

# Analisis Kluster Pengelompokan Kecamatan di Sulawesi Barat Berdasarkan Indikator Pendidikan

Hikmah<sup>\*1</sup>, Fardinah<sup>2</sup>, Laila Qadrini<sup>3</sup>, Elviani Tande<sup>4</sup>

<sup>1,2,3,4</sup>Uniiiversitas Sulawesi Barat

e-mail: <sup>1\*</sup>[hikmah@unsulbar.ac.id](mailto:hikmah@unsulbar.ac.id), <sup>2</sup>[fardinah@unsulbar.ac.id](mailto:fardinah@unsulbar.ac.id), <sup>3</sup>[laila.qadrini@unsulbar.ac.id](mailto:laila.qadrini@unsulbar.ac.id),  
<sup>4</sup>[elvianitande@unsulbar.ac.id](mailto:elvianitande@unsulbar.ac.id)

## Abstrak

Berdasarkan Biro Pusat Statistik 2018, IPM Sulawesi Barat berada di peringkat ke 31 dari 34 provinsi di Indonesia, yang meliputi komposisi pencapaian pendidikan, kesehatan, dan penghasilan per kepala. Salah satu faktor yang mempengaruhi adalah pemerataan dan karakteristik tingkat pendidikan. Tujuan penelitian ini adalah mengkluster kecamatan di Sulawesi Barat berdasarkan indikator pendidikan dengan menggunakan analisis kluster metode ward. Analisis kluster merupakan salah satu metode analisis yang bertujuan untuk mengkluster objek-objek berdasarkan kemiripan karakteristik yang dimilikinya. Dari hasil penelitian diperoleh 3 kluster, yaitu kluster 1 (tinggi) terdiri atas 3 kecamatan, kluster 2 (sedang) terdiri atas 19 kecamatan, dan kluster 3 (rendah) terdiri atas 20 kecamatan. Pengklasteran ini menggunakan jumlah siswa dan jumlah guru masing-masing dari KB, TK, SD, SMP, SMA, dan SMK. Dengan adanya hasil penelitian ini, semoga pemerintah dapat memperhatikan pemerataan kecamatan khususnya berdasarkan indikator pendidikan.

**Kata kunci:** Analisis Kluster, Pendidikan, Metode Ward

## 1. PENDAHULUAN

Kualitas Pendidikan di Indonesia khususnya di Sulawesi Barat masih sangat memprihatinkan. Hal ini dinyatakan oleh Badan Program Pembangunan di bawah PBB (*United Nations Development Programme/ UNDP*) dalam laporan *Human Development Report* 2016 yang mencatat Indeks Pembangunan Manusia (IPM) Indonesia berapa di peringkat 113, yang sebelumnya berada di posisi 110 tahun 2015. Berdasarkan Biro Pusat Statistik 2018, IPM Sulawesi Barat berada di peringkat ke 31 dari 34 provinsi di Indonesia, yang meliputi komposisi pencapaian pendidikan, kesehatan, dan penghasilan per kepala. Rusman (2019). Salah satu faktor yang mempengaruhi tingkat IPM adalah pemerataan pendidikan. Untuk melihat sejauh mana pemerataan pendidikan atau karakteristik tingkat pendidikan kabupaten atau kota dapat dilakukan kecamatan di Sulawesi Barat berdasarkan indikator pendidikan.

Analisis kluster sering dijumpai dalam kehidupan sehari-hari, baik dalam bidang kesehatan, sosial, pertanian, dan lainnya. Analisis kluster adalah salah satu metode analisis yang bertujuan untuk mengkan objek-objek berdasarkan kemiripan karakteristik yang dimilikinya. Metode yang digunakan dalam analisis kluster yaitu metode hierarki, yang terdiri atas *single linkage*, *complete linkage*, *average linkage*, *ward*, dan *centroid* dengan ukuran jarak *Euclidean*. L.S., Dewi, dkk. (2021).

Penelitian sebelumnya dilakukan oleh Awaliah, R. (2018) dengan judul “Analisis Cluster dengan *Average Lincage Method* dan *Ward’s Method* untuk Data Responden Nasabah Asuransi Jiwa Unit Link”. Pada penelitian ini dapat disimpulkan bahwa alasan dokter, dosen, guru, pedagang, pengusaha, pengacara, teknisi, dan PNS memutuskan untuk membeli asuransi jiwa unit link adalah untuk persiapan pendidikan anak. Penelitian lain dilakukan oleh L.S., Dewi, dkk (2021) dengan judul “Analisis Kluster untuk Pengelompokan

Kabupaten/ Kota di Provinsi Maluku Berdasarkan Indikator Pendidikan dengan Menggunakan Metode Ward". Hasil penelitiannya menyebutkan bahwa terdapat 3 tingkat pendidikan, yaitu tinggi, sedang, dan rendah.

Berdasarkan uraian di atas, peneliti tertarik melakukan penelitian pengklasteran kecamatan di Provinsi Sulawesi Barat berdasarkan indikator pendidikan dengan menggunakan metode Ward untuk melihat sejauh mana tingkat pemerataan pendidikan dan karakteristik tingkat pendidikan kecamatan di Sulawesi Barat. Dengan adanya pengelompokan ini diharapkan pemerintah dapat memberikan perhatian khusus bagi kabupaten atau daerah yang masih berada pada rendah.

### 1.1 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. Secara umum analisis multivariat dibagi menjadi dua, yaitu analisis dependensi dan analisis interdependensi. Ciri dari analisis dependensi adalah adanya satu atau beberapa variabel yang berfungsi sebagai variabel tergantung dan variabel bebas, seperti, analisis regresi linear berganda, analisis diskriminan, analisis logit, dan analisis korelasi kanonik. Ciri dari analisis interdependensi adalah semua variabelnya bersifat independen. Berikut ini yang termasuk dalam analisis interdependensi adalah analisis faktor, analisis kluster dan *multidimensional scaling*. Sarwono (2007).

Menurut Fadhli (2011), data dalam analisis multivariat dapat dinyatakan dalam bentuk matriks dimana terdapat  $n$  objek dan  $p$  variabel. Misalkan pada suatu pengamatan terdapat  $p$  variabel yaitu  $X = (X_1, X_2, \dots, X_p)$  dari banyaknya  $n$  objek, maka data tersebut dapat disajikan dalam bentuk matriks seperti di bawah ini:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} & x_{21} & x_{22} & \dots & x_{2p} & \vdots & \vdots & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

### 1.2 Analisis Kluster

Analisis kluster merupakan suatu teknik analisis multivariat yang bertujuan untuk mengklusterkan data observasi ataupun variabel-variabel ke dalam kluster sedemikian rupa sehingga masing-masing bersifat homogen sesuai dengan faktor yang digunakan untuk melakukan pengklasteran. Karena yang diinginkan adalah untuk mendapatkan kluster yang sehomogen mungkin, maka yang digunakan sebagai dasar untuk mengklusterkan adalah kesamaan skor nilai yang dianalisis. Data mengenai ukuran kesamaan tersebut dapat dianalisis dengan analisis kluster sehingga dapat ditentukan siapa yang masuk kluster mana. Gudono (2011).

Tujuan analisis kluster adalah mengelompokkan objek yang mirip ke dalam kluster yang sama. Oleh karena itu memerlukan beberapa ukuran untuk mengetahui seberapa mirip atau berbeda objek-objek tersebut. Pendekatan yang biasa digunakan adalah mengukur kemiripan yang dinyatakan dalam jarak (*distance*) antara pasangan objek. Pada analisis kluster terdapat tiga ukuran untuk mengukur kesamaan antar objek, yaitu ukuran asosiasi, ukuran korelasi, dan ukuran kedekatan.

Proses pembentukan kluster dapat dilakukan dengan dua cara, yaitu dengan metode hierarki dan non hierarki. Pada metode hierarki terdiri dari metode *agglomerative* dan metode defisit. Metode *agglomerative* sendiri terdiri dari 3 metode, yaitu metode *linkage*, metode *variance*, dan metode *centroid*, dimana *linkage* terdiri dari metode *single linkage*, *complete linkage*, dan *average linkage*. Sedangkan pada metode *variance* terdiri dari metode *Ward*. Metode non hierarki terdiri dari 3 metode yaitu, metode *sequential threshold*, metode *parallel*, dan metode *optimizing partitioning*. Simamora (2005).

#### 1.2.1 Metode Hierarki

Metode hierarki (*hierarchical method*) adalah suatu metode pada analisis kluster yang membentuk tingkatan tertentu seperti pada struktur pohon karena proses pengklasterannya dilakukan secara bertingkat/ bertahap. Hasil pengklasteran dengan metode hierarki dapat disajikan dalam bentuk dendogram. Dendogram adalah representasi visual dari langkah-langkah dalam analisis kluster yang menunjukkan bagaimana terbentuk dan nilai koefisien jarak pada setiap langkah. Angka di sebelah kanan adalah objek penelitian, di mana objek-objek tersebut dihubungkan oleh garis dengan objek yang lain sehingga pada akhirnya akan membentuk satu. (Simamora, 2005).

Tahap-tahap pengklasteran data dengan menggunakan metode hierarki adalah (Gudono, 2011):

- 1) Menentukan  $k$  sebagai jumlah kluster yang ingin dibentuk
- 2) Setiap data objek dianggap sebagai kluster sehingga  $n = N$
- 3) Menghitung jarak antar kluster
- 4) Mencari dua yang mempunyai jarak antar paling minimal dan menggabungkannya (berarti  $N = n - 1$ )
- 5) Jika  $n > k$ , maka kembali ke langkah 3.

Metode-metode yang bisa digunakan dalam metode hierarki adalah metode agglomerative (*agglomerative method*) dan metode defisif (*divisive method*).

Metode varians bertujuan untuk memperoleh kluster yang memiliki varians internal yang sekecil mungkin. Metode varians yang umum dipakai adalah metode *Ward* dimana rata-rata untuk setiap kluster dihitung. Selanjutnya, menghitung jarak *Euclidean* antara setiap objek dan nilai rata-rata. Pada setiap tahap, dua kluster yang memiliki kenaikan '*sum of squares dalam cluster*' yang terkecil digabungkan. Simamora (2005).

Metode *Ward* merupakan suatu metode pembentukan kluster yang didasari oleh hilangnya informasi akibat penggabungan objek menjadi kluster. Hal ini diukur dengan menggunakan jumlah total dari deviasi kuadrat pada *mean cluster* untuk setiap pengamatan. *Error sum of squares (SSE)* digunakan sebagai fungsi objektif. Dua objek akan digabungkan jika mempunyai fungsi objektif terkecil di antara kemungkinan yang ada.

$$SSE = \sum_{j=1}^p \left( \sum_{i=1}^n X_{ij}^2 - \frac{1}{n} \left( \sum_{i=1}^n X_{ij} \right)^2 \right)$$

dengan:

$X_{ij}$ : nilai untuk objek ke- $i$  pada kluster ke- $j$

$p$  : banyaknya variabel yang diukur

$n$  : banyaknya objek dalam kluster yang terbentuk

Langkah penyelesaian dengan metode *Ward*:

- a) Dimulai dengan memperhatikan  $N$  yang mempunyai satu responden perkluster (semua responden dinggap sebagai kluster). Pada tahap pertama ini SSE bernilai nol.
- b) Kluster pertama dibentuk dengan memilih dua dari  $N$  kluster yang memiliki nilai SSE terkecil.
- c)  $N-1$  kumpulan kluster kemudian diperhatikan kembali untuk menentukan dua dari kluster ini yang bisa meminimumkan heterogenan. Dengan demikian  $N$  kluster secara sistematis dikurangi  $N-1$ .
- d) Mengulangi langkah (c) dan (d), sampai diperoleh satu atau semua responden bergabung menjadi satu .

### 1.2.2 Metode Non Hierarki

Metode non hierarki sering disebut sebagai metode *k-means*. Prosedur pada metode non hierarki dimulai dengan memilih sejumlah nilai awal sesuai dengan jumlah yang diinginkan, kemudian objek pengamatan digabungkan ke dalam kluster-kluster tersebut. Metode non hierarki ini meliputi metode *sequential threshold*, *parallel threshold*, dan *optimizing partitioning*. Gudono (2011).

Pada metode *Sequential Threshold* dimulai dengan pemilihan satu kluster dan menempatkan semua objek yang berada pada jarak terdekat ke dalam kluster tersebut. Jika semua objek yang berada pada ambang batas tertentu telah dimasukkan, kemudian kluster yang kedua dipilih dan menempatkan semua objek yang berada pada jarak terdekat ke dalamnya. Kemudian kluster ketiga dipilih dan proses dilanjutkan seperti yang sebelumnya.

Secara prinsip sama dengan metode *sequential threshold*, hanya saja pada metode *parallel threshold* dilakukan pemilihan terhadap beberapa objek awal sekaligus dan kemudian melakukan penggabungan objek ke dalamnya secara bersamaan. Pada saat proses berlangsung, jarak terdekat dapat ditentukan untuk memasukkan beberapa objek ke dalam kluster-kluster.

Metode *Optimization partitioning* hampir mirip dengan metode *Sequential Threshold* dan metode *Parallel Threshold* yang membedakan adalah metode *optimization* ini memungkinkan untuk menempatkan kembali objek-objek ke dalam kluster yang lebih dekat atau dengan melakukan optimasi pada penempatan objek yang ditukar untuk kluster lainnya dengan pertimbangan kriteria optimasi.

Masalah utama dalam analisis kluster adalah menentukan berapa banyaknya kluster. Sebetulnya tidak ada aturan yang baku untuk menentukan berapa sebetulnya banyaknya kluster, namun demikian ada beberapa petunjuk yang bisa dipergunakan, yaitu:

- 1) Pertimbangan teoretis, konseptual, praktis, mungkin bisa diusulkan/disarankan untuk menentukan berapa banyaknya kluster yang sebenarnya. Sebagai contoh, jika tujuan pengklasteran untuk mengenali/mengidentifikasi segmen pasar, manajemen mungkin menghendaki kluster dalam jumlah tertentu (katakan 3, 4, atau 5).
- 2) Besarnya relatif kluster seharusnya berguna/bermanfaat. Supranto (2004)

### 1.3 Prosedur Analisis Kluster

Secara umum dalam melakukan analisis kluster diperlukan proses sebagai berikut:

1. Merumuskan masalah, dalam hal ini dilakukan pemilihan variabel-variabel yang akan dipergunakan untuk membentuk kluster. Hal yang paling penting di dalam masalah analisis kluster adalah pemilihan variabel-variabel yang akan dipergunakan untuk pengklasteran (pembentukan kluster). Memasukkan satu atau dua variabel yang tidak relevan dengan masalah pengklasteran sehingga akan menyebabkan penyimpangan hasil pengklasteran yang kemungkinan besar sangat bermanfaat. Supranto (2004).

2. Melakukan uji asumsi

- a. Menguji kecukupan sampel

Untuk mengetahui apakah sampel yang diambil benar-benar dapat mewakili populasi yang ada dibutuhkan nilai *Keiser Meyer Olkin* (KMO). Nilai KMO kurang dari 0,5 menandakan bahwa sampel yang diambil tidak dapat mewakili populasi yang ada. Machfudhoh, S. dan Wahyuningsi, N. (2013)

- b. Menguji multikolinearitas

Data yang saling berhubungan akan menyebabkan hasil yang bias pada variabel yang saling berhubungan. Apabila terjadi multikolinearitas maka sebaiknya mengeluarkan salah satunya, atau dengan menggunakan alternatif lain, yaitu melakukan *composit variabels* yang dapat dianalisis lebih lanjut. Mattjik, Ahmad Ansori (2011).

Untuk mendeteksi ada atau tidaknya multikolinearitas dapat dilakukan dengan:

- 1) Menghitung koefisien korelasi sederhana antara variabel bebas, jika terdapat nilai yang mencapai atau melebihi 0,8 maka terjadi multikolinearitas.
- 2) Menghitung nilai *Variance Inflation Factor* (VIF), jika nilai VIF melebihi 10 maka hal tersebut menunjukkan multikolinearitas.

3. Menentukan ukuran kemiripan objek

Dalam menjelaskan atau mengukur kemiripan/ kesamaan antar objek digunakan pendekatan ukuran jarak (*distance*). Semakin besar nilai jarak antara objek mengindikasikan semakin besar pula perbedaan karakteristik antara objek. Sedangkan semakin kecil nilai jarak antara objek berarti semakin tinggi pula kemiripan/ kesamaan karakteristik antara objek. Hasrul (2018).

Misalkan terdapat variabel  $X_1, X_2, X_3, \dots, X_p$ . Hasil pengukuran variabel  $X_j$  pada objek  $a_i$  dinyatakan dengan  $d_{ij}$  (untuk setiap  $i = 1, 2, \dots, n$  dan  $j = 1, 2, \dots, p$ ). Selanjutnya dapat dibentuk matriks

$$d_{ij} = [d_{11} \ d_{12} \ \dots \ d_{1p} \ d_{21} \ d_{22} \ \dots \ d_{2p} \ \vdots \ \dots \ \vdots \ d_{n1} \ d_{n2} \ \dots \ d_{np}]$$

$d_{ij}$  menjelaskan jarak antara individu ke- $i$  dan ke- $j$ , dengan jumlah objek adalah  $n$  dan  $p$  sama. Misalkan  $a$  dan  $b$  adalah jarak antar objek. Fungsi jarak  $d(a, b)$  mempunyai sifat:

- Simetri,  $d(a, b) = d(b, a)$
- Positif,  $d(a, b) \geq 0$
- $d(a_1, a_2) = 0$
- $(a, b)$  meningkat seiring tidak miripnya  $a$  dan  $b$
- $(a, c) \leq d(a, b) + d(b, c)$

Apabila nilai  $d$  semakin kecil, maka semakin besar pula kemiripan antara kedua objek tersebut. Sebaliknya jika nilai  $d$  besar, maka semakin besar pula ketidakmiripan dari kedua objek tersebut. Safa'at Yulianto dan Kishera (2014).

Terdapat dua pengamatan dalam ruang  $p$  dimensi  $x' = (x_1, x_2, \dots, x_p)$  dan  $y' = (y_1, y_2, \dots, y_p)$ . Jarak dari dua pengamatan tersebut dapat dihitung dengan berbagai cara, seperti jarak *Euclid*, statis, atau metrik *minkowski*. Rumus perhitungan jarak *Euclid* dapat dituliskan dengan

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2}$$
$$= \sqrt{(x - y)'(x - y)}$$

4. Melakukan pengklasteran dengan metode ward.
5. Menginterpretasikan kluster. Pada tahap interpretasi meliputi pengujian pada masing-masing kluster yang terbentuk untuk memberikan nama atau keterangan secara tepat sebagai gambaran sifat dari kluster tersebut, menjelaskan bagaimana mereka bisa berbeda secara relevan pada tiap dimensi. Ketika memulai proses interpretasi digunakan rata-rata (*centroid*) setiap kluster pada setiap variabel.

## 2. METODE PENELITIAN

### 2.1 Jenis Penelitian

Jenis penelitian ini adalah penelitian kuantitatif. Melakukan pengklasteran kecamatan di Sulawesi Barat berdasarkan indikator pendidikan dengan menggunakan data sekunder yang diperoleh dari Kementerian Pendidikan, Kebudayaan Riset dan Teknologi pada tahun 2021. Dalam penelitian ini diklaster berdasarkan banyak siswa dan guru mulai dari KB sampai SMA sederajat.

### 2.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah variabel yang menjadi kriteria dalam pengelompokan kecamatan berdasarkan indikator pendidikan. Adapun indikator-indikatornya, yaitu:

$X_1$ = Jumlah siswa TK	$X_7$ = Jumlah guru KB
$X_2$ = Jumlah siswa KB	$X_8$ = Jumlah guru TK
$X_3$ = Jumlah siswa SD	$X_9$ = Jumlah guru SD
$X_4$ = Jumlah siswa SMP	$X_{10}$ = Jumlah guru SMP
$X_5$ = Jumlah siswa SMA	$X_{11}$ = Jumlah guru SMA
$X_6$ = Jumlah siswa SMK	$X_{12}$ = Jumlah guru SMK

### 2.3 Langkah-langkah Penelitian

Langkah-langkah yang dilakukan dalam pelaksanaan penelitian ini, yaitu:

- 1) Melakukan statistik deskriptif.
- 2) Melakukan pengujian asumsi analisis kluster
  - a. Sampel harus mewakili populasi (dilakukan dengan menggunakan nilai *Kaiser Mayer Olkin* atau *KMO*)
  - b. Multikolinearitas (dilakukan dengan menggunakan korelasi pearson)
- 3) Melakukan analisis kluster dengan metode ward
- 4) Menentukan banyaknya kluster.
- 5) Melakukan interpretasi analisis kluster dengan menggunakan metode ward

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Deskripsi Data

Penelitian ini menggunakan data sekunder, yaitu data yang diperoleh dari data dapodik kemdikbud. (RI 2021). Data yang digunakan yaitu jumlah siswa KB, TK, SD, SMP, SMA, dan SMK serta jumlah guru KB, TK, SD, SMP, SMA, SMK pada semester ganjil 2021/2022.

Tabel 1 Analisis deskriptif

Variabel	Minimum	Kuartil 1	Median	Mean	Maksimum
$X_1$	41	221,5	390	429	1188
$X_2$	46	112	220	354,3	2291
$X_3$	255	1070	1924	2323	7528
$X_4$	16	574	838	978	3318
$X_5$	75	244	367	598	3251
$X_6$	2	166	272	501,4	1994
$X_7$	7	19,5	41,5	39,43	110
$X_8$	2	9	18	25,24	142
$X_9$	64	136,2	185	202,4	504
$X_{10}$	25	52	75	89,52	201
$X_{11}$	8	21,5	31	45,81	212
$X_{12}$	0	16,5	31,5	42,69	150

Dari tabel di atas, diperoleh bahwa jumlah maksimal dan minimum siswa TK berturut-turut 1.188 orang dan 41 orang dengan rerata 429 orang. Jumlah maksimal dan minimum siswa KB berturut-turut 2291 orang dan 46 orang dengan rerata  $354,3 \approx 355$  orang. Jumlah maksimal dan minimum siswa SD berturut-turut 7528 orang dan 255 orang dengan rerata 2323 orang. Jumlah maksimal dan minimum siswa SMP berturut-turut 3318 orang dan 16 orang dengan rerata 978. Jumlah maksimal dan minimum siswa SMA berturut-turut 3251 orang dan 75 orang dengan rerata 598 orang. Jumlah maksimal dan minimum siswa SMK berturut-turut 1994 orang dan 2 orang dengan rerata  $501,4 \approx 50$  orang. Jumlah maksimal dan minimum guru TK berturut-turut 110 orang dan 7 orang dengan rerata  $39,43 \approx 40$  orang. Jumlah maksimal dan minimum guru KB berturut-turut 142 orang dan 2 orang dengan rerata  $25,25 \approx 26$  orang. Jumlah maksimal dan minimum guru SD berturut-turut 504 orang dan 64 orang dengan rerata  $202,4 \approx 203$ . Jumlah maksimal dan minimum guru SMP berturut-turut 201 orang dan 25 orang dengan rerata  $89,52 \approx 90$  orang. Jumlah maksimal dan minimum guru SMA berturut-turut 212 orang dan 8 orang dengan rerata  $45,81 \approx 46$  orang. Jumlah maksimal dan minimum guru SMK berturut-turut 150 orang dan 0 orang dengan rerata  $42,69 \approx 43$  orang.

#### 3.2 Hasil dan Pembahasan

##### 3.2.1 Uji Asumsi Analisis Kluster

Analisis pertama, asumsi kecukupan sampel menggunakan nilai *Kaiser Meyer Olkin* (KMO). Dengan menggunakan R dilakukan analisis KMO dan diperoleh nilai KMO sebesar 0,92866. Karena nilai KMO  $> 0,5$  artinya sampel cukup dan layak untuk di.

Asumsi kedua adalah multikolinearitas. Untuk mengetahui ada tidaknya multikolinearitas dapat dilihat dari nilai-nilai korelasi pada matriks korelasi. Dikatakan terdapat multikolinearitas apabila nilai koefisien korelasinya lebih besar dari 0,8. Dengan menggunakan program R diperoleh nilai-nilai korelasi sebagai berikut.

Tabel 2 Nilai koefisien korelasi antar variabel

Variabel	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>	X <sub>10</sub>	X <sub>11</sub>	X <sub>12</sub>
X <sub>1</sub>	1	0,32	0,50	0,67	0,72	0,67	0,84	0,37	0,67	0,74	0,69	0,69
X <sub>2</sub>	0,32	1	0,10	0,38	0,22	0,21	0,30	0,81	0,42	0,40	0,26	0,16
X <sub>3</sub>	0,50	0,10	1	0,76	0,62	0,65	0,53	0,32	0,74	0,67	0,77	0,57
X <sub>4</sub>	0,67	0,38	0,76	1	0,79	0,67	0,66	0,63	0,77	0,76	0,86	0,66
X <sub>5</sub>	0,72	0,22	0,62	0,79	1	0,74	0,68	0,54	0,60	0,70	0,87	0,77
X <sub>6</sub>	0,67	0,21	0,65	0,67	0,74	1	0,68	0,39	0,66	0,66	0,75	0,80
X <sub>7</sub>	0,84	0,30	0,53	0,66	0,68	0,68	1	0,39	0,75	0,81	0,72	0,66
X <sub>8</sub>	0,37	0,81	0,32	0,63	0,54	0,39	0,39	1	0,53	0,46	0,53	0,37
X <sub>9</sub>	0,67	0,42	0,74	0,77	0,60	0,66	0,75	0,53	1	0,80	0,75	0,66
X <sub>10</sub>	0,47	0,40	0,67	0,76	0,70	0,66	0,81	0,46	0,80	1	0,80	0,68
X <sub>11</sub>	0,69	0,26	0,77	0,86	0,87	0,75	0,72	0,53	0,75	0,80	1	0,73
X <sub>12</sub>	0,69	0,16	0,57	0,66	0,77	0,80	0,66	0,37	0,66	0,68	0,73	1

Berdasarkan tabel 2 diperoleh bahwa variabel X<sub>1</sub> dengan X<sub>7</sub>, X<sub>4</sub> dengan X<sub>11</sub>, serta X<sub>5</sub> dengan X<sub>11</sub> mempunyai koefisien korelasi lebih dari 0,8. Artinya tiga pasang variabel tersebut mengalami multikolinearitas. Oleh karena itu, selanjutnya akan dilakukan analisis komponen utama atau *Principal Component Analysis* (PCA). Sebelum melakukan analisis PCA, terlebih dahulu melakukan analisis *Measure of Sampling Adequacy* (MSA). Nilai MSA dilakukan untuk mengetahui variabel-variabel apa saja yang layak dianalisis PCA. Dengan menggunakan program R diperoleh.

Tabel 3 Nilai MSA

Variabel	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>	X <sub>10</sub>	X <sub>11</sub>	X <sub>12</sub>
MSA	0,8	0,5	0,8	0,9	0,8	0,9	0,8	0,6	0,8	0,8	0,9	0,8
	6	4	8	4	6	2	6	8	9	9	3	9

Pada tabel 3 menunjukkan nilai MSA seluruh variabel lebih dari 0,5. Hal ini menunjukkan bahwa variabel tersebut layak untuk dianalisis PCA. Dengan menggunakan program R diperoleh hasil sebagai berikut.

Tabel 4 Nilai atau *score principal component* (PC)

Kecamatan	<i>Principal Component</i> (PC)	
	PC <sub>1</sub>	PC <sub>2</sub>
Mamuju	-5843.5437	-59.90754
Kalukku	-4349,58	-1486.10229
Simboro	-1142.3764	-930.85513
Tapalang	1832,2508	910,073
Tommo	-342.9611	-411.28354
Papalang	-119.4994	-268.90347
Kalumpang	979.2244	-246.72920
⋮	⋮	⋮
Kec. Pangale	1098,6874	-143,21353

Pada tabel 4 di atas diperoleh nilai atau *score Principal Component* untuk setiap kecamatan. Nilai tersebut akan digunakan untuk melakukan pengelompokan kecamatan di Sulawesi Barat dengan menggunakan metode ward.

Tahap selanjutnya adalah menghitung jarak masing-masing objek. Dengan menggunakan program R diperoleh hasil pada lampiran. Secara ringkas, dapat dilihat pada tabel 5

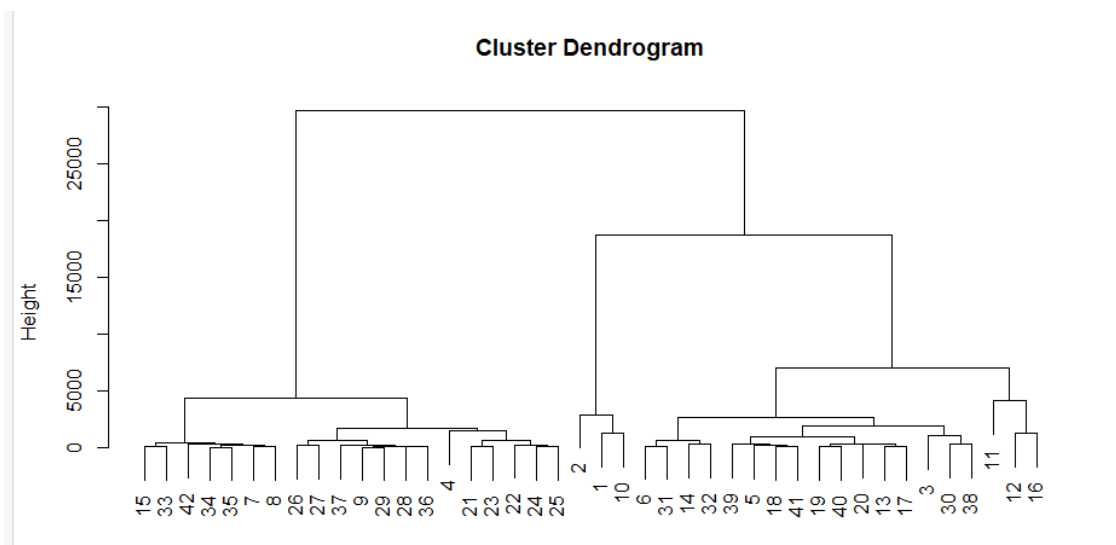
Tabel 5 Jarak *Euclid* Kecamatan

Kecamatan	Mamuju	Kalukku	...	Pangale
Mamuju	0	2065.42054	...	2883.16867
Kalukku	2065.42054	0	...	4221.12257
⋮	⋮	⋮	⋮	⋮
Pangale	2883.16867	4221.12257	...	0

Selanjutnya adalah melakukan analisis kluster dengan menggunakan metode ward. Hasil pengklasteran dengan program R dapat dilihat pada gambar 1 dendrogram dengan metode ward.

Berdasarkan gambar 1, indeks D menunjukkan terdapat 42 kecamatan. Dari gambar tersebut terlihat bahwa terdapat tiga kluster yang terbentuk dengan menggunakan metode ward. Selanjutnya, dengan menggunakan program R, diperoleh hasil pada tabel 6.

Berdasarkan tabel 6 terdapat tiga kluster beserta anggotanya, kluster 1 terdiri atas 3 anggota yaitu Mamuju, Kalukku, Polewali, kluster 2 terdiri atas 19 anggota yaitu Simboro, Tommo, Papalang, Campalagian, Wonomulyo, Balanipa, Tapango, Banggae Timur, Sendana, Pamboang, Malunda, Mamasa, Pasangkayu, BAMBALAMUTU, Baras, Topoyo, Karossa, Tobadak, Budong-Budong, serta kluster 3 terdiri atas 20 anggota yaitu Tapalang, Kalumpang, Tapalang Barat, Bonehau, Alu, Sumarorong, Mambi, Tabulahan, Bambang, Tandukalua, Rantebulahan Timur, Buntumalangka, Aralle, Sesenapadang, Tikke Raya, Dapurang, Sarudu, Sarjo, Lariang, Pangale.



Gambar 1 dendrogram

Tabel 6. beserta anggotanya dengan metode ward

Kluster	Kecamatan
1	Mamuju, Kalukku, Polewali.
2	Simboro, Tommo, Papalang, Campalagian, Wonomulyo, Balanipa, Tapango, Banggae Timur, Sendana, Pamboang, Malunda, Mamasa, Pasangkayu, BAMBALAMUTU, Baras, Topoyo, Karossa, Tobadak, Budong-Budong.
3	Tapalang, Kalumpang, Tapalang Barat, Bonehau, Alu, Sumarorong, Mambi, Tabulahan, Bambang, Tandukalua, Rantebulahan Timur,



	Buntumalangka, Aralle, Sesenapadang, Tikke Raya, Dapurang, Sarudu, Sarjo, Lariang, Pangale.
--	---

#### 4. KESIMPULAN

Klaster 1 terdiri atas 3 kecamatan, yaitu Mamuju, Kalukku, Polewali. Klaster 2 terdiri atas 19 kecamatan, yaitu Simboro, Tommo, Papalang, Campalagian, Wonomulyo, Balanipa, Tapango, Banggae Timur, Sendana, Pamboang, Malunda, Mamasa, Pasangkayu, Bambalamutu, Baras, Topoyo, Karossa, Tobadak, Budong-Budong. Klaster 3 terdiri atas 20 kecamatan, yaitu Tapalang, Kalumpang, Tapalang Barat, Bonehau, Alu, Sumarorong, Mambi, Tabulahan, Bambang, Tandukalua, Rantebulahan Timur, Buntumalangka, Aralle, Sesenapadang, Tikke Raya, Dapurang, Sarudu, Sarjo, Lariang, Pangale. Nilai *centroid* tertinggi untuk semua variabel berada pada klaster 1. Ketiga klaster memiliki nilai *centroid* tertinggi pada variabel  $X_3$  yaitu jumlah siswa SD, sedangkan nilai *centroid* terendah pada variabel  $X_8$  yaitu jumlah guru KB

#### DAFTAR PUSTAKA

- Awaliah, R., 2018, Analisis Clustering untuk Mengelompokkan Tingakt Kesejahteraan Kabupaten/ Kota Berdasarkan Sosial Ekonomi Rumah Tangga di Wilayah Provinsi Sulawesi Selatan, *Skripsi*, Program Sarjana Matematika, UIN Alauddin, Makassar.
- Fadhli. 2011, Analisis *Cluster* untuk Pemetaan Mutu Pendidikan di Aceh. Tesis. Program Pasca Sarjana. Universitas Gadjah Mada, Yogyakarta.
- Gudono, 2011, Analisis Data Multivariat Edisi Pertama, Yogyakarta: BPFPE.
- Hasrul, M., 2018, Analisis Kluster untuk Pengelompokan Kabupaten/ Kota Di Provinsi Sulawesi Selatan Berdasarkan Indikator Sosial Kesejahteraan Rakyat, *Skripsi*, Program Sarjana Matematika, UIN Alauddin, Makassar.
- L.S., Dewi, dkk, 2021, Analisis untuk Pengelompokan Kabupaten/ Kota di Provinsi Maluku Berdasarkan Indikator Pendidikan dengan Menggunakan Metode Ward. *Jurnal Statistika dan Aplikasinya*. Vol. 5, Issue 1, hal. 51-60.
- Machfudhoh, S dan Wahyuningsih, N. 2013, Analisis Cluster Kabupaten/Kota Berdasarkan Pertumbuhan ekonomi Jawa Timur, *Jurnal Sains dan Seni Pomits*, Vol.2 No.1. hal 4-15.
- Mattjik, A.A. dan Sumertajaya, I.M., 2011, Sidik Peubah Ganda Dengan Menggunakan SAS. Bogor: IPB Press.
- RI, Kementerian Pendidikan dan Kebudayaan. *dapo.kemdikbud.go.id*. Juli 15, 2021. <https://dapo.kemdikbud.go.id/pd/1/330000> (accessed 2021).
- Rusman. 2019, Kualitas Pendidikan di Sulawesi Barat. *Jurnal Mandar (Management Development and Applied Research)* vol 1, No.2. hal 18-27.
- Supranto, 2004, Analisis Multivariat Arti dan Interpretasi Edisi Pertama, Jakarta: Rineka Cipta.