



\*Corresponding author: Muhammad Nusran, Department of Statistics, Universitas Negeri Makassar, Makassar, Indonesia

E-mail: [muh.nusrang@unm.ac.id](mailto:muh.nusrang@unm.ac.id)

## RESEARCH ARTICLE

# K-Means Cluster Analysis for Grouping Districts in South Sulawesi Province Based on Village Potential

Azrahwati, Muhammad Nusrang\*, Muhammad Kasim Aidid, & Zulkifli Rais

Department of Statistics, Universitas Negeri Makassar, Makassar, Indonesia

**Abstract:** Cluster analysis is an analysis in multivariable statistics that is used to group objects that have the same characteristics. One of the methods in cluster analysis used to group relatively large amounts of data is the K-Means method. In this study, the K-Means method was applied to classify sub-districts in South Sulawesi Province based on village potential. The variables used are the number of: Elementary School/Equivalent degree, Junior High School/Equivalent degree, Senior High School/Vocational School/Equivalent degree, Community Health Center/Pustu, Families without electricity, Villages/Urbans according to market presence, Villages/Towns that are passed by public transportation and Villages/Kelurahan that have lighting main road. The results of this study are that 3 groups are formed where the first cluster consists of 107 sub-districts that have high village potential, the second cluster consists of 16 sub-districts that have medium village potential and the third cluster consists of 184 sub-districts that have low village potential.

**Keywords:** Cluster Analysis, K-Means, Village Potential

## 1. Introduction

Indonesia merupakan suatu negara kepulauan yang terdiri dari 5 pulau besar salah satunya yaitu pulau Sulawesi. Provinsi Sulawesi Selatan merupakan salah satu dari 6 provinsi yang ada di pulau tersebut. Berdasarkan posisi geografisnya, provinsi ini memiliki batas wilayah yaitu utara berbatasan dengan Sulawesi Tengah dan Sulawesi Barat; timur berbatasan dengan Teluk Bone dan Sulawesi Tenggara; selatan berbatasan dengan Laut Flores; dan barat berbatasan dengan Selat Makassar sedangkan secara administratif, Provinsi Sulawesi Selatan terdiri atas 21 Kabupaten dan 3 Kota serta 307 Kecamatan dan 3049 wilayah administrasi setingkat Desa (BPS, 2019).

Menurut Fathia, et.al. (2016), Setiap wilayah memiliki karakteristik tertentu dari segi fisik maupun non fisik. Kedua hal ini dapat digunakan untuk melihat keragaman wilayah administrasi pemerintahan satu diantaranya yaitu wilayah kecamatan. Berdasarkan buku pedoman pencacah desa/kelurahan Podes 2018, data hasil potensi desa hingga saat ini merupakan satu-satunya sumber data tematik berbasis wilayah yang mampu menggambarkan potensi yang dimiliki setiap wilayah administrasi pemerintahan. Adapun potensi wilayah tersebut meliputi: sosial, ekonomi, serta sarana dan prasarana wilayah. Pendataan potensi desa dilaksanakan oleh Badan Pusat Statistik Provinsi Sulawesi Selatan setiap 3 kali dalam 10 tahun. Akan tetapi, dari hasil pendataan ini tidak dijelaskan mengenai wilayah yang memiliki potensi desa yang tinggi maupun rendah. Oleh karena itu, untuk dapat mengetahui keadaan potensi desa di setiap wilayah kecamatan dapat digunakan



analisis dalam statistika yaitu analisis cluster atau yang dikenal dengan analisis pengelompokan.

Analisis cluster merupakan salah satu analisis dalam statistika peubah banyak yang digunakan dalam pengelompokan objek yang memiliki kesamaan sehingga anggota dalam satu cluster bersifat homogen dan antar cluster itu sendiri bersifat heterogen (Johnson & Wichern, 2002). Adapun metode dalam analisis cluster terbagi atas dua yaitu metode hierarki dan metode non-hierarki. Salah satu metode non-hierarki dalam analisis cluster yang sering digunakan yaitu metode K-Means. Metode K-Means merupakan suatu metode dalam analisis cluster yang berusaha membagi data yang ada ke dalam bentuk satu atau lebih cluster (Bastian, et.al., 2018).

Data yang digunakan dalam penelitian ini terdiri dari 307 kecamatan di Provinsi Sulawesi Selatan yang datanya relatif banyak sehingga metode dalam analisis cluster yang dapat digunakan yaitu metode K-Means. Hal ini sesuai dengan hasil penelitian yang dilakukan oleh Dewi Rachmatin dan Kania Sawitri pada tahun 2019 yaitu mengenai perbandingan antara metode agglomerative, metode divisif, dan metode K-Means dalam analisis cluster yang menghasilkan kesimpulan bahwa metode hierarki tidak efektif untuk mengelompokkan objek yang memiliki ukuran sampel yang besar sehingga disarankan untuk menggunakan metode non hierarki yaitu metode K-Means.

Berdasarkan uraian latar belakang tersebut, maka peneliti tertarik untuk melakukan penelitian yang berjudul analisis cluster metode K-Means untuk pengelompokan kecamatan di Provinsi Sulawesi Selatan berdasarkan potensi desa.

## 2. Literature Review

### 2.1. Analisis Cluster

Analisis *cluster* merupakan salah satu analisis dalam statistika peubah banyak yang digunakan untuk mengelompokkan objek-objek yang memiliki karakteristik yang sama sehingga anggota dalam satu *cluster* bersifat homogen dan antar *cluster* itu sendiri bersifat heterogen (Johnson & Wichern, 2002). Seperti yang diketahui, analisis *cluster* akan membagi data menjadi beberapa kelompok. Adapun ciri sebuah *cluster* yang baik adalah sebagai berikut:

- Homogenitas yaitu kesamaan yang tinggi antara anggota dalam satu kelompok atau biasa disebut *within cluster*.
- Heterogenitas yaitu perbedaan yang tinggi antara *cluster* satu dengan kelompok yang lain atau biasa disebut *between cluster* (Alfira, 2018).

### 2.2. Ukuran Jarak

Menurut Johnson & Wichern (2002), jarak *Euclidean* merupakan salah satu ukuran jarak dalam analisis *cluster* yang paling sering digunakan untuk mengukur jarak antar obyek dan pusat *cluster*. Semakin besar jarak antara satu objek ke objek yang lain maka semakin besar pula perbedaan antara kedua objek tersebut, sehingga cenderung untuk tidak ditempatkan kedalam kelompok yang sama (Mongi, 2015). Rumus untuk menghitung ukuran jarak sebagai berikut (Johnson & Wichern, 2002):

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

$p$  adalah banyaknya peubah,  $d_{ij}$  adalah jarak antara objek ke- $i$  dan objek ke- $j$ ,  $X_{ik}$  adalah data dari objek ke  $i$  pada peubah ke  $k$ ,  $X_{jk}$  adalah data dari objek ke- $j$  pada peubah ke- $k$ .

### 2.3. Metode K-Means



Metode *K-Means* ditemukan pada tahun 1976 oleh ilmuwan bernama MacQueen JB. Metode *K-Means* merupakan salah satu metode pengelompokan data yang berusaha membagi data yang ada dalam bentuk satu atau lebih kelompok dimana data yang memiliki karakteristik yang sama di kelompokan dalam satu *cluster* sedangkan data yang memiliki karakteristik yang berbeda di kelompokan dalam *cluster* yang lain (Talakua, 2017). Adapun tahapan pengelompokan dengan metode *K-Means* yaitu sebagai berikut (Rivani, 2010):

- 1) Menentukan nilai  $k$  sebagai jumlah *cluster* yang ingin dibentuk
- 2) Membangkitkan  $k$  *centroid* (titik pusat *cluster*) secara acak
- 3) Menghitung jarak setiap *cluster* ke *centroid* terdekat dengan menggunakan jarak *euclidean*.
- 4) Kelompokkan setiap data berdasarkan jarak *centroid* terdekat.
- 5) Tentukan posisi *centroid* baru ( $C_k$ ) dengan cara menghitung nilai rata-rata dari data yang ada pada *centroid* yang sama.

$$C_k = \left(\frac{1}{n_k}\right) \sum d_i \tag{2}$$

Dimana  $n_k$  adalah jumlah anggota dalam *cluster*  $k$  dan  $d_i$  adalah anggota dalam *cluster*  $k$ .

- 6) Kembali ke langkah ke 3 jika posisi *centroid* baru dan *centroid* yang lama berubah.

#### 2.4. Davies Bouldin Index

*Davies Bouldin Index* (DBI) pertama kali diperkenalkan oleh oleh ilmuwan bernama David L. Davies dan Donald W. Bouldin pada tahun 1979. Index ini digunakan untuk mengukur validitas *cluster* pada metode pengelompokan. DBI memaksimalkan jarak inter-*cluster* dan meminimalkan jarak antara data dan titik pusat *cluster*. Semakin kecil nilai DBI yang diperoleh maka semakin baik *cluster* yang diperoleh. Rumus menghitung *Davies Bouldin Index* yaitu sebagai berikut (Jumadi, 2018):

$$DBI = \frac{1}{k} \cdot \sum_{i=1}^k \max R_{i,j} \tag{3}$$

dimana:

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \tag{4}$$

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i) \tag{5}$$

dan

$$SSB_{i,j} = d(c_i, c_j) \tag{6}$$

Dimana  $SSW$  (*Sum of Square Within-cluster*) yaitu jumlah dari kedekatan data terhadap titik pusat *cluster*,  $SSB$  (*Sum of Square Between-cluster*) yaitu jarak antar titik pusat *cluster*,  $R_{i,j}$  yaitu untuk mengetahui nilai perbandingan antara *cluster* ke- $i$  dan *cluster* ke- $j$ ,  $k$  yaitu jumlah *cluster*,  $c_i$  yaitu *centroid cluster* ke- $i$ ,  $m_i$  yaitu jumlah data dalam *cluster* ke- $i$ ,  $d(x_j, c_i)$  yaitu jarak *euclidean* setiap data ke *centroid*,  $d(c_i, c_j)$  yaitu jarak antar *centroid*.

### 2.5. Potensi Desa

Menurut UU No.6 tahun 2014, Desa adalah kesatuan masyarakat hukum yang memiliki batas wilayah yang berwenang untuk mengatur dan mengurus urusan pemerintahan, kepentingan masyarakat setempat berdasarkan prakarsa masyarakat, hak asal usul, dan/atau hak tradisional yang diakui dan dihormati dalam sistem Negara Kesatuan Republik Indonesia.

Potensi desa merupakan potensi atau kemampuan yang dimiliki oleh suatu desa yang dapat digunakan untuk memajukan desa sehingga meningkatkan kesejahteraan masyarakat (Fathia, et.al., 2016).

## 3. Research Method

The template is used to format your manuscript. Please do not alter prescribed margins, column widths, line spaces, and text fonts. Provided specifications anticipate your manuscript as part of the proceedings rather than a stand-alone document. Please do not modify any of the designations provided. Sub sub-titles should be avoided unless absolutely necessary.

### 3.1. Jenis Penelitian

Jenis penelitian yang digunakan dalam penelitian ini yaitu penelitian terapan dengan pendekatan kuantitatif yaitu dengan mengambil atau mengumpulkan data yang diperlukan dan menganalisisnya menggunakan analisis *cluster* metode *K-Means*.

### 3.2. Sumber Data

Data yang digunakan dalam penelitian ini yaitu data potensi desa tahun 2018 yang bersumber dari Badan Pusat Statistik Provinsi Sulawesi Selatan.

### 3.3. Definisi Operasional Peubah

Peubah yang digunakan dalam penelitian ini terdiri atas jumlah : SD/MI ( $X_1$ ), SMP/MTs ( $X_2$ ), SMA/SMK/MA ( $X_3$ ), puskesmas/pustu ( $X_4$ ), keluarga tanpa listrik ( $X_5$ ), desa/kelurahan menurut keberadaan pasar ( $X_6$ ), desa/kelurahan yang dilewati angkutan umum ( $X_7$ ) dan desa/kelurahan yang memiliki penerangan di jalan utama ( $X_8$ ).

### 3.4. Teknik Analisis Data

Adapun tahapan analisis data yang dilakukan dalam penelitian ini yaitu sebagai berikut:

- 1) Melakukan analisis deskriptif terhadap peubah-peubah yang digunakan.
- 2) Melakukan standarisasi pada data
- 3) Melakukan pengelompokan dengan metode *K-Means*. Adapun langkah-langkahnya antara lain:
  - a. Menentukan jumlah kelompok
  - b. Menentukan pusat *cluster* awal atau *centroid*.
  - c. Mengalokasikan masing-masing data ke pusat *cluster* terdekat. Ukuran jarak yang digunakan yaitu jarak *Euclidean*.
  - d. Menghitung kembali pusat *cluster* berdasarkan kelompok yang terbentuk. Iterasi dilakukan sampai tidak ada perpindahan objek antar *cluster*.
- 4) Menentukan *cluster* terbaik dengan menggunakan *Davies Bouldin Index*.
- 5) Melakukan interpretasi dari hasil *cluster* yang terbentuk.

## 4. Results and Discussion

### 4.1. Analisis Deskriptif

Tabel 1. Statistika Deskriptif Potensi Desa 307 Kecamatan di Provinsi Sulawesi Selatan

Peubah	Minimum	Maksimum	Rerata	Simpangan Baku
$X_1$	4	77	23	12,42
$X_2$	1	25	8	4,32
$X_3$	0	30	4	4,67
$X_4$	1	22	6	3,31
$X_5$	0	1.914	149	241,58
$X_6$	0	11	3	2,17
$X_7$	0	27	9	4,26
$X_8$	0	25	8	4,60

### 4.2. Cluster Metode K-Means

Cluster metode *K-Means* dapat digunakan untuk mengelompokkan kecamatan di Provinsi Sulawesi Selatan berdasarkan potensi desa. Adapun tahapan pengelompokannya adalah sebagai berikut:

a) Jumlah *Cluster*

Pada pengelompokan kecamatan menurut potensi desa akan dibuat menjadi 3 sampai dengan 7 *cluster*, hal ini dilakukan agar nantinya ditentukan jumlah *cluster* yang terbaik dalam mengelompokkan kecamatan berdasarkan potensi desa.

b) Menentukan pusat *cluster* awal atau *centroid*

Misalkan akan dibuat 3 kelompok, maka ditentukan pusat *cluster* awal atau *centroid* sebanyak 3 pula. Diperlihatkan pusat *cluster* awal atau *centroid* pada tabel 2.

Tabel 2. *Centroid* awal

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$
C1	0,97	0,85	0,55	0,64	-0,07	0,75	0,86	0,94
C2	-0,19	0,04	-0,32	0,48	3,43	0,44	0,06	-0,14
C3	-0,55	0,50	-0,29	-0,41	-0,26	-0,47	-0,50	-0,53

c) Mengalokasikan data ke pusat *cluster* terdekat

Proses pengelompokan data ke dalam suatu *cluster* dapat menggunakan ukuran jarak. Jarak paling dekat antara satu data dengan pusat *cluster* tertentu akan menentukan suatu data masuk pada *cluster* mana. Adapun ukuran jarak yang digunakan adalah jarak *Euclidean*.

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

Data 1

$$C_1 = \sqrt{(-1,06 - 0,97)^2 + (-1,09 - 0,85)^2 + (-0,72 - 0,55)^2 + \dots + (-0,69 - 0,94)^2}$$

$$= \sqrt{4,120 + 3,763 + 1,612 + \dots + 1,573} = 6,036$$

$$C_2 = \sqrt{(-1,06 - (-0,19))^2 + (-1,09 - 0,04)^2 + (-0,72 - (-0,32))^2 + \dots + (-0,69 - (-0,14))^2}$$

$$= \sqrt{0,756 + 1,276 + 1,081 + \dots + 0,302} = 2,991$$

$$C_3 = \sqrt{(-1,06 - (-0,55))^2 + (-1,09 - 0,50)^2 + (-0,72 - (-0,29))^2 + \dots + (-0,69 - (-0,53))^2}$$

$$= \sqrt{0,260 + 2,528 + 0,184 + \dots + 0,025} = 4,336$$

Berdasarkan hasil perhitungan jarak *euclidean* memperlihatkan bahwa jarak terdekat pada data 1 adalah dengan titik pusat *cluster* kedua. Oleh karena itu, data pertama menjadi anggota *cluster* 2. Proses perhitungan ini dilakukan seterusnya sampai dengan data ke 307.

d) Menghitung kembali pusat *cluster* berdasarkan kelompok yang terbentuk

Berdasarkan hasil pengelompokan untuk *cluster* 1 sampai dengan 3 maka langkah selanjutnya dihitung pusat *cluster* yang baru.

$$C_k = \left(\frac{1}{n_k}\right) \sum d_i$$

*Cluster* 1 ( $C_1$ )

$$x_1 = \frac{-0,09 - (-0,49) + 2,23 + 0,14 + 0,95 + 1,51 + \dots + 1,19}{107} = 0,974$$

$$x_2 = \frac{0,06 + (-0,39) + 3,30 + 0,76 + 0,52 + (0,39) + \dots + 2,37}{107} = 0,853$$

$$x_3 = \frac{-0,50 + (-0,50) + 0,98 + 2,48 + (-0,08) + (-0,08) + \dots + 2,70}{107} = 0,553$$

$$x_4 = \frac{2,14 + 0,93 + 1,23 + (-0,87) + 1,23 + 2,14 + \dots + (-0,27)}{107} = 0,641$$

$$x_5 = \frac{-0,61 + (-0,24) + (-0,46) + (-0,60) + (-0,22) + (-0,61) + \dots + (-0,61)}{107} = -0,069$$

$$x_6 = \frac{1,79 + 2,25 + 3,62 + 0,41 + 1,33 + 2,71 + \dots + (-0,96)}{107} = 0,745$$

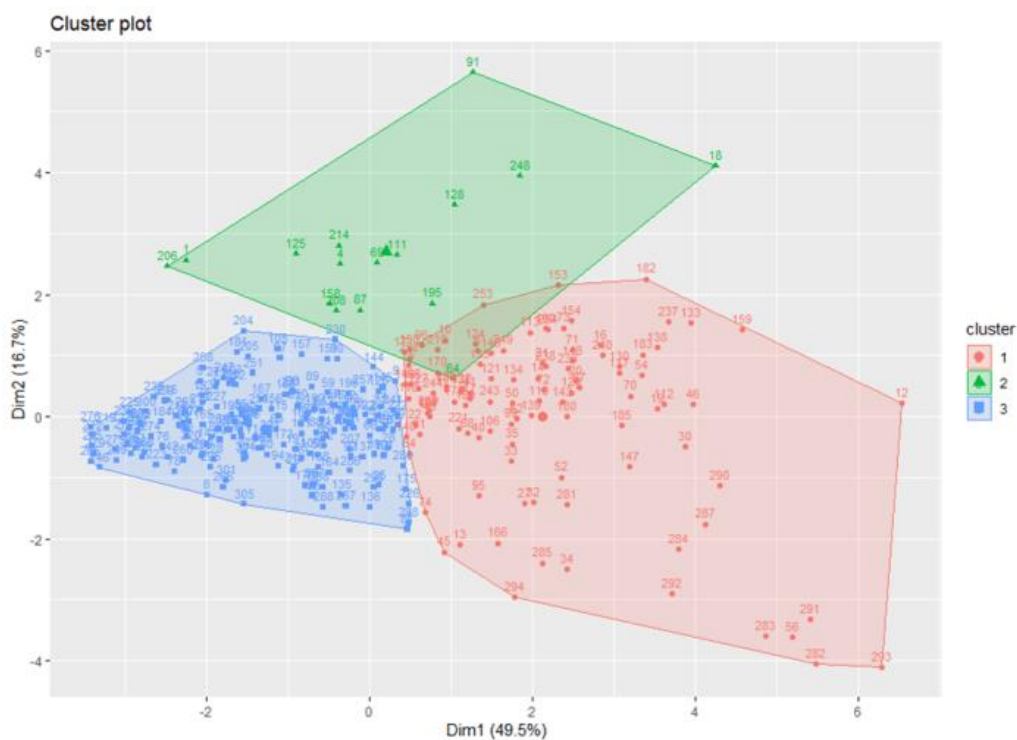
$$x_7 = \frac{0,76 + 0,76 + 2,87 + 0,06 + 0,99 + 0,99 + \dots + (-0,40)}{107} = 0,859$$

$$x_8 = \frac{-0,03 + 0,39 + 2,78 + 0,18 + 0,61 + 1,05 + \dots + (-0,03)}{107} = 0,935$$

Begitu pula perhitungan *centroid cluster* 2 dan 3. *Centroid* baru ini digunakan untuk mencari jarak terdekat antara data dan *centroid*. Proses perhitungan *centroid* dan penentuan jarak ke pusat *cluster* akan terus mengalami perulangan hingga pengelompokan data menjadi tetap.

Berikut ditampilkan plot hasil pengelompokan mulai dari kelompok sebanyak 3 sampai dengan kelompok sebanyak 7.

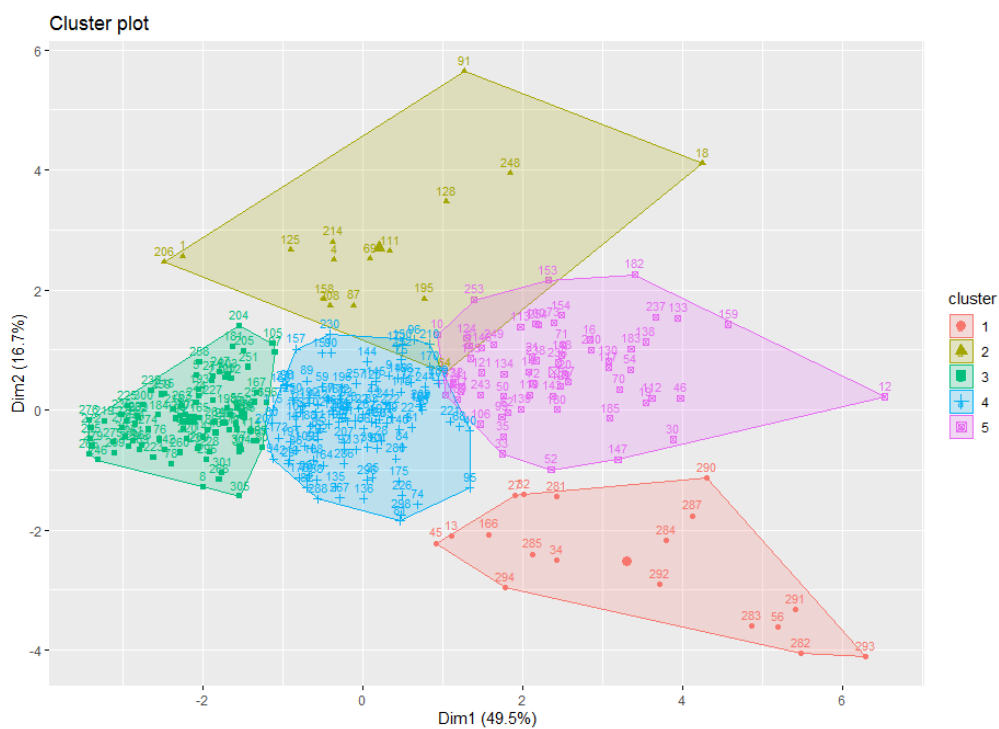
Berdasarkan gambar 1-7 terlihat bahwa hasil plot yang diperoleh mengalami irisan, hal ini disebabkan oleh adanya outlier pada data yaitu pada peubah keluarga tanpa listrik dimana nilai rata-ratanya sangat jauh berbeda dari rata-rata peubah yang lain.



Gambar 1. Plot K=3



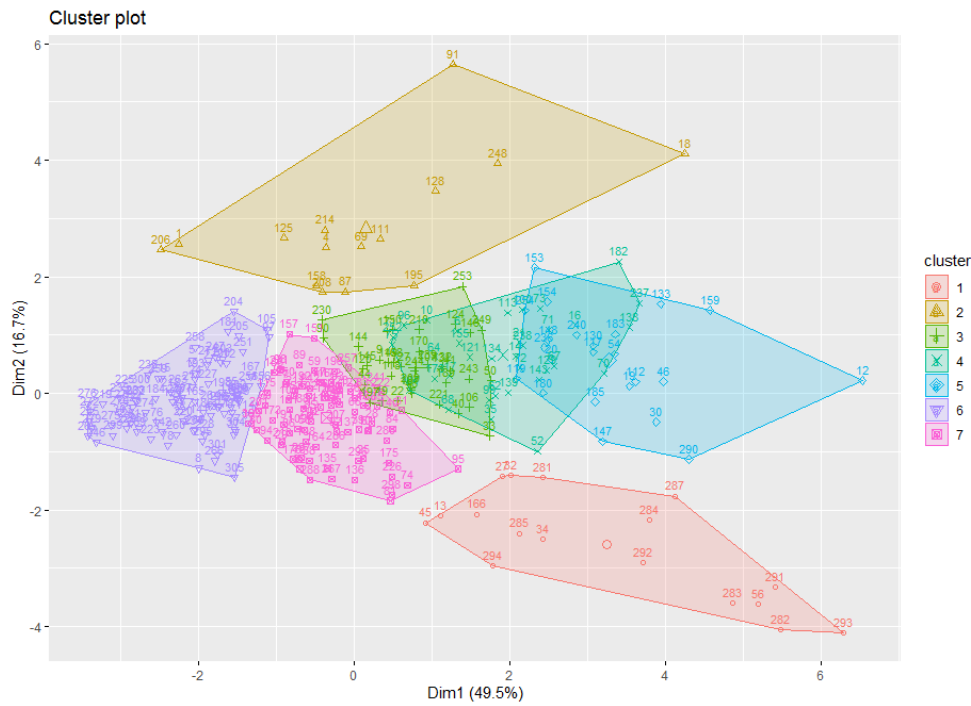
Gambar 2. Plot K=4



Gambar 3. Plot K=5



Gambar 4. Plot K=6



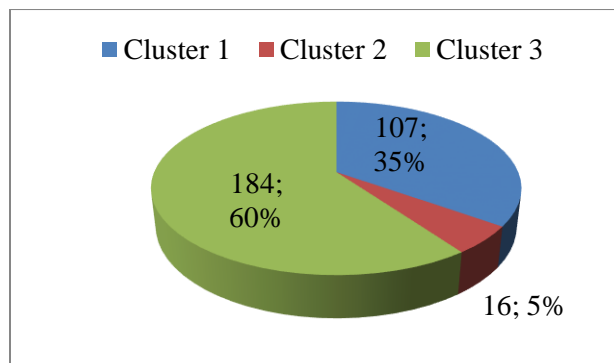
Gambar 5. Plot K=7

### 4.3. Penentuan Cluster Terbaik

Tabel 3. Hasil Davies Bouldin Index

Cluster	Davies Bouldin Index
K3	1,218
K4	1,393
K5	1,292
K6	1,475
K7	1,429

Berdasarkan tabel 3 diperoleh hasil DBI paling kecil pada saat  $K = 3$  yaitu sebesar 1,218, sehingga jumlah cluster terbaik yaitu sebanyak 3 kelompok. Berikut ditampilkan pada Gambar 6 hasil pengelompokan akhir kecamatan di Provinsi Sulawesi Selatan berdasarkan peubah potensi desa yaitu sebanyak 3 kelompok .



Gambar 6. Jumlah dan Persentase Cluster yang Terbentuk

#### 4.4. Interpretasi Keadaan Kecamatan Berdasarkan Hasil Cluster

Dengan menggunakan metode *K-Means* setelah kelompok terbentuk yaitu sebanyak 3 kelompok, untuk melihat keadaan kecamatan berdasarkan hasil pengelompokan yang terbentuk maka dari seluruh objek penelitian sebanyak 307 kecamatan di Provinsi Sulawesi Selatan diambil rata-rata dari masing-masing peubah potensi desa ( $\bar{X}$ ) dan masing-masing kelompok diambil nilai rata-rata dari setiap peubahnya pula ( $\bar{X}_c$ ).

Tabel 4. Deskriptif Kelompok Berdasarkan Rata-Rata

Peubah	$\bar{X}$	$\bar{X}_c$		
		Cluster 1	Cluster 2	Cluster 3
$X_1$	23	35	21	16
$X_2$	8	11	8	6
$X_3$	4	7	3	3
$X_4$	6	8	7	5
$X_5$	149	132	978	86
$X_6$	3	5	4	2
$X_7$	9	12	9	7
$X_8$	8	12	7	6

Kecamatan yang masuk pada *cluster 1* dicirikan oleh peubah yang memiliki nilai rata-rata yang tinggi pada jumlah SD/MI, jumlah SMP/MTs, jumlah SMA/SMK/MA, jumlah Puskesmas/pustu, jumlah Desa/Kelurahan menurut keberadaan pasar, jumlah Desa/Kelurahan yang dilewati angkutan umum, jumlah Desa/Kelurahan yang memiliki penerangan di jalan utama serta dicirikan oleh peubah yang memiliki nilai yang rendah pada jumlah keluarga tanpa listrik.

Pada *Cluster 2*, kecamatan yang termasuk dalam kelompok ini merupakan kecamatan yang dicirikan oleh peubah yang memperoleh nilai rata-rata yang rendah pada jumlah SD/MI, SMA/SMK/MA, dan jumlah Desa/Kelurahan yang memiliki penerangan di jalan utama serta dicirikan pula oleh rata-rata peubah keluarga tanpa listrik yang tinggi.

Kecamatan yang berada pada *cluster 3* dikelompokkan berdasarkan peubah yang memiliki nilai rata-rata yang rendah pada setiap peubah nya. Sehingga kecamatan yang berada pada *cluster* ini merupakan kelompok dengan potensi desa terendah dari kelompok lainnya.

## 5. Conclusion

Berdasarkan hasil dan pembahasan yang telah dilakukan untuk mengelompokkan kecamatan di Provinsi Sulawesi Selatan berdasarkan Potensi Desa maka diperoleh kesimpulan sebagai berikut:

- 1) Hasil *cluster* yang terbentuk yaitu sebanyak 3 dimana *cluster* pertama terdiri atas 107 kecamatan dengan persentase sebesar 35%, *cluster* kedua terdiri atas 16 kecamatan dengan persentase sebesar 5%, dan *cluster* ketiga terdiri atas 184 kecamatan dengan persentase sebesar 60%.
- 2) Kecamatan yang berada pada *cluster* pertama merupakan kecamatan yang memiliki potensi desa yang tinggi, *cluster* kedua merupakan kecamatan yang memiliki potensi desa yang sedang dan kecamatan yang berada pada *cluster* ketiga memiliki potensi desa yang rendah.

## References

- Alfira, A. (2018). Analisis *Hybrid Mutual Clustering* Menggunakan Jarak *Square Euclidean*.  
Badan Pusat Statistika. (2018). *Potensi Desa Provinsi Sulawesi Selatan 2018*.



- Badan Pusat Statistika. (2018). *Booklet Pedoman Pencacah Desa/Kelurahan Podes 2018*.
- Badan Pusat Statistika. (2019). *Booklet Provinsi Sulawesi Selatan dalam Angka 2019*. Unggahan: Badan Pusat Statistik Provinsi Sulawesi Selatan.
- Bastian, dkk. (2018). Penerapan Algoritma *K-Means Clustering Analysis* pada Penyakit Menular Manusia. (Studi Kasus Kabupaten Majalengka). *Jurnal Sistem Informasi*, 26-32.
- Fathia, A. N. dkk (2016). Analisis Klaster Kecamatan di Kabupaten Semarang Berdasarkan Potensi Desa Menggunakan Metode *Ward* dan *Single Linkage*. *Jurnal Gaussian*, 801-810.
- Johnson & Wichern. (2002). *Applied Multivariate Analysis*, Fifth Edition. New Jersey: Prentice Hall, Inc.
- Jumadi, B. (2018). Peningkatan Hasil Evaluasi *Clustering Davies Bouldin Index* dengan Penentuan Titik Pusat *Cluster* Awal algoritma *K-Means*.
- Mongi, C. E. (2015). Penggunaan Analisis Two Step Clustering untuk Data Campuran. *JdC Vol.4 No.1*.
- Rachmatin, D. (2019). Perbandingan Antara Metode Agglomeratif, Metode Divisif dan Metode *K-Means* dalam Analisis Klaster.
- Rivani, E. (2010). Aplikasi *K-Means Cluster* untuk Pengelompokan Provinsi Berdasarkan Produksi Padi, Jagung, Kedelai, dan Kacang Hijau Tahun 2009. *Jurnal Mat Stat*, 122-134.
- Talakua, M. W. (2017). Analisis *Cluster* dengan Menggunakan Metode *K-Means* untuk Pengelompokan Kabupaten/Kota di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014. *Jurnal Ilmu Matematika dan Terapan*, 119-128.