio perubahan jarak) mengidentifikasi solusi klaster terbaik.

Menurut [3] *BIC* menghasilkan penduga awal yang baik bagi jumlah klaster maksimum. Jumlah klaster maksimum adalah banyaknya klaster yang memiliki rasio *BIC* yang pertama kali lebih kecil dari *c₁*. Nilai *c₁* = 0,04, berdasarkan studi simulasi [1].

Jumlah klaster yang terbentuk dapat diketahui dengan menggunakan perbandingan antar jarak untuk *k* klaster, dengan rumus perbandingannya sebagai berikut :

$$R(k) = \frac{d_{k-1}}{d_k} \tag{5}$$

$$d_k = l_{k-1} - l_k$$

Keterangan :

$$l_j = \frac{(m_j \log N - BIC(J))}{2}$$

$$J = k, k - 1$$

d_{k-1} = Jarak jika k kluster digabungkan dengan $k-1$ kluster

Jumlah kluster optimal diperoleh berdasarkan ketentuan ditemukannya perbedaan yang nyata pada rasio perubahan kluster. Rasio perubahan kluster dihitung sebagai berikut:

$$\frac{R(k_1)}{R(k_2)} \quad (6)$$

Untuk dua nilai terbesar dari $R(k_1)$ ($k = 1, 2, \dots, k_{max}$; k_{max} didapatkan dari langkah pertama). Jika rasio perubahan lebih besar daripada nilai batas c_2 (*SPSS* menetapkan nilai $c_2 = 1,15$) maka jumlah kluster di tetapkan sama dengan k_1 . Dengan cara lain penentuan jumlah kluster terbaik dapat didefinisikan sebagai solusi dari maksimum (k_1, k_2) [1].

Variabel Berpengaruh

Penentuan variabel yang berpengaruh dalam pembentukan setiap kluster dilakukan pada kedua jenis variabel [7]. Untuk menentukan variabel yang berpengaruh ditentukan melalui uji, statistik uji *Chi-Square* untuk variabel kategorik sedangkan variabel numerik menggunakan uji *t*.

- Uji untuk variabel kategorik:
 - $H_0: \pi_{jkl} = \pi_{kl}$ (Variabel kategorik tidak berpengaruh terhadap kluster)
 - $H_1: \pi_{jkl} \neq \pi_{kl}$ (Variabel kategorik berpengaruh terhadap kluster)
- Dengan tingkat signifikansi α yaitu 0,05 % dan statistik uji *Chi-Square* hitung didapatkan berdasarkan :

$$\chi^2 = \sum_{l=1}^{L_k} \left(\frac{N_{jkl} - N_{kl}}{N_{kl}} \right)^2 \quad (7)$$

Keterangan :

N_{jkl} :Jumlah data didalam kluster ke- j untuk variabel kategorik ke- k pada kategorik ke- l

N_{kl} : Jumlah data untuk variabel kategorik ke- k pada kategorik ke- l

L_k : Jumlah kategorik untuk variabel kategorik ke- k

- Apabila Statistik $\chi^2_{hit} > \chi^2_{(\alpha, L_k - 1)}$ maka hipotesis nol ditolak. Hal tersebut berarti bahwa variabel kategorik berpengaruh terhadap kluster. Pada software *SPSS* menampilkan statistik ini dalam bentuk grafik batang dimana nilai kritis digambarkan dengan garis putus-putus.
- Uji untuk variabel numerik :
 - $H_0: \mu_{jk} = \mu_k$ (Variabel numerik tidak berpengaruh terhadap kluster)
 - $H_1: \mu_{jk} \neq \mu_k$ (Variabel numerik berpengaruh terhadap kluster)
- Dengan tingkat signifikansi α yaitu 0,05 % dan statistik uji *t* hitung didapatkan berdasarkan :

$$t_{hit} = \frac{\hat{\mu}_{jk} - \hat{\mu}_k}{\hat{\sigma}_{jk} / \sqrt{N_{jk}}} \quad (8)$$

Keterangan :

$\hat{\mu}_k$: Rataan dari variabel numerik ke- k

$\hat{\mu}_{jk}$: Rataan dari variabel numerik ke- k di dalam kluster ke- j

$\hat{\sigma}_{jk}$: Simpangan baku variabel numerik ke- k di dalam kluster ke- j

N_{jk} : Jumlah observasi pada kluster ke- j untuk variabel numerik ke- k

- Apabila Statistik $t_{hit} > t_{(\frac{\alpha}{2}, N_{jk} - 1)}$ maka hipotesis nol ditolak. Hal tersebut berarti bahwa variabel kategorik berpengaruh terhadap kluster. Pada Software *SPSS* menampilkan

statistik ini dalam bentuk grafik batang dimana nilai kritis digambarkan dengan garis putus-putus.

3. Metode

Data yang digunakan dalam penelitian ini, merupakan data sekunder yaitu data potensi desa yang diperoleh langsung dari situs Badan Pusat Statistika (BPS) dengan jumlah variabel sebanyak 19 yang terdiri dari 2 variabel kategorik dan 17 variabel numerik sedangkan jumlah data sebanyak 329 kelurahan/desa [2].

Variabel Penelitian

Variabel – variabel yang digunakan dalam penelitian ini disesuaikan dengan ketersediaan yang ada. Adapun variabel kategorik yang digunakan adalah sebagai berikut : Status desa (X_1) yaitu : (1 = Kelurahan), (2 = desa) Lokasi lokasi (X_2) yaitu : (1 = Non pesisir), (2 = pesisir), Sedangkan variabel numerik yaitu Luas wilayah (Ha) (X_3), Jarak desa ke kecamatan (km) (X_4), Jarak desa ke kabupaten (km) (X_5), Jarak desa ke provinsi (km) (X_6), Jumlah penduduk (X_7), Jumlah kepadatan penduduk per-km² (X_8), Jumlah rumah tembok (unit) (X_9), Jumlah fasilitas pendidikan (unit) (X_{10}), Jumlah fasilitas kesehatan (unit) (X_{11}), Jumlah tenaga kesehatan (unit) (X_{12}), Jumlah tempat ibadah (unit) (X_{13}), Jumlah industri kecil rumah tangga (X_{14}), Jumlah RT pengguna PLN (X_{15}), Jumlah RT pengguna gas (X_{16}), Jumlah sepeda motor (X_{17}), Jumlah pengguna TV (X_{18}) dan Jumlah kios (X_{19}).

Prosedur Penelitian

Prosedur penelitian yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Mengumpulkan data
Pengumpulan data dilakukan dengan cara mengambil langsung dari website BPS kabupaten Konawe
2. Mendeskripsikan potensi desa/kelurahan kabupaten Konawe
3. Melakukan pengklasteran *Two Step Cluster* terhadap variabel-variabel kontinu dan kategorik, dengan langkah-langkah sebagai berikut:
 - Tahap pertama :
 1. Memilih objek untuk dijadikan sub-klaster awal secara random, serta menentukan jumlah batas maksimum objek yang dapat bergabung pada sub-klaster.
 2. Memasukkan objek satu persatu untuk membentuk sub-klaster dengan menghitung jarak berdasarkan ukuran jarak *Log Likelihood*. Sub-klaster terdiri dari kombinasi objek anggota-anggota yang masuk sebagai sub klaster.
 3. Penggabungan terus digabungkan hingga semua obyek anggota bergabung pada suatu sub-klaster. Sehingga diperoleh pembentukan klaster awal yang merupakan hasil dari pembentukan sub-sub klaster pada tahap pertama.
 - Tahap kedua :
 1. Menghitung nilai tengah sub-sub klaster yang terbentuk pada tahap pertama.
 2. Hasil nilai tengah sub-sub klaster pada tahap pertama diklasterkan menggunakan Metode klaster berhirarki dengan metode penggabungan (*agglomerative*).

3. Melakukan pemilihan jumlah kluster optimum dari hasil pembentukan *Cluster Feature (CF) Tree*. Pemilihan kluster dilakukan berdasarkan kriteria nilai *BIC* terkecil pada persamaan (4) atau dengan rasio perubahan jarak terbesar pada persamaan (6).
4. Hasil kluster-kluster yang terbentuk
7. Menjelaskan Karakteristik Kluster-kluster yang terbentuk
8. Kesimpulan

4. Hasil dan Pembahasan

Menentukan jumlah kluster optimal

Dalam menentukan jumlah kluster optimal digunakan kriteria penentuan optimum dengan melihat kriteria *BIC* (*Bayesian Information Criterion*) dengan menggunakan persamaan (2.9). Pada kasus ini, nilai *BIC* terus menurun seiring dengan meningkatnya jumlah kluster. Maka solusi terbaik memilih dari nilai rasio perubahan jarak (*Ratio of Distance Measures*) dengan menggunakan persamaan (2.1) yang terbesar.

Tabel 1 *BIC* (*Bayesian Information Criterion*)

Number of cluster	Schwarz's Bayesian Criterion (BIC)	BIC Change	Ratio of BIC Changes	Ratio of Distance Measure
1	4566,159			
2	3773,031	-793,128	1	2,486
3	3578,791	-194,240	0,245	1,314
4	3480,881	-97,910	0,123	1,212
5	3436,622	-44,258	0,056	1,428
6	3468,158	31,535	-0,040	1,274
7	3537,802	69,644	-0,088	1,246
8	3634,931	97,129	-0,122	1,436
9	3765,903	130,973	-0,165	1,050
10	3900,545	134,641	-0,170	1,095
11	4041,607	141,062	-0,178	1,174
12	4192,680	151,072	-0,190	1,250
13	4355,255	162,575	-0,205	1,050
14	4520,029	164,774	-0,208	1,080
15	4688,064	168,036	-0,212	1,000

Berdasarkan **Tabel 1**, menjelaskan kluster yang dihasilkan pada tahap pertama sebanyak 6 kluster. Hal ini dilihat dari nilai *Ratio of BIC Changes* dimana nilai kurang atau sama dengan 0,04 ($c_1=0,04$, didapat dari simulasi penelitian yang dilakukan oleh Bacher et al, penulis dari *SPSS Two Step Cluster* pada tahun 2004), pada kluster 5 nilai *Ratio of BIC Changes* tersebut sebesar -0,04. Sedangkan pada penentuan kluster optimal dilihat dari 2 nilai $R(k)$ terbesar yaitu 2,486 pada kluster 2 dan 1,436 pada kluster 8, apabila rasio dari ke dua $R(k)$ lebih besar atau sama dengan 1,15 ($R(k)$ dilihat dari nilai *Ratio of Distance Measures* dan nilai $C_2 = 1,15$ didapat pula dari simulasi penelitian yang dilakukan oleh Bacher et al, penulis *SPSS Two Step Cluster* pada tahun 2004), maka solusi kluster optimal adalah yang memiliki nilai $R(k)$ terbesar. Didasarkan pada nilai Rasio antara kedua nilai $R(k)$ adalah 1,73 dan lebih besar dari nilai konstanta $c_2 = 1,15$. Oleh karena itu, dalam kasus ini 2 kluster merupakan solusi kluster optimal.

Hasil klaster yang terbentuk

Distribusi anggota dari masing-masing klaster yang terbentuk untuk lebih jelasnya dapat dilihat pada **Tabel 2**. Dari 329 kelurahan/desa yang berada di Sulawesi Tenggara, sebanyak 247 atau 75,1 % kelurahan/desa berasal dari klaster 1, dan 82 atau 24,9 % kelurahan/desa berasal dari klaster 2.

Tabel 2 Distribusi klaster potensi kelurahan/desa Kabupaten Konawe

		N	% of Combined	% of Total
Cluster	1	247	75,10%	75,10%
	2	82	24,90%	24,90%
	Combined	329	100,00%	100,00%
Total		329		100,00%

Karakteristik Klaster kelurahan/desa Kabupaten Konawe

Karakteristik untuk variabel-variabel kategorik dan numerik potensi kelurahan/desa untuk di Kabupaten Konawe yang terdiri dari 2 klaster yang terbentuk.

Tabel 3. Karakteristik dan variabel-variabel berpengaruh

Klaster		X1	X2	X4	X5	X7	X8	X9	X10
1	Mean			4,9449	48,3057	507,923	94,4693	34,0729	1,251
	Variabel Berpengaruh	✓	-	-	-	✓	✓	✓	✓
2	Mean			3,0402	16,3927	1250,26	239,222	130,671	2,9878
	Variabel Berpengaruh	✓	✓	✓	✓	✓	✓	✓	✓

Klaster		X11	X13	X14	X15	X16	X17	X18	X19
1	Mean	1,4	1,6235	12,68	82,915	81,7247	64,8381	74,2996	9,9636
	Variabel Berpengaruh	✓	✓	✓	-	-	✓	✓	✓
2	Mean	1,6	3,2561	18,854	260,976	199,866	143,39	236,878	24,244
	Variabel Berpengaruh	✓	✓	✓	✓	✓	✓	✓	✓

Berdasarkan **Tabel 3**, menjelaskan bahwa variabel yang digunakan ada 19 variabel, 2 variabel kategorik yaitu X_1 dan X_2 sedangkan variable numerik X_3 - X_{19} . Sebagai berikut:

- Klaster 1 dipengaruhi oleh variabel kategorik hanya pada X_1 (Status desa), dan untuk variabel numerik sebanyak 11 variabel yaitu jumlah penduduk (X_7), jumlah kepadatan penduduk per- km^2 (X_8), jumlah rumah tembok (X_9), jumlah fasilitas pendidikan (X_{10}), jumlah tenaga kesehatan (X_{12}), jumlah tempat ibadah (X_{13}), jumlah pengguna gas (X_{16}), jumlah sepeda motor (X_{17}), jumlah pengguna TV (X_{18}) dan jumlah kios (X_{19}). Sedangkan semua variabel memiliki nilai mean rendah. Hal ini menandakan potensi desa yang berada klaster 1 memiliki banyak kekurangan, sehingga pemerintah harus lebih memprioritaskan kelurahan/desa pada klaster 1 terutama variabel-variabel yang dapat dikembangkan lebih

oleh pemerintah seperti jumlah rumah tembok, fasilitas dan tenaga kesehatan serta pendidikan, industri kecil rumah tangga dan jumlah kios sehingga kelurahan/desa pada klaster 1 dapat berkembang lebih baik.

- Klaster 2 dipengaruhi oleh kedua dua variabel kategorik yaitu X_1 dan X_2 (Status desa dan lokasi), dan untuk variabel numerik sebanyak 16 variabel yaitu jarak desa ke kecamatan (X_4), jarak desa ke kabupaten (X_5), jumlah penduduk (X_7), jumlah penduduk (X_7), jumlah kepadatan penduduk per- km^2 (X_8), jumlah rumah tembok (X_9), jumlah tenaga kesehatan (X_{12}), jumlah tempat ibadah (X_{13}), jumlah industri kecil rumah tangga (X_{14}), jumlah pengguna PLN (X_{15}), jumlah pengguna gas (X_{16}), jumlah sepeda motor (X_{17}), jumlah pengguna TV (X_{18}) dan jumlah kios (X_{19}). Sedangkan nilai mean klaster 2 semua bernilai besar yaitu sebanyak 12 variabel.

Dapat disimpulkan bahwa klaster 2 lebih baik dari pada klaster 1, sehingga pemerintah harus lebih memprioritaskan kelurahan/desa pada klaster 1 agar potensi kelurahan/desa lebih baik lagi dan dapat mensejahterakan masyarakat yang berada pada kelurahan/desa klaster.

5. Kesimpulan

Metode Two Step Cluster merupakan analisis klaster dua tahap yang baik digunakan untuk ukuran data yang besar dan tipe data yang berbeda yaitu kontinu (numerik) dan kategorik. Pengklasteran *Two Step Cluster* pada data Potensi Desa Kabupaten Konawe tahun 2016 dengan menggunakan 19 variabel yang terdiri dari 2 variabel kategorik dan 17 variabel numerik menghasilkan 2 klaster optimal.

Klaster 2 memiliki potensi desa lebih baik daripada klaster 1, sehingga pemerintah harus meningkatkan potensi desa pada klaster 1 agar kelurahan/desa yang berada di klaster 1 dapat meningkatkan potensi fisik seperti sarana, prasarana, dan akses maupun potensi non fisik seperti sosial budaya, ekonomi dan penduduk, untuk mendapatkan kesejahteraan secara merata di Kabupaten Konawe.

Ucapan Terima Kasih

Penyusunan tulisan ini telah melibatkan banyak pihak dengan memberikan doa dan dukungan kepada penulis. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan terima kasih kepada seluruh dosen pengajar dan pimpinan Jurusan Statistika Institut Sains & Teknologi AKPRIND Yogyakarta.

Daftar Pustaka

- [1] Bacher J, Wenzig K, Vogler M, 2004, SPSS two step cluster – a first evaluation. *RC33 Sixth International Conference on Social Science Methodology: Recent Developments and Applications on Social Science Research Methodology*; Amsterdam, Netherlands. Tersedia pada <http://www.statisticalinnovations.com/products/Two Step. Pdf>
- [2] Badan Pusat Statistik. 2017. *Kabupaten Konawe dalam Angka 2017*. BPS, Jakarta.
- [3] Chiu T, Fang D, Chen J, Wang Y., and Jeris C. 2001. *A robust and scalable clustering algorithm for mixed type attributes in large database environment*. Di dalam: *Doheon Lee, Mario Schkolnick, Foster J Provost, Ramakrishnan Srikant. Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2001)*; 2001 Agus 26-29; San Francisco, United States. New York (US): ACM Press. hlm 263-264.
- [4] Johnson,R.A, and Wichern,D.W, 2002, *Applied Multivariat Statistics*, Ed ke-5, New Jersey : prantice Hall.
- [5] Li H, Sun J. 2011, *Mining business failure predictive knowledge using two-step clustering*. *AJBM*. 5(11):4107-4120.
- [6] Norusis,M.J, 2004, *SPSS 19.0 Statistical Procedures Companion*. Upper Saddle River, NJ (US): Prentice-Hall. hlm 375-404.

- [7] Schiopu, D. 2010, *Applying two step cluster analysis for identifying bank customers' profile*. EI-TC. 62(3):66-75.
- [8] SPSS Inc, (2001), *The SPSS Two Step Cluster Component. A scalable Component Enabling More Efficient Customer Segmentation*. White paper-technical report, Schweiz. [terhubung berkala].