

Received: 7 Maret 2019

Revised: 25 Maret 2019

Accepted: 13 April 2019

Published: 28 Juni 2019

# PENGLASIFIKASIAN PENGGUNA INTERNET LINGKUNGAN PEDESAAN MENURUT JENJANG PENDIDIKAN DI INDONESIA MENGGUNAKAN METODE *CLUSTER AVERAGE LINKAGE*

Hemelia<sup>1,a)</sup>, Bagus Sumargo<sup>2,b)</sup><sup>1</sup>Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta<sup>2</sup>Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri JakartaEmail: a)[hemelia.memey@yahoo.com](mailto:hemelia.memey@yahoo.com), b)[bagusumargo63@gmail.com](mailto:bagusumargo63@gmail.com)

## Abstract

Internet users are increasing every year, therefore cluster analysis is carried out in classifying internet users in each province. This is done to determine the characteristics of each cluster. In this journal, the cluster method used is the Average Linkage method. The purpose of this journal is to determine the clustering process with the Average Linkage method and find out the results of the cluster that is formed using the method. In this method, data on rural internet users according to the level of education in Indonesia are similar in their similarity to be cluster. The similarity was calculated using Euclidean distance, then formed a cluster, and calculated standard deviation between groups and standard deviation in groups to find out the cluster method that has good performance.

**Keywords:** Cluster, Average Linkage, Internet User.

## Abstrak

Pengguna internet semakin meningkat setiap tahunnya oleh karena itu, dilakukan analisis cluster dalam pengklasifikasian pengguna internet setiap provinsi. Hal tersebut dilakukan untuk mengetahui karakteristik setiap kelompok yang terbentuk. Pada jurnal ini, metode *cluster* yang digunakan adalah metode *Average Linkage*. Tujuan dari jurnal ini adalah untuk mengetahui proses *clustering* dengan metode *Average Linkage* dan mengetahui hasil *cluster* yang terbentuk dengan menggunakan metode tersebut. Dalam metode ini, data pengguna internet daerah pedesaan menurut jenjang pendidikan di Indonesia dihitung kemiripannya untuk dilakukan pengelompokan. Kemiripan tersebut dihitung menggunakan jarak *Euclidean*, kemudian terbentuk *cluster*, dan dilakukan perhitungan simpangan baku antar kelompok dan simpangan baku dalam kelompok untuk mengetahui metode *cluster* yang memiliki kinerja yang baik.

**Kata-kata kunci:** *Cluster, Average Linkage, Pengguna Internet.*

## PENDAHULUAN

Pengguna internet di Indonesia mengalami peningkatan setiap tahunnya. Peningkatan tersebut bertambah secara drastis dari tahun ke tahun. Berdasarkan hal tersebut maka dapat dilakukan pengelompokan berdasarkan kesamaan karakteristik. Pengelompokan ini bertujuan agar informasi terkait pengguna internet tersebar tiap provinsi menjadi lebih spesifik sehingga dapat terlihat provinsi mana yang memiliki pengguna internet terbanyak. Oleh sebab itu dilakukan analisis untuk mengetahui pengguna internet di provinsi mana yang banyak memanfaatkan internet dan pada provinsi mana saja yang perlu dikembangkan akses pengguna internet di Indonesia. Salah satu cara untuk melakukan pengelompokan tersebut adalah dengan cara identifikasi dengan analisis *cluster*.

Terdapat dua metode pengklasteran dalam analisis *cluster*, yaitu metode hierarki dan metode non hierarki. Dalam metode hierarki terdapat beberapa macam, diantaranya yaitu, metode *Single Linkage*, metode *Average Linkage*, metode *Complete Linkage*, metode *Ward*, dan metode *Centroid*. Mengacu pada hal ini, maka dilakukan analisis *cluster* metode *Average Linkage* untuk menganalisis data persentase pengguna internet daerah pedesaan menurut jenjang pendidikan di Indonesia pada tahun 2017. Sehingga dapat dilakukan analisis provinsi mana saja yang banyak memanfaatkan internet dan provinsi mana yang perlu dikembangkan akses internetnya.

## KAJIAN TEORI

### Analisis Multivariat

Analisis multivariat adalah analisis statistik yang digunakan untuk menganalisis data yang terdiri dari beberapa variabel dan variabel-variabel tersebut saling berkorelasi satu sama lain. Terdapat dua jenis teknik dalam analisis multivariat, yaitu teknik dependen dan teknik interdependen.

Teknik dependen yaitu jika variabel dependen dipengaruhi oleh variabel independen, atau bisa diartikan adanya satu atau beberapa variabel yang berfungsi sebagai variabel terikat dan variabel bebas. Menurut Supranto (2004:20), metode-metode yang termasuk dalam analisis dependensi diantaranya analisis Regresi Linear Berganda, analisis Diskriminan Berganda, analisis Multivariat Varian, dan analisis Korelasi Kanonik.

Teknik interdependen yaitu jika semua variabel saling berpengaruh. Dengan kata lain, dalam teknik interdependen semua variabel adalah independen. Menurut Supranto (2004:20), metode-metode yang termasuk dalam teknik ini adalah analisis Faktor, analisis *Cluster*, *Multi Dimensional Scaling Analysis* (MDS) dan analisis *Categorical*.

### Analisis Cluster

Analisis *cluster* merupakan suatu teknik multivariat yang bertujuan mengelompokkan data observasi ataupun objek-objek berdasarkan karakteristik yang dimilikinya. Analisis *cluster* mengklasifikasi objek sehingga setiap objek yang memiliki sifat yang mirip (paling dekat kesamaannya) akan mengelompok ke dalam satu *cluster* (kelompok) yang sama. Masing-masing *cluster* bersifat homogen sesuai dengan faktor yang digunakan untuk melakukan pengklasteran.

*Cluster* membangun kelompok dengan sifat homogen dari sampel besar heterogen. Kelompok-kelompok tersebut harus sehomogen mungkin dan perbedaan di antara berbagai kelompok tersebut harus seluas mungkin. Proses *clustering* pada dasarnya mengelompokkan objek yang mirip satu dengan yang lain, maka kriteria mirip (*similarity*) adalah dasar dari metode *clustering*. Semakin besar tingkat kemiripan (atau homogenitas) di dalam satu grup dan semakin besar tingkat perbedaan di antara grup, maka semakin baik *cluster* tersebut.

Tahap-tahap dalam pengelompokan objek menjadi satu atau beberapa *cluster* adalah sebagai berikut.

1. Melakukan Proses Standarisasi Data

Hal pertama yang perlu dilakukan sebelum proses *clustering* adalah melakukan standardisasi dengan mengubah data yang akan digunakan menjadi data *z-score*. Proses standardisasi akan menjadikan data-data dengan perbedaan satuan yang lebar akan otomatis menjadi menyempit. Standardisasi data dengan nilai *z* dapat menggunakan rumus berikut.

$$z = \frac{x - \bar{X}}{\sigma} \quad (1)$$

dengan *z* merupakan nilai data yang telah distandardisasi, *x* merupakan nilai data asli,  $\bar{X}$  merupakan nilai rata-rata keseluruhan dan  $\sigma$  merupakan simpangan baku.

## 2. Menentukan Ukuran Kemiripan

Tujuan analisis *cluster* adalah mengklasifikasi objek sehingga setiap objek yang mirip akan mengelompok ke dalam *cluster* yang sama. Pada analisis *cluster* terdapat beberapa ukuran kemiripan diantaranya adalah ukuran asosiasi, ukuran korelasi, dan ukuran jarak.

Ukuran kemiripan dengan mengitung jarak antar dua objek. Dalam mengukur kedekatan antara dua objek, dapat digunakan metode pengukuran *Euclidean Distance* yang merupakan ukuran jarak atau ketidaksamaan antar objek ke-*i* dengan objek ke-*j*, disimbolkan dengan  $d_{ij}$  dan variabel ke-*k* = 1,2,..., *p*, nilai  $d_{ij}$  diperoleh perhitungan jarak *Euclidean* sebagai berikut.

$$d_{ij} = \sqrt{\sum_{k=1}^n (X_{ik} - X_{jk})^2} \quad (2)$$

dengan  $d_{ij}$  merupakan jarak antara objek ke-*i* dan objek ke-*j*, *n* merupakan jumlah variabel yang diamati,  $X_{ik}$  merupakan data dari subjek ke-*i* pada variabel ke-*k* dan  $X_{jk}$  merupakan data dari subjek ke-*j* pada variabel ke-*k*.

## 3. Memilih Prosedur Pengklasteran

Proses pembentukan *cluster* dapat dilakukan dengan dua cara, yaitu dengan metode Hierarki dan Non Hierarki. Pada metode Hierarki terdiri dari metode *Agglomerative* (Penggabungan) dan metode *Devisive* (Pemecahan). Sedangkan metode Non Hierarki terdiri yaitu metode K-Means.

## 4. Menentukan Banyaknya Cluster

Masalah utama dalam analisis *cluster* adalah menentukan berapa banyaknya *cluster*. Sebetulnya tidak ada aturan yang baku untuk menentukan berapa banyaknya *cluster*, namun demikian ada beberapa petunjuk yang bisa dipergunakan, yaitu (Supranto, 2004).

- a) Pertimbangan teoretis, konseptual, praktis, mungkin bisa diusulkan/disarankan untuk menentukan berapa banyaknya *cluster* yang sebenarnya.
- b) Besarnya relatif *cluster* seharusnya berguna/bermanfaat.

Metode *clustering* dikatakan memiliki kinerja yang baik dapat diukur dengan menghitung simpangan baku dalam kelompok ( $S_w$ ) dan simpangan baku antar kelompok ( $S_b$ ) (Bunkers, dkk.1996).

Nilai  $S_w$  dapat dicari dengan menggunakan rumus.

$$S_w = \frac{1}{K} \sum_{k=1}^n S_k \quad (3)$$

dengan *K* merupakan banyaknya kelompok yang terbentuk dan  $S_k$  merupakan simpangan baku kelompok ke-*k*.

$$S_k = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X}_k)^2} \quad (4)$$

dengan  $x_i$  merupakan rata-rata objek pada setiap variabel, dan  $\bar{X}_k$  merupakan rata-rata seluruh kelompok.

Nilai  $S_b$  dapat dicari dengan menggunakan rumus.

$$S_b = \sqrt{\frac{1}{K-1} \sum_{k=1}^K (\bar{X}_k - \bar{X})^2} \quad (5)$$

dengan  $K$  merupakan banyaknya kelompok yang terbentuk,  $\bar{X}_k$  merupakan rata-rata kelompok ke- $k$ , dan  $\bar{X}$  merupakan rata-rata seluruh kelompok.

Semakin kecil nilai simpangan baku dalam kelompok ( $S_w$ ) dan semakin besar nilai simpangan baku antar kelompok ( $S_b$ ) maka metode tersebut memiliki kinerja yang baik.

### 5. Mengintrepretasikan Hasil *Cluster*

Pada tahap terakhir adalah interpretasi pada masing-masing *cluster* yang terbentuk. Ketika memulai proses interpretasi digunakan rata-rata (*centroid*) setiap *cluster* pada setiap variabel.

#### Metode *Cluster* Hierarki

Metode *cluster* hierarki merupakan suatu metode pada analisis *cluster* yang membentuk tingkatan tertentu seperti pada struktur pohon dikarenakan proses pengklasterannya dilakukan secara bertingkat atau bertahap. Pada metode *cluster* hierarki terdiri dari metode *Agglomerative* (Penggabungan) dan metode *Devisive* (Pemecahan). Metode *Agglomeratif* sendiri terdiri dari beberapa metode yaitu metode *Single Linkage*, *Average Linkage*, *Complete Linkage*, *Ward*, dan *Centroid*.

Hasil pengklasteran dengan metode *cluster* hierarki dapat disajikan dalam bentuk *dendogram*. *Dendogram* adalah representasi visual dari langkah-langkah dalam analisis *cluster* yang menunjukkan bagaimana *cluster* terbentuk dan nilai koefisien jarak pada setiap langkah. Angka disebelah kanan adalah objek penelitian, dimana objek-objek tersebut dihubungkan oleh garis dengan objek yang lain sehingga pada akhirnya akan membentuk suatu *cluster*.

#### Tahap-Tahap Pembentukan *Cluster* Hierarki *Agglomeratif*

Berikut tahap-tahap pembentukan *cluster* dengan menggunakan metode hierarki adalah sebagai berikut.

1. Dimulai dengan  $N$  kelompok, masing-masing mengandung kesatuan yang tunggal dan matriks simetris  $N \times N$  dari jarak (kemiripan) yaitu,  $D = \{d_{ij}\}$ .
2. Dicari matriks jarak untuk pasangan kelompok terdekat (yang paling banyak kesamaan). Dimisalkan jarak antara kelompok  $U$  dan  $V$  yang paling sama dinotasikan dengan  $d_{uv}$ .
3. Gabungkan kelompok  $U$  dan  $V$ . Gabungan tersebut dinotasikan dengan  $(UV)$ . Letakkan objek pada matriks jarak dengan.
  - a. Menghapus baris dan kolom yang berkorespondensi dengan kelompok  $U$  dan  $V$
  - b. Menambahkan baris dan kolom yang terdapat jarak antara kelompok  $(UV)$  dan kelompok yang tertinggal.
4. Ulangi langkah 2 dan 3 sebanyak  $(N - 1)$  kali, dengan  $N$  adalah jumlah objek atau responden. Sehingga semua objek akan berada pada satu *cluster* saat algoritma terakhir.

Proses pengklasteran dengan metode *Average Linkage* adalah dengan menghitung jarak rata-rata objek yang bergabung menjadi satu *cluster* dengan objek yang lain. Penggabungan berikutnya terjadi

pada *cluster-cluster* yang paling mirip, sehingga membentuk *cluster* yang kedua. Kemudian dihitung jarak rata-rata dengan menggunakan rumus.

$$d_{(UV)W} = \frac{\sum_i \sum_j d_{ij}}{N_{UV} N_W} \quad (6)$$

dengan  $d_{ij}$  merupakan jarak antara objek  $i$  dalam *cluster* (UV) dan objek  $k$  dalam *cluster* W. Sedangkan  $N(UV)$  dan  $N(W)$  berturut-turut merupakan jumlah objek dalam *cluster* (UV) dan (W).

### HASIL DAN PEMBAHASAN

Dalam pembahasan ini akan mengklasterkan data persentase pengguna internet yang mengakses internet menurut jenjang pendidikan dari mulai SD, SMP, SMA dan Perguruan Tinggi di pedesaan pada tahun 2017. Data yang akan digunakan adalah 33 provinsi di Indonesia, 1 provinsi yaitu DKI Jakarta tidak termasuk ke dalam data persentase pengguna internet daerah pedesaan.

#### Standarisasi Data

Standarisasi data dilakukan sebelum melakukan proses *clustering*. Standarisasi dilakukan dengan transformasi variabel ke bentuk *z-score*. Diperoleh nilai rata-rata dan simpangan baku setiap variabel sebagai berikut:  $\bar{X}_{SD} = 4,8042$ ,  $\bar{X}_{SMP} = 36,0964$ ,  $\bar{X}_{SMA} = 65,9785$ ,  $\bar{X}_{PT} = 73,4952$ ,  $\sigma_{SD} = 2,86714$ ,  $\sigma_{SMP} = 15,77563$ ,  $\sigma_{SMA} = 16,45381$  dan  $\sigma_{PT} = 11,78047$ . Maka, dihitung nilai standarisasi masing-masing provinsi dengan menggunakan persamaan (1) sebagai berikut.

Untuk provinsi Aceh.

$$Z_{SD_{ACEH}} = \frac{1,58 - 4,8042}{2,86714} \approx -1,12455$$

$$Z_{SMP_{ACEH}} = \frac{18,45 - 36,0964}{15,77563} \approx -1,11858$$

$$Z_{SMA_{ACEH}} = \frac{43,91 - 65,9785}{16,54381} \approx -1,34124$$

$$Z_{PT_{ACEH}} = \frac{72,41 - 73,4952}{11,78047} \approx -0,09211$$

dan seterusnya sehingga semua data pengguna internet menurut jenjang pendidikan per provinsi tahun 2017 dilakukan standarisasi (*z-score*).

#### Menghitung Jarak Euclidean

Jarak *Euclidean* merupakan jarak antara objek ke- $i$  dengan ke- $j$ , yang akan diukur kemiripannya. Semakin kecil jarak *Euclidean* antar dua objek maka semakin mirip dua objek tersebut. Dengan menggunakan persamaan (2), maka diperoleh perhitungan jarak *Euclidean* untuk provinsi Aceh dan Sumatera Utara sebagai berikut.

$$d_{ij} = \sqrt{\sum_{k=1}^n (X_{ik} - X_{jk})^2}$$

$$= \sqrt{2,933951539}$$

$$= 1,712878145 \approx 1,713$$

Proses perhitungan ini dilakukan untuk menghitung jarak *Euclidean* dari masing-masing provinsi di Indonesia.

Tabel matriks *Proximity*, dengan angka yang tertera adalah jarak antara dua buah variabel. Sebagai contoh, jarak provinsi Aceh dan provinsi Sumatera Utara sebesar 1,713. Sedangkan jarak antara provinsi Aceh dan provinsi Sumatera Barat sebesar 2,653. Hal ini berarti karakteristik provinsi Aceh lebih mirip provinsi Sumatera Utara dibandingkan provinsi Sumatera Barat. Dengan demikian seterusnya untuk penafsiran data yang lain, dengan acuan semakin kecil jarak kedua objek, maka semakin mirip kedua data tersebut.

### Pembentukan Cluster Metode Average Linkage

Dalam melakukan pengelompokan data dimulai dengan mencari nilai rata-rata dari tiap objek. Setelah didapat nilai rata-rata setiap objek, maka dapat dilakukan proses *clustering* dengan menggunakan metode *Average Linkage*.

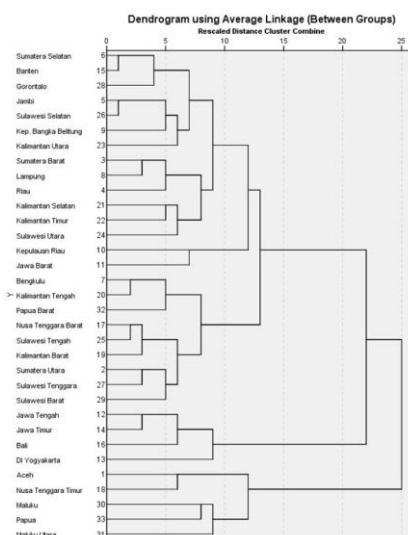
Cara hirarki berarti pengelompokan yang dimulai dengan menganggap semua objek mempunyai *clusternya* masing-masing. Kemudian, pengelompokan dimulai secara bertingkat, satu demi satu, dimulai dari 2 objek yang memiliki jarak yang paling dekat dijadikan satu *cluster*. Selanjutnya *cluster* tersebut dicari kembali objek yang mendekati jarak *cluster* tersebut. Proses tersebut dilakukan berulang kali hingga akhirnya mendapatkan satu *cluster* tunggal. Proses ini dinamakan Agglomerasi.

Proses pengklasteran metode *Average Linkage* adalah dengan menghitung jarak rata-rata antara objek yang bergabung dengan objek yang lain. Maka untuk menghitung jarak rata-rata yang terjadi antara objek terakhir yang bergabung yakni contoh provinsi Gorontalo dan dua objek sebelumnya, yakni provinsi Sumatera Selatan dan provinsi Banten. Dengan melihat jarak pada output *Proximity Matrix* maka dapat dihitung jarak rata-rata dengan persamaan (6) sebagai berikut.

- Jarak antara objek 28 (Gorontalo) dan objek 6 (Sumatera Selatan) adalah 0,621.
- Jarak antara variabel 28 (Gorontalo) dan variabel 15 (Banten) adalah 0,598.

$$\text{Jarak rata - rata} = \frac{(0,621 + 0,598)}{2} = 0,608$$

Proses pengklasteran di atas juga dapat diilustrasikan dalam bentuk *dendrogram*. *Dendrogram* dibaca dari kiri ke kanan dimana garis vertikal menunjukkan *cluster* yang digabung bersama, sedangkan garis pada skala menunjukkan jarak *cluster* yang digabungkan. Dapat dilihat *cluster* yang terbentuk melalui *dendrogram* berikut ini.



GAMBAR 1. Dendrogram Hasil Clustering dengan Metode Average Linkage

### Menentukan Jumlah Cluster

Bedasarkan tabel *Cluster Membership*, maka dapat ditentukan metode *clustering* dengan kinerja yang baik. Metode *clustering* dikatakan memiliki kinerja yang baik dapat diukur dengan menghitung simpangan baku dalam kelompok ( $S_w$ ) dan simpangan baku antar kelompok ( $S_b$ ). Dengan menggunakan persamaan (3) dan (4) maka diperoleh hasil perhitungan sebagai berikut.

- 1) Untuk pembentukan dua *cluster* diperoleh nilai  $S_w = 0,498221322$  dan nilai  $S_b = 1,227942039$ ,  
 $\frac{S_w}{S_b}$   
 maka rasio  $\frac{S_w}{S_b}$  yang diperoleh yaitu 0,40573684.
- 2) Untuk pembentukan tiga *cluster* diperoleh nilai  $S_w = 0,383835403$  dan nilai  $S_b = 1,49448773$ ,  
 $\frac{S_w}{S_b}$   
 maka rasio  $\frac{S_w}{S_b}$  yang diperoleh yaitu 0,256834094.
- 3) Untuk pembentukan empat *cluster* diperoleh nilai  $S_w = 0,284165844$  dan nilai  $S_b = 1,267671813$ ,  
 $\frac{S_w}{S_b}$   
 maka rasio  $\frac{S_w}{S_b}$  yang diperoleh yaitu 0,224163573.

Dengan memperhatikan hasil perhitungan di atas, maka metode *Average Linkage* memiliki kinerja yang baik yaitu dengan pembentukan 4 *cluster*. Karena semakin kecil nilai simpangan baku dalam kelompok dan semakin besar nilai simpangan baku antar kelompok tersebut maka dikatakan memiliki kinerja yang baik.

### Mengintrepretasikan Hasil Cluster

Dengan menggunakan metode *Average Linkage* maka diperoleh klasifikasi pengguna internet yang mengakses internet menurut jenjang pendidikan di pedesaan tiap provinsi pada tahun 2017 sebanyak 4 *cluster*. Diketahui bahwa pada *cluster* 1 terdiri dari 5 objek, *cluster* 2 terdiri dari 9 objek, *cluster* 3 terdiri dari 15 objek, dan *cluster* 4 terdiri dari 4 objek. Berikut ini merupakan rincian anggota tiap *cluster* dengan karakteristiknya.

1. Anggota *cluster* 1 adalah provinsi Aceh, Nusa Tenggara Timur, Maluku, Maluku Utara dan Papua.  
*Cluster* 1 memiliki karakteristik pengguna internet yang sangat rendah, dengan rata-rata pengguna internet sebesar 26,98%.
2. Anggota *cluster* 2 adalah provinsi Sumatera Utara, Bengkulu, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Tengah, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Barat dan Papua Barat.  
*Cluster* 2 memiliki karakteristik pengguna internet yang cukup rendah, dengan rata-rata pengguna internet sebesar 39,73%.
3. Anggota *cluster* 3 adalah provinsi Sumatera Barat, Riau Jambi, Sumatera Selatan, Lampung, Kep. Bangka Belitung, Kepulauan Riau, Jawa Barat, Banten, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Selatan dan Gorontalo.  
*Cluster* 3 memiliki karakteristik pengguna internet yang cukup tinggi, dengan rata-rata pengguna internet sebesar 49,91%.
4. Anggota *cluster* 4 adalah provinsi Jawa Tengah, DI Yogyakarta, Jawa Timur dan Bali.  
*Cluster* 4 memiliki karakteristik pengguna internet yang sangat tinggi, dengan rata-rata pengguna internet sebesar 61,73%.

## KESIMPULAN DAN SARAN

### Kesimpulan

Dengan metode *Average Linkage* maka diperoleh hasil *cluster* yang terbentuk yaitu 4 *cluster* dengan *cluster* 1 terdiri dari 5 provinsi memiliki karakteristik pengguna internet yang sangat rendah, dengan rata-rata pengguna internet sebesar 26,98%. *Cluster* 2 terdiri dari 9 provinsi cukup rendah

dengan rata-rata pengguna internet sebesar 39,73%. Untuk *cluster* 3 terdiri dari 15 provinsi cukup tinggi dengan rata-rata pengguna internet sebesar 49,91%. Sedangkan, *cluster* 4 terdiri dari 4 provinsi memiliki karakteristik sangat tinggi, dengan rata-rata pengguna internet sebesar 61,73%. Dari *cluster* yang terbentuk tersebut diperoleh rasio antara simpangan baku dalam *cluster* dan simpangan baku antar *cluster* sebesar 0,224163573. Hal ini berarti *cluster* yang terbentuk dikatakan memiliki kinerja yang baik.

#### Saran

Sebagai *open problem* dari hasil penelitian ini adalah dapat menggunakan metode lain untuk melakukan proses *clustering*. Diharapkan pembaca juga dapat membandingkan hasil masing-masing metode hierarki untuk menentukan metode yang lebih baik. Selain itu, pembaca juga dapat melakukan uji multikolinearitas dan uji *outlier* pada data yang akan dianalisis.

#### REFERENSI

- Bunkers, M.J., Miller Jr, J.R. and DeGaetano, A.T. (1996) 'Definition of climate regions in the Northern Plains using an objective cluster modification technique', *Journal of Climate*, 9(1), pp.130-146.
- Haryono, S. (2017) Analisis Multivariat. Yogyakarta, Universitas Muhammadiyah Yogyakarta.
- Johnson, R.A. and Wichern, D.W. (2007) *Applied Multivariate Statistical Analysis sixth edition*. United States of America: Pearson Education, Inc.
- Laeli, Sofyah (2014) Analisis Cluster dengan Average Linkage Method dan Wards Method untuk Data Responden Nasabah Asuransi Jiwa Unit Link. Yogyakarta, Universitas Negeri Yogyakarta.
- Laraswati, T.F. (2014) *Perbandingan Kinerja Metode Complete, Linkage Metode Average Linkage, dan Metode K-means Dalam Menentukan Hasil Analisis Cluster*, Yogyakarta: Universitas Negeri Yogyakarta.
- Supranto, J. (2004) *Analisis Multivariat Arti dan Interpretasi*, Jakarta: Rineka Cipta.
- Ulwan, M Nashihun (2014) 'Cara Analisis Cluster Metode Hirarkis dengan SPSS' [online]. Available at: <http://www.portal-statistik.com/2014/02/cara-analisis-cluster-metode-hirarkis.htm> (Accessed: 20 Oktober 2017).