

PENGGEROMBOLAN DESA/KELURAHAN BERDASARKAN INDIKATOR KEMISKINAN DENGAN MENERAPKAN ALGORITMA TSC DAN K-PROTOTYPES*

Andrew Donda Munthe¹, I Made Sumertajaya^{2‡}, Utami D. Syafitri³

¹Department of Statistics, Bogor Agricultural University (IPB), Indonesia, dondaandrew@gmail.co.id

²Department of Statistics, Bogor Agricultural University (IPB), Indonesia, imsjaya.stk@gmail.com

³Department of Statistics, Bogor Agricultural University (IPB), Indonesia, utamids@apps.ipb.ac.id

‡corresponding author

Indonesian Journal of Statistics and Its Applications (eISSN:2599-0802)

Vol 2 No 2 (2018), 63 - 76

Copyright © 2018 Andrew Donda Munthe, I Made Sumertajaya, and Utami Dyah Syafitri. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Statistic Indonesia (BPS) noted that in 2014 there were 3.270 villages in Nusa Tenggara Timur Province. Most of them have a high percentage of poverty. Therefore, the village clustering based on poverty indicators is very important. The clustering algorithm that can be used on large data size and with mixed variables are Two Step Cluster (TSC) and K-Prototypes. The purpose of this research is to compare of TSC and K-Prototypes algorithm for village clustering in Nusa Tenggara Timur Province based on poverty indicators. The data were taken from 2014 village potential data (PODES 2014) collected by BPS. The best selection criteria for the cluster is the minimum ratio between variance within groups and variance between groups. The result showed that the best clustering algorithm was TSC which had the smallest ratio (2.6963). The best clustering showed that villages in Nusa Tenggara Timur Province divided into six groups with different characteristics.

Keywords: clustering, K-Prototypes algorithm, Two Step Cluster, villages.

1. Pendahuluan

Salah satu agenda prioritas pembangunan Nasional (2014–2019) yang merupakan butir ketiga dari kebijakan Nawacita adalah “Membangun Indonesia dari pinggiran

* Received Okt 2018; Accepted Nov 2018; Published online on Nov 2018

dengan memperkuat daerah-daerah dan desa dalam kerangka negara kesatuan.” Hal ini dimaksudkan untuk menjawab persoalan dan kerentanan akibat dari ketimpangan pembangunan yang telah dilakukan (Kemendes PDTT, 2015).

Karakteristik wilayah desa/kelurahan yang berbeda-beda berakibat pada pendekatan program kebijakan pembangunan yang berbeda juga antar wilayah. Pengelompokan desa/kelurahan dapat mengatasi permasalahan karakteristik wilayah yang berbeda-beda tersebut. Hal ini perlu dilakukan agar program kebijakan yang dilakukan oleh pemerintah dapat lebih fokus dan terarah sesuai dengan karakteristik dari hasil pengelompokan desa/kelurahan yang terbentuk.

Algoritma yang mampu mengatasi permasalahan pada ukuran data yang besar dan peubah campuran adalah algoritma *K-Prototypes* dan *Two Step Cluster* (TSC). Algoritma *K-Prototypes* (Huang, 1997 & 1998) merupakan salah satu algoritma yang pertama kali diperkenalkan untuk menerapkan analisis gerombol pada data berukuran besar dengan peubah campuran. Algoritma ini mengkombinasikan ukuran jarak pada algoritma *k-means* (peubah numerik) dan *k-modes* (peubah kategorik).

Algoritma lain yang dikembangkan selain *K-Prototypes* adalah algoritma *Two Step Cluster* (TSC). Algoritma ini mulai diperkenalkan pada awal tahun 2000-an. Penggerombolan dengan algoritma ini dilakukan dalam dua tahapan yaitu tahap membentuk gerombol awal dan kemudian dilanjutkan dengan tahap pembentukan gerombol optimal (Bacher et al., 2004).

Kebaikan dari kualitas penggerombolan diantara kedua algoritma dapat diketahui melalui rasio simpangan baku dalam kelompok (S_w) dan simpangan baku antar kelompok (S_B). Algoritma penggerombolan dikategorikan semakin baik apabila menghasilkan rasio antara S_w dan S_B yang nilainya semakin kecil (Bunkers et al., 1996).

Penerapan analisis gerombol untuk mengelompokkan wilayah desa/kelurahan di Provinsi Nusa Tenggara Timur sangat diperlukan. Provinsi Nusa Tenggara Timur (NTT) dipilih sebagai lokasi penelitian karena memiliki permasalahan yang sangat kompleks apabila dibandingkan dengan daerah lain. Salah satu isu utama di provinsi ini adalah terkait dengan angka kemiskinan yang tinggi. Persentase penduduk miskin di NTT pada September 2017 merupakan yang tertinggi ketiga di Indonesia yaitu mencapai 21.38 persen dari total penduduknya (BPS NTT (2014), BPS (2016) & BPS (2018)).

Kendala dalam menyajikan data kemiskinan hingga level desa/kelurahan telah dikaji oleh beberapa peneliti. Aminah (2017) menggunakan metode *Spasial Empirical Best Linear Unbiased Prediction* (SEBLUP) dalam melakukan pendugaan proporsi penduduk miskin di Kabupaten Wonosobo pada level kecamatan. Sulistiyono (2016) melakukan pemetaan kemiskinan di Kabupaten Pati pada level administrasi pemerintahan yang terendah yaitu hingga tingkat desa. Penggunaan peubah-peubah dari kedua penelitian di atas dapat diterapkan untuk mengelompokkan desa/kelurahan di NTT berdasarkan indikator kemiskinan. Oleh karena itu, penelitian ini akan menerapkan analisis gerombol dengan algoritma TSC dan algoritma *K-Prototypes* untuk mengelompokkan desa/kelurahan di Provinsi Nusa Tenggara Timur berdasarkan indikator kemiskinan pada level desa/kelurahan.

2. Metodologi Penelitian

2.1 Data

Data yang digunakan dalam penelitian ini adalah data sekunder hasil pendataan Potensi desa (Podes) tahun 2014 yang dikumpulkan oleh Badan Pusat Statistik (BPS). Satuan pengamatan dalam penelitian ini adalah seluruh desa dan kelurahan di Provinsi Nusa Tenggara Timur dengan total sebanyak 3.270 desa/kelurahan dengan rincian 2.951 desa dan 319 kelurahan.

Peubah yang digunakan sebanyak 13 peubah campuran (9 peubah numerik dan 4 peubah kategorik). Penggunaan peubah numerik didasarkan pada penelitian terkait indikator kemiskinan yang dilakukan oleh Aminah (2017) dan Sulistiyono (2016). Penggunaan peubah kategorik didasarkan pada konsep kebutuhan dasar (Green, 1978) serta mempertimbangkan ketersediaan data penelitian. Peubah-peubah yang akan digunakan pada penelitian ini adalah sebagai berikut:

- Persentase keluarga pertanian (X1)
- Persentase keluarga pengguna listrik (X2)
- Rasio sarana pendidikan per 1000 penduduk (X3)
- Rasio sarana kesehatan per 1000 penduduk (X4)
- Rasio tenaga kesehatan per 1000 penduduk (X5)
- Persentase penduduk penerima kartu Jamkesmas/Jamkesda (X6)
- Rasio penerima layanan Surat Keterangan Tidak Mampu (SKTM) per 1000 penduduk (X7)
- Jarak kantor desa/kelurahan ke kantor kecamatan (X8)
- Jarak kantor desa/kelurahan ke kantor bupati/walikota (X9)
- Topografi wilayah desa/kelurahan (X10)
- Tempat buang air besar sebagian besar keluarga (X11)
- Sumber air untuk minum sebagian besar keluarga (X12)
- Keberadaan angkutan umum yang melewati desa/kelurahan (X13)

2.2 Metode

Langkah-langkah analisis data yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Melakukan pengecekan kelengkapan data hasil pendataan Potensi desa (Podes) tahun 2014 Provinsi Nusa Tenggara Timur.
2. Melakukan penggerombolan data desa/kelurahan dengan menggunakan algoritma *Two Step Cluster* (TSC). Sebelum proses TSC dijalankan maka terlebih dahulu dilakukan pembakuan terhadap semua peubah numerik. Selanjutnya proses TSC dijalankan dengan tahapan (Chaudhary & Sharma, 2014):
 - a. Melakukan penggerombolan awal (*preclustering*) melalui pendekatan sekuensial dengan membentuk *Cluster Feature (CF) Tree*. Artinya objek diamati satu per satu berdasarkan ukuran jaraknya. Selanjutnya ditentukan apakah objek tersebut terkategori dalam gerombol yang telah terbentuk atau justru membentuk gerombol baru. Tahapan ini bertujuan untuk meminimalisasi jumlah amatan yang relatif besar. Penggerombolan awal pembentukan *Cluster Feature (CF) Tree* menggunakan ukuran jarak *log-*

likelihood dilakukan untuk menghasilkan *subcluster* dengan rumus perhitungan sebagai berikut:

$$d(j,s) = \xi_j + \xi_s - \xi_{(j,s)} \quad (1)$$

dimana,

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

$$\xi_s = -N \left(\sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 + \hat{\sigma}_{sk}^2) - \sum_{k=1}^{K^B} \sum_{l=1}^{L_k} \frac{N_{skl}}{N_j} \log \left(\frac{N_{skl}}{N_j} \right) \right)$$

$$\xi_{(js)} = -N \left(\sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 + \hat{\sigma}_{(js)k}^2) - \sum_{k=1}^{K^B} \sum_{l=1}^{L_k} \frac{N_{(js)kl}}{N_j} \log \left(\frac{N_{(js)kl}}{N_j} \right) \right)$$

dengan:

$d(j,s)$ = jarak antara gerombol ke- j dengan gerombol ke- s .

N = banyaknya objek

N_j = jumlah objek di dalam gerombol ke- j .

N_{jkl} = jumlah objek di dalam gerombol ke- j untuk peubah kategorik ke- k dengan kategori ke- l

$\hat{\sigma}_k^2$ = ragam dugaan untuk peubah kontinu ke- k untuk keseluruhan objek

$\hat{\sigma}_{jk}^2$ = ragam dugaan untuk peubah kontinu ke- k untuk keseluruhan objek dalam gerombol ke- j

K^A = banyaknya peubah kontinu

K^B = banyaknya peubah kategorik

L_k = banyaknya kategori untuk peubah kategorik ke- k

b. Melakukan proses penggerombolan akhir (gerombol optimal)

Pada tahapan ini, hasil *subcluster* yang terbentuk pada tahapan sebelumnya digerombolkan menggunakan metode hirarki penggabungan (*agglomerative*). Tujuan yang ingin dicapai pada tahapan ini adalah membentuk gerombol optimal. Suatu gerombol dikatakan optimal apabila memiliki jarak antar gerombolnya paling jauh dan jarak antar objek dalam gerombol tersebut paling dekat. Semakin dekat jarak antar objek maka semakin besar kemiripan antar obyek dalam satu gerombol.

Terdapat dua langkah yang harus dilakukan untuk menentukan jumlah gerombol optimal. Langkah yang pertama yaitu menghitung BIC (*Bayesian Information Criterion*) atau AIC (*Akaike's Information Criterion*) untuk tiap-tiap gerombol. Rumus BIC dan AIC untuk Gerombol J adalah sebagai berikut:

$$\text{BIC}(J) = -2 \sum_{j=1}^J \xi_j + m_j \log(N) \quad (2)$$

$$\text{AIC}(J) = -2 \sum_{j=1}^J \xi_j + 2m_j \quad (3)$$

dengan:

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

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

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

N = jumlah total observasi

N_j = jumlah observasi di dalam gerombol j

N_{jkl} = jumlah data di gerombol ke- j , peubah kategorik ke- k , kategori ke- l

$\hat{\sigma}_k^2$ = ragam dugaan untuk peubah kontinu ke- k untuk keseluruhan observasi

$\hat{\sigma}_{jk}^2$ = ragam dugaan untuk peubah kontinu ke- k dalam gerombol ke- j

K^A = banyaknya peubah kontinu

K^B = banyaknya peubah kategorik

L_k = banyaknya kategori untuk peubah kategorik ke- k

Langkah kedua adalah mencari peningkatan jarak terbesar antara dua gerombol terdekat pada masing-masing tahapan pengelompokan. Berdasarkan nilai BIC dan AIC dengan rumus diatas maka hasil perhitungannya dapat dipergunakan untuk menduga jumlah gerombol.

Apabila nilai BIC atau AIC tidak terus menurun seiring bertambahnya banyaknya gerombol maka gerombol optimal ditentukan berdasarkan nilai AIC atau BIC terkecil. Akan tetapi, apabila nilai BIC atau AIC terus menurun, penentuan gerombol maksimal menggunakan rasio perubahan BIC yang paling mendekati $c_1=0.04$ dan penentuan gerombol optimal menggunakan rasio perubahan jarak yang paling maksimum (Bacher *et al.*, 2004). Banyaknya gerombol optimal dilakukan dengan menemukan perbedaan nyata pada rasio perubahan jarak. Rumus menghitung rasio perubahan jarak adalah sebagai berikut:

$$R(k) = d_{k-1}/d_k \tag{4}$$

$$d_k = l_{k-1} - l_k \tag{5}$$

dengan:

$l_v = (m_j \log(N) - BIC_v)/2$ atau

$l_v = (2m_j - AIC_v)/2$

$v = k, k-1$

N = jumlah total observasi

d_k = jarak untuk k gerombol

d_{k-1} = jarak jika k gerombol digabungkan dengan $k-1$ gerombol

$R(k)$ = rasio perubahan jarak

Jumlah gerombol optimal diperoleh berdasarkan ketentuan ditemukannya perbedaan yang nyata pada rasio perubahan gerombol yaitu dengan melakukan perhitungan pada dua rasio perubahan jarak terbesar. Rumus perhitungannya adalah sebagai berikut:

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

dengan :

$R(k_1)$ = rasio perubahan jarak terbesar pertama

$R(k_2)$ = rasio perubahan jarak terbesar kedua

Jika rasio perubahan lebih besar dari nilai batas c_2 yaitu sebesar 1.15 (Bacher et al., 2014) maka jumlah gerombol optimal ditetapkan sama dengan k_1 . Selain dari hasil tersebut maka gerombol optimal sama dengan maksimum $\{k_1, k_2\}$.

3. Melakukan penggerombolan desa/kelurahan menggunakan algoritma *K-Proyotypes*. Sama seperti pada algoritma TSC, terlebih dahulu dilakukan pembakuan terhadap semua peubah numerik kemudian baru dilanjutkan dengan tahapan sebagai berikut:
 - a. Menentukan jumlah gerombol (k) yang akan dibentuk.
 - b. Menentukan inisial pada sejumlah gerombol (k) yang sudah ditentukan yaitu Z_1, Z_2, \dots, Z_k sebagai pusat cluster di masing-masing gerombol.
 - c. Melakukan perhitungan jarak seluruh data observasi pada dataset terhadap inisial gerombol awal. Ukuran jarak yang digunakan adalah ukuran jarak campuran dengan rumus:

$$d_2(X, Y) = \sum_{j=1}^p (x_j - y_j)^2 + \gamma \sum_{j=p+1}^m \delta(x_j, y_j) \quad (7)$$

dengan:

$d_2(X, Y)$ = ukuran jarak antara objek X dan Y (atribut data campuran)

$\sum_{j=1}^p (x_j - y_j)^2$ = ukuran jarak untuk atribut data yang bertipe numerik

$\gamma \sum_{j=p+1}^m \delta(x_j, y_j)$ = ukuran jarak untuk atribut data yang bertipe nominal/kategorik

γ = parameter penimbang

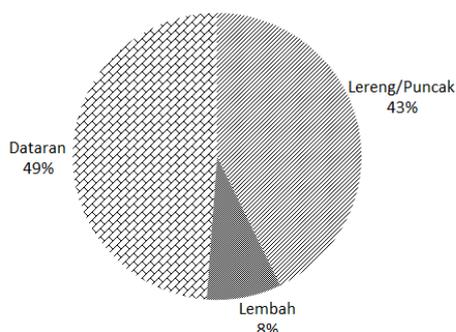
- d. Melakukan alokasi semua observasi ke dalam gerombol yang memiliki jarak prototype terdekat dengan objek yang diukur.
- e. Melakukan perhitungan titik pusat gerombol yang baru setelah semua objek dialokasikan. Selanjutnya merealokasikan semua data observasi pada dataset terhadap prototype yang baru.
- f. Apabila titik pusat gerombol tidak berubah atau sudah konvergen maka proses algoritma berhenti. Akan tetapi jika titik pusat masih berubah-ubah secara signifikan maka proses kembali ke tahap (2) sampai dengan (5) hingga iterasi maksimum tercapai atau dengan kata lain sudah tidak ada lagi perpindahan objek.

- g. Melakukan proses penggerombolan algoritma ini pada ukuran k yang lain. Dalam penelitian ini ukuran k yang digunakan dalam penggerombolan K -Prototypes adalah $k=2$ hingga $k=9$.
 - h. Gerombol optimal pada algoritma K -Prototypes adalah gerombol dengan rasio keragaman yang terkecil dari sejumlah k yang digunakan pada penelitian ini.
4. Hasil penggerombolan dari kedua algoritma dievaluasi dengan melakukan perhitungan rasio simpangan baku dalam kelompok (S_W) dan simpangan baku antar kelompok (S_B). Gerombol terbaik adalah gerombol dengan nilai rata-rata rasio S_W dan S_B terkecil untuk keseluruhan peubah.
 5. Melakukan visualisasi dan interpretasi karakteristik gerombol terbaik dalam mengelompokkan desa/kelurahan berdasarkan indikator kemiskinan. Visualisasi tersebut dilakukan dengan penerapan:
 - a. *Chernoff Face*
Chernoff Face adalah visualiasi hasil penggerombolan untuk peubah-peubah numerik dengan menghitung nilai rata-rata dari semua peubah pada masing-masing gerombol. *Chernoff Face* akan menghasilkan bentuk wajah (kepala, mulut, hidung, telinga) yang mewakili nilai rata-rata variabel numerik dengan bentuk dan ukuran sesuai dengan karakteristik gerombol yang terbentuk (Chernoff, 1973).
 - b. Peta Tematik

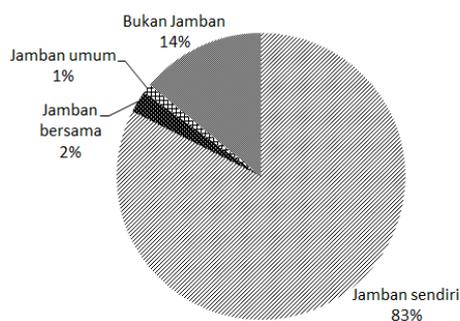
3. Hasil dan Pembahasan

3.1 Deskripsi Data

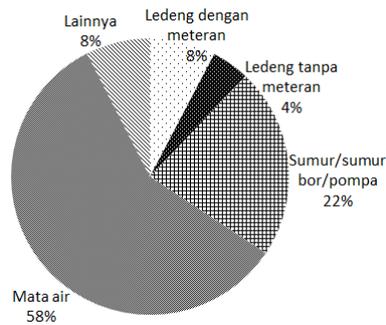
Topografi di sebagian besar wilayah NTT adalah dataran dan juga lereng/puncak (Gambar 1a). Desa/kelurahan di NTT yang penggunaan jamban sendiri sebagai tempat buang air besar, persentasenya cukup tinggi yaitu mencapai 83 persen (Gambar 1b). Dilihat dari sumber air minum maka desa/kelurahan yang masyarakatnya memanfaatkan mata air sebagai sumber air minum adalah kategori dengan persentase terbesar (Gambar 1c). Keberadaan angkutan umum dengan trayek tetap maupun tanpa trayek tetap sudah dinikmati di sebagian besar wilayah di NTT. Akan tetapi yang perlu menjadi catatan adalah masih terdapat 319 wilayah desa yang tidak memiliki angkutan umum atau persentasenya mencapai 10 persen (Gambar 1d).



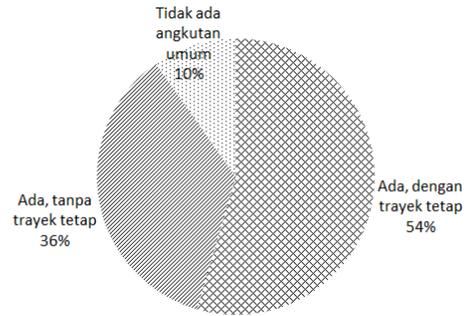
Gambar 1 (a) Persentase topografi wilayah desa/kelurahan



Gambar 1 (b) Persentase tempat buang air besar sebagian keluarga

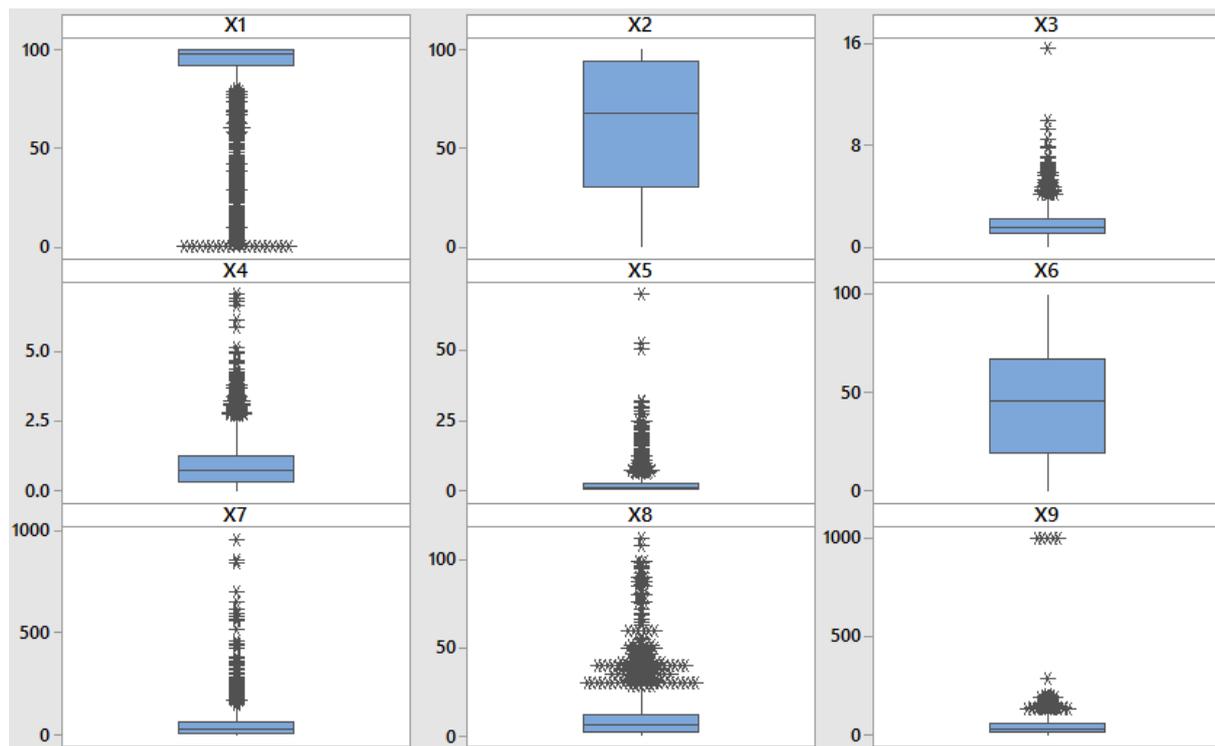


Gambar 1 (c) Persentase sumber air untuk minum sebagian besar keluarga



Gambar 1 (d) Persentase keberadaan angkutan umum yang melewati desa/kelurahan

Gambaran karakteristik dari peubah-peubah numerik untuk mengetahui pemusatan dan penyebaran data menggunakan diagram kotak garis (*Boxplot*) seperti ditunjukkan pada Gambar 2. Secara visual dapat terlihat bahwa peubah X1 memiliki pola sebaran positif yang berarti data pada peubah ini terkumpul pada nilai-nilai yang tinggi. Sebaliknya peubah X3, X4, X5, X7, X8, X9 memiliki pola sebaran negatif karena data pada peubah-peubah ini terkumpul pada nilai-nilai yang rendah.



Gambar 2 Diagram kotak garis (*Boxplot*) peubah numerik (X1-X9)

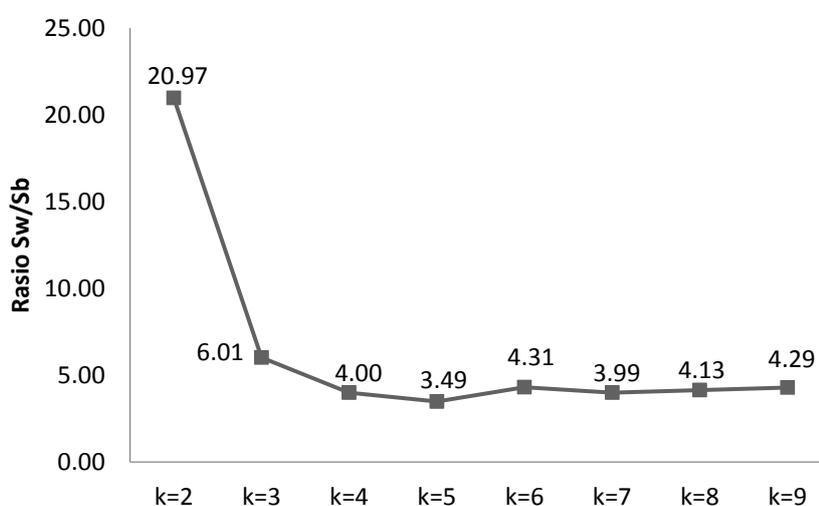
Gambar 2 juga menunjukkan bahwa dari 9 peubah numerik, hanya terdapat 2 peubah (X2 dan X6) yang tidak memiliki pencilan (*outlier*) yaitu nilai yang memiliki perbedaan besar dibandingkan dengan nilai lainnya. Nilai pencilan pada masing-masing peubah disebut dengan *univariate outlier*. Pada peubah X1 terdapat pencilan yang disebut dengan pencilan bawah sedangkan pada peubah X3, X4, X5, X7, X8, dan X9 terdapat pencilan atas.

3.2 Penggerombolan dengan *Two Step Cluster (TSC)*

Penentuan jumlah gerombol optimal pada penelitian ini adalah dengan menggunakan kriteria BIC (*Bayesian Information Criterion*). Langkah pertama dalam tahapan penggerombolan akhir adalah menentukan gerombol maksimal yang dapat terbentuk. Hal ini dengan melihat nilai rasio perubahan BIC yang paling mendekati $c_1=0.04$. Rasio perubahan BIC yang paling mendekati c_1 dan merupakan gerombol maksimum berada pada jumlah gerombol 18. Jumlah gerombol optimal diperoleh berdasarkan ketentuan ditemukannya perbedaan yang nyata pada rasio perubahan gerombol. Dua nilai $R(k)$ terbesar terdapat pada solusi 3 gerombol yaitu $R(k_1)=1.541$ dan 6 gerombol yaitu $R(k_2) = 1.350$. Rasio antara $R(k_1)$ dan $R(k_2)$ adalah sebesar 1.1415, nilai rasio ini lebih kecil dari nilai batas c_2 yaitu sebesar 1.15. Dengan demikian maka jumlah gerombol optimal adalah maksimum $\{k_1, k_2\}$ sehingga k optimal adalah k_2 dengan 6 gerombol.

3.3 Penggerombolan dengan *K-Prototypes*

Langkah pertama dalam mengaplikasikan algoritma ini adalah dengan menentukan jumlah gerombol (k) yang akan dibentuk. Pada penelitian ini, jumlah gerombol yang diolah dengan menggunakan algoritma *K-Prototypes* adalah mulai dari $k=2$ hingga $k=9$. Penentuan jumlah gerombol optimal algoritma *K-Prototypes* dilakukan dengan menghitung rasio keragaman dalam kelompok (S_W) dan keragaman antar kelompok (S_B). Rasio antara S_W dan S_B dengan nilai terkecil merupakan solusi gerombol optimal.



Gambar 3 Rasio S_W/S_B hasil penggerombolan *K-Prototypes*

Berdasarkan hasil perhitungan rasio keragaman maka gerombol optimal yang dihasilkan oleh algoritma *K-Prototypes* adalah pada $k=5$. Nilai rasio S_W/S_B pada solusi gerombol ini merupakan yang terendah daripada gerombol lainnya yaitu sebesar 3.49 (Gambar 3).

3.4 Perbandingan antara *Two Step Cluster (TSC)* dan *K-Prototypes*

Penggerombolan terbaik dari kedua algoritma adalah dengan melihat nilai rasio keragaman dalam kelompok (S_W) dan keragaman antar kelompok (S_B) seperti

ditunjukkan pada Tabel 1. Secara rata-rata untuk keseluruhan peubah, nilai rasio keragaman pada algoritma *Two Step Cluster* (TSC) adalah sebesar 2.6963. Sedangkan nilai rasio keragaman pada algoritma *K-Prototypes* adalah sebesar 3.4852.

Tabel 1 Perbandingan rasio simpangan baku dalam gerombol dan simpangan baku antar gerombol (S_W/S_B) algoritma *Two Step Cluster* (TCS) dan algoritma *K-Prototypes*

Peubah	Tipe Peubah	<i>Two Step Cluster</i> (TSC) dengan $k=6$			<i>K-Prototypes</i> dengan $k=5$		
		S_W	S_B	S_W/S_B	S_W	S_B	S_W/S_B
X1	Numerik	15.3482	10.7124	1.4328	16.4027	13.3662	1.2272
X2	Numerik	30.7049	13.4604	2.2811	31.6930	19.0959	1.6597
X3	Numerik	1.1476	0.6186	1.8552	1.1899	0.9515	1.2506
X4	Numerik	0.8670	0.4435	1.9548	0.7295	0.2781	2.6229
X5	Numerik	3.9097	2.7703	1.4113	4.0509	0.7604	5.3274
X6	Numerik	27.1804	5.6544	4.8069	22.1632	15.5738	1.4231
X7	Numerik	73.0525	41.8922	1.7438	53.5690	21.6270	2.4770
X8	Numerik	12.6461	7.6946	1.6435	9.0002	3.3503	2.6864
X9	Numerik	56.6850	21.7117	2.6108	13.5136	478.0101	0.0283
X10	Kategorik	21.7746	9.5932	2.2698	25.7324	8.2481	3.1198
X11	Kategorik	11.3358	8.4895	1.3353	21.9276	1.4176	15.4682
X12	Kategorik	23.5979	9.1850	2.5692	27.5992	7.3650	3.7474
X13	Kategorik	29.5381	3.2325	9.1378	27.5387	6.4491	4.2701
Rata-Rata		23.6760	10.4199	2.6963	19.6238	44.3456	3.4852

Rasio keragaman pada algoritma TSC menghasilkan nilai yang lebih kecil dibandingkan dengan algoritma *K-Prototypes*. Dengan demikian maka gerombol optimal algoritma *Two Step Cluster* (TSC) merupakan solusi terbaik untuk menggerombolkan desa/kelurahan di Provinsi Nusa Tenggara Timur berdasarkan indikator kemiskinan yang digunakan dalam penelitian ini.

3.5 Karakteristik dan Visualisasi Hasil Penggerombolan Terbaik

Peubah-peubah yang berpengaruh terhadap pembentukan gerombol diperoleh dengan melakukan uji *t-student* (peubah numerik) dan uji *chi-square* (peubah kategorik) pada setiap gerombol. Hasil uji *t-student* dapat diketahui bahwa peubah X1 merupakan peubah yang berpengaruh di seluruh gerombol. Sedangkan hasil uji *chi-square* untuk gerombol 1, 2, 3, 5, 6 diperoleh hasil bahwa semua peubah kategorik dalam penelitian ini memiliki pengaruh dalam pembentukan gerombol masing-masing. Pada gerombol 4, peubah kategorik yang berpengaruh dalam membentuk gerombol ini hanya X12 (Sumber air minum sebagian besar keluarga) dan X11 (Tempat buang air besar sebagian besar keluarga). Secara lengkap hasil uji *t-student* dan uji *chi-square* seperti ditunjukkan pada tabel 2.

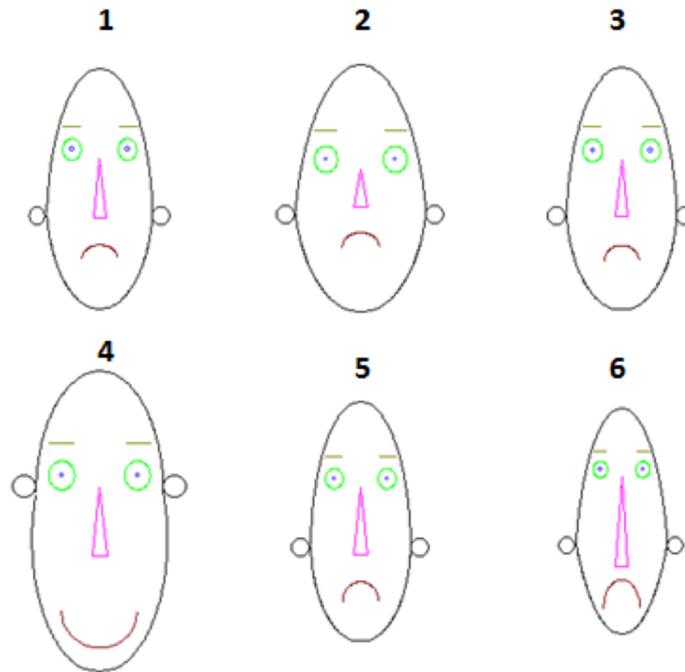
Tabel 2 Peubah-peubah yang berpengaruh dalam pembentukan gerombol terbaik

Gerombol	Peubah Numerik	Peubah Kategorik
1	X5, X7, X1	X12, X13, X11, X10
2	X1, X5, X2, X6, X3, X8,	X11, X13, X12, X10
3	X1, X5, X2	X10, X12, X11, X13
4	X1, X3, X4, X5, X8, X7, X6, X9	X12, X11
5	X1, X7	X10, X12, X11, X13
6	X2, X8, X1, X4, X9, X3, X5, X6, X7	X12, X10, X11, X13

Visualisasi hasil penggerombolan terbaik ($k=6$) dengan *Chernoff Face* seperti ditunjukkan pada Gambar 4, sedangkan deskripsi peubahnya dijelaskan pada Tabel 3. Berdasarkan Gambar 3 dapat terlihat bahwa *Chernoff Face* gerombol 1, gerombol 3, dan gerombol 5 memiliki bentuk wajah yang tidak terlalu jauh berbeda. Gerombol-gerombol ini dicirikan dengan bentuk wajah yang tidak terlalu lebar, posisi telinga yang rendah, bentuk mulut kecil dan melengkung. *Chernoff Face* gerombol 2 dicirikan dengan wajah yang lebar dan bentuk hidung yang kecil. Bentuk hidung yang kecil mempresentasikan peubah X2 yaitu persentase keluarga pengguna listrik.

Tabel 3 Deskripsi peubah sesuai dengan tampilan *Chernoff Face*

Peubah	Ciri Wajah	Keterangan
X1	Wajah bagian bawah (<i>eccentricity of lower face</i>)	Semakin lebar wajah bagian bawah maka semakin tinggi pula persentase keluarga pertanian
X2	Panjang hidung (<i>length of nose</i>)	Semakin panjang hidung maka semakin tinggi pula persentase keluarga pengguna listrik
X3	Posisi mulut (<i>position of center of mouth</i>)	Semakin tengah posisi mulut maka rasio sarana pendidikan semakin baik
X4	Kelengkungan mulut (<i>Curvature of mouth</i>)	Semakin melengkung bentuk mulut maka rasio sarana kesehatan semakin baik
X5	Panjang mulut (<i>Length of mouth</i>)	Semakin panjang mulut maka rasio tenaga kesehatan semakin baik
X6	Lebar wajah (<i>face width</i>)	Semakin lebar wajah maka semakin tinggi persentase penduduk penerima kartu Jamkesmas/Jamkesda.
X7	Posisi telinga (<i>ear level</i>)	Semakin tinggi posisi telinga maka semakin tinggi ratio penerima layanan Surat Keterangan Tidak Mampu (SKTM)
X8	Tinggi dari setengah wajah (<i>half-face height</i>)	Semakin tinggi bentuk dari setengah wajah maka semakin besar jarak kantor desa/kelurahan ke kantor kecamatan
X9	Wajah bagian atas (<i>eccentricity of upper face</i>)	Semakin lebar wajah bagian atas maka semakin besar jarak kantor desa/kelurahan ke kantor bupati/walikota



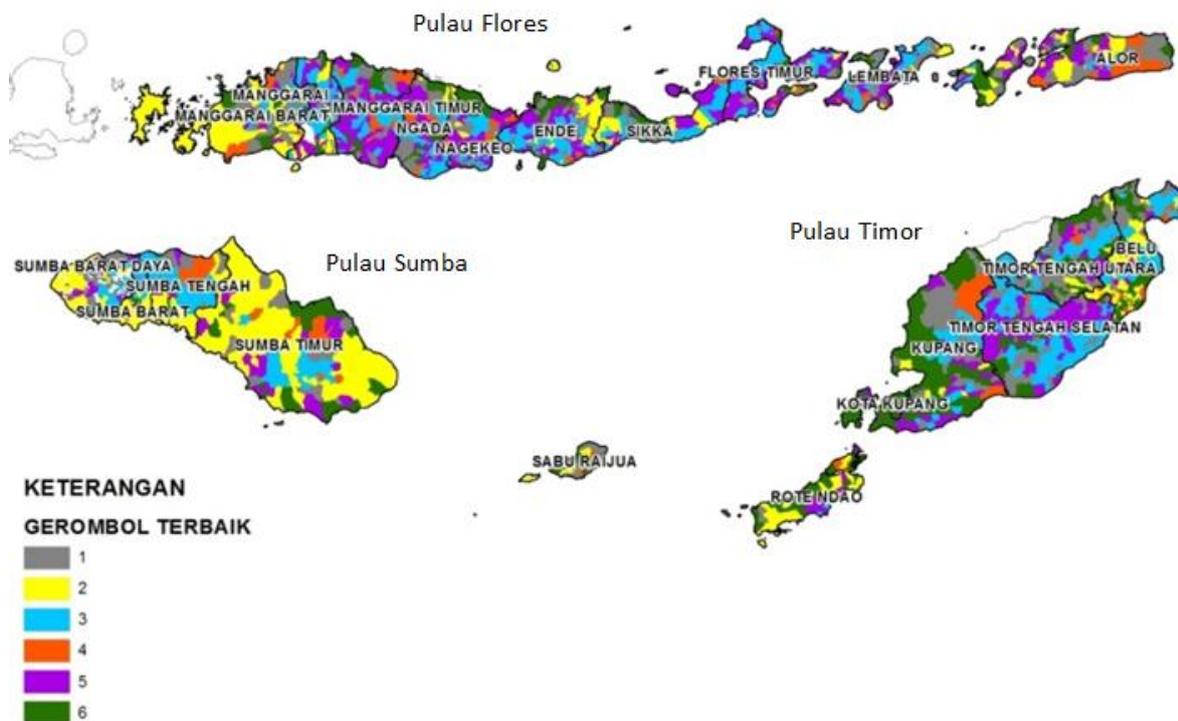
Gambar 4 Tampilan *Chernoff Face* hasil penggerombolan terbaik ($k=6$)

Chernoff Face gerombol 4 dapat terlihat bahwa wajah lebar, posisi telinga tinggi, lengkungan mulut paling lebar dibandingkan dengan gerombol-gerombol lainnya. Artinya rasio sarana pendidikan, rasio sarana kesehatan, dan rasio tenaga kesehatan pada gerombol ini sudah tergolong cukup baik.

Chernoff Face gerombol 6 dicirikan dengan wajah yang paling sempit dibandingkan dengan gerombol lainnya, hidung yang panjang, dan bentuk mulut yang melengkung. Bentuk wajah yang sempit berarti bahwa X_1 (persentase keluarga pertanian) pada gerombol ini adalah yang terendah. Bentuk hidung yang panjang menggambarkan peubah X_2 yaitu persentase keluarga pengguna listrik pada gerombol ini sangat tinggi.

Provinsi Nusa Tenggara Timur merupakan wilayah kepulauan dengan tiga pulau utamanya yaitu Pulau Flores, Pulau Sumba, dan Pulau Timor. Hasil penggerombolan terbaik bila divisualisasikan dalam bentuk peta tematik seperti ditunjukkan pada Gambar 5.

Wilayah desa/kelurahan yang masuk dalam gerombol 1 (warna abu-abu) menyebar merata di sebagian kecil Pulau Sumba, Pulau Flores, maupun juga Pulau Timor. Wilayah desa/kelurahan yang masuk dalam gerombol 2 (warna kuning) lebih dominan berada di wilayah Pulau Sumba. Sedangkan wilayah gerombol 3 (warna biru) dan gerombol 5 (warna ungu) dominan tersebar di Pulau Timor dan Pulau Flores. Gerombol 4 (warna coklat) merupakan wilayah gerombol yang terkecil dibandingkan dengan gerombol lain. Secara visual, gerombol ini terlihat dominan di Kabupaten Alor serta Kabupaten Kupang. Terakhir, gerombol 6 (warna hijau) terlihat sangat dominan di wilayah Kabupaten Kupang dan Kota Kupang.



Gambar 5 Peta tematik hasil penggerombolan terbaik ($k=6$)

4. Simpulan

Hasil perbandingan secara menyeluruh antara dua algoritma didapatkan bahwa nilai rasio keragaman algoritma *Two Step Cluster* (TSC) lebih kecil dibandingkan dengan algoritma *K-Prototypes*. Pada kasus ini maka hasil penggerombolan algoritma TSC dengan rasio keragaman sebesar 2.6963 merupakan solusi terbaik dalam menggerombolkan desa/kelurahan di Provinsi Nusa Tenggara Timur berdasarkan indikator kemiskinan. Gerombol terbaik yang dihasilkan adalah sebanyak 6 gerombol. Akan tetapi hasil gerombol terbaik berdasarkan algoritma TSC ini ternyata tidak menghasilkan karakteristik yang berbeda antara setiap gerombol yang terbentuk. Karakteristik gerombol 1, Gerombol 3, dan Gerombol 5 secara visual berdasarkan *Chernoff Face* tidak berbeda.

Daftar Pustaka

- Aminah A.S. (2017). Penerapan Metode Spasial Emperical Best Linear Unbiased Prediction Dengan Prosedur Restricted Maximum Likelihood dan Bootstrap Untuk Estimasi Persentase Penduduk Miskin Tingkat Kecamatan di Kabupaten Wonosobo [tesis]. Bandung (ID): Universitas Padjajaran.
- Bacher, J., Wenzig, K., & Vogler, M. (2004). SPSS TwoStep Cluster-A First Evaluation. *Lehrstuhl fur Soziologie Arbeits- und Diskussionspapiere* 2:p.1–20

- [BPS] Badan Pusat Statistik Provinsi Nusa Tenggara Timur. (2014). *Statistik Potensi Desa Provinsi Nusa Tenggara Timur Tahun 2014*. Kupang (ID): Badan Pusat Statistik Provinsi Nusa Tenggara Timur.
- [BPS] Badan Pusat Statistik. (2016). *Perhitungan dan Analisis Kemiskinan Makro Indonesia 2016*. Jakarta (ID): BPS.
- [BPS] Badan Pusat Statistik. (2018). Profil Kemiskinan di Indonesia September 2017 No. 05/01/Th.XXI, 2 Januari 2018. Jakarta (ID): BPS.
- Bunkers, M. J., Miller Jr, J. R., & DeGaetano, A. T. (1996). Definition of climate regions in the Northern Plains using an objective cluster modification technique. *Journal of Climate*, **9**(1), 130-146.
- Chaudhary, K., Sharma, A. (2014). Implementation of Two Steps Clustering Using Telecommunication System. *International Journal of Information Technology*, **7**(2),42–48.
- Chernoff, H. (1973). The use of faces to represent points in k-dimensional space graphically. *Journal of the American statistical Association*, **68**(342), 361-368.
- Green, R. H. (1978). Basic human needs: concept or slogan, synthesis or smokescreen?. *The IDS Bulletin*, **9**(4), 7-11.
- Huang, Z. (1997). Clustering large data sets with mixed numeric and categorical values. In Proceedings of the 1st pacific-asia conference on knowledge discovery and data mining,(PAKDD) (pp. 21-34).
- Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data mining and knowledge discovery*, **2**(3), 283-304.
- [Kemendes PDTT] Kementerian Desa, Pembangunan Daerah Tertinggal, dan Transmigrasi. (2015). *Indeks Desa Membangun 2015*. Jakarta (ID): Kemendes PDTT.
- Sulistiyono, D. (2016). Bootstrap Spasial Empirical Best Linear Unbiased Prediction Untuk Pemetaan Kemiskinan Tingkat Desa di Kabupaten Pati [tesis]. *Surabaya (ID): Institut Teknologi Sepuluh November*.