

**PERBANDINGAN METODE AVERAGE LINKAGE  
DAN K-MEANS DALAM MENGELOMPOKKAN PERSEBARAN PENYAKIT  
MULUT DAN KUKU DI INDONESIA**

**COMPARISON OF AVERAGE LINKAGE  
AND K-MEANS METHODS IN GROUPING THE SPREAD OF  
FOOT AND MOUTH DISEASES IN INDONESIA**

**ANGELINA<sup>1\*</sup>, LISA HARSYIAH<sup>2</sup>, NUR ASMITA PURNAMASARI<sup>3</sup>**

<sup>1</sup> Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mataram.  
Jl. Majapahit 62, Mataram, 83125, Indonesia. email: [angelinayoo07@gmail.com](mailto:angelinayoo07@gmail.com).

<sup>2</sup> Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mataram.  
Jl. Majapahit 62, Mataram, 83125, Indonesia. email: [lisa\\_harsiyah@unram.ac.id](mailto:lisa_harsiyah@unram.ac.id)

<sup>3</sup> Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mataram.  
Jl. Majapahit 62, Mataram, 83125, Indonesia. email: [asmitapurnamasari@unram.ac.id](mailto:asmitapurnamasari@unram.ac.id)

**Abstrak.** Analisis *cluster* merupakan teknik peubah ganda yang memiliki tujuan utama untuk mengelompokkan objek-objek berdasarkan kemiripan karakteristik yang dimilikinya. Tujuan dari penelitian ini adalah untuk menentukan hasil pengelompokan persebaran kasus Penyakit Mulut dan Kuku (PMK) di Indonesia menggunakan metode *average linkage* dan *k-means*. Selain itu, penelitian ini juga bertujuan untuk menentukan metode terbaik dalam mengelompokkan persebaran kasus PMK di Indonesia di antara kedua metode yang digunakan. Data yang digunakan dalam penelitian ini adalah data 27 provinsi yang terdampak di Indonesia yang terdiri dari enam variabel persebaran kasus PMK di Indonesia. Berdasarkan hasil validasi *cluster*, jumlah *cluster* optimal yang terbentuk pada metode *average linkage* sebanyak 4 *cluster*, sedangkan pada metode *k-means* sebanyak 3 *cluster*. Berdasarkan nilai rasio simpangan baku diperoleh hasil bahwa hasil pengelompokan dengan metode *average linkage* lebih baik dibandingkan dengan hasil pengelompokan dengan metode *k-means*. Hal ini karena pada metode *average linkage* diperoleh nilai rasio simpangan baku yang lebih kecil dibandingkan metode *k-means*, yaitu sebesar 0,035, sedangkan pada metode *k-means* diperoleh nilai rasio simpangan baku sebesar 0,258. Oleh karena itu, diperoleh metode *average linkage* lebih baik dibandingkan metode *k-means* dalam mengelompokkan persebaran kasus PMK di Indonesia.

**Kata kunci:** Analisis *Cluster*, *Average Linkage*, *K-Means*, PMK

**Abstract.** Cluster analysis is a multivariate technique the main aim is to classify objects based on similar characteristics. This study aimed to determine the results of clustering the spread of Foot and Mouth Disease (FMD) cases in Indonesia using the average linkage and k-means methods. In addition, this study also aims to determine the best way for classifying the distribution of FMD cases in Indonesia between the two methods used. The data used in this study is data from 27 affected provinces in Indonesia and consist of six variables of the distribution of FMD cases in Indonesia. Based on the results of cluster validation, the optimal number of clusters formed in the average linkage method is 4 clusters, while in the k-means method, there are 3 clusters. Based on the value of the standard deviation ratio, obtained results of grouping with the average linkage method were better than the results of classifying with the k-means method. It is because, in the average linkage method, the standard deviation ratio is smaller than the k-means method, which is 0.035, while the k-means metho's, standard deviation ratio is 0.258. Therefore, we obtained that the average linkage method is better than the k-means method in classifying the distribution of FMD cases in Indonesia.

**Key words:** Average Linkage, Cluster Analysis, FMD, K-Means

## PENDAHULUAN

Menurut Direktorat Kesehatan Hewan (2022), Penyakit Mulut dan Kuku (PMK) atau yang disebut juga *Foot and Mouth Disease* (FMD) adalah penyakit yang disebabkan oleh infeksi virus yang bersifat akut dan sangat menular pada hewan berkuku genap/belah (*cloven-hoofed*). Penyebaran PMK di Indonesia saat ini bukanlah hal yang pertama kali terjadi, karena penyakit ini telah memasuki Indonesia sejak tahun 1887. Kemudian, Indonesia dinyatakan sebagai negara bebas PMK sejak tahun 1986 (Ditkeswan, 2022).

Pada tanggal 27 April 2022, kasus PMK pertama untuk tahun 2022 terjadi di Kabupaten Gresik pada 402 sapi potong (CIVAS, 2022). Kemudian, wabah ini menyebar dan menyerang daerah-daerah lain di Indonesia. Peningkatan kasus PMK di Indonesia saat ini menyebabkan kekhawatiran dalam kalangan masyarakat umum dan kalangan peternak khususnya. Setelah terjadinya kasus PMK di Indonesia, penyakit ini memiliki potensi penyebaran yang cepat dan dapat menjangkit populasi hewan rentan di Indonesia. Berdasarkan hal tersebut, maka saat ini pemerintah Indonesia perlu melakukan berbagai tindakan dan kebijakan untuk menekan penyebaran dan peningkatan kasus PMK di Indonesia.

Selain dengan melakukan vaksinasi terhadap hewan ternak, melaksanakan strategi yang telah direncanakan juga merupakan salah satu cara yang efisien untuk menekan kasus PMK. Strategi utama apabila wabah PMK terjadi di Indonesia seharusnya adalah melakukan *stamping out* dan sistem *zoning* (perwilayahan) sehingga daerah lain yang tidak tertular tetap dipertahankan bebas PMK dan perdagangan di daerah bebas PMK tersebut dapat terus berjalan (Ditkeswan, 2022). Untuk mencegah penyebaran virus ke provinsi lain yang belum terjangkit PMK, maka perlu dilakukan pengelompokan provinsi di Indonesia ke dalam suatu kelompok/*cluster* yang memiliki kemiripan atau karakteristik yang hampir sama.

Metode yang dapat digunakan untuk mengelompokkan provinsi di Indonesia adalah analisis *cluster*. Menurut Hair *et al.* (2006) dalam Widyadhana *et al.* (2021), analisis *cluster* adalah metode multivariat yang dirancang untuk mengklasifikasikan objek berdasarkan karakteristik yang sama. Analisis *cluster* digunakan karena memiliki kelebihan, yaitu (a) dapat mengelompokkan data observasi dalam jumlah besar dan variabel yang relatif banyak, dan (b) dapat dipakai dalam skala ordinal, interval, dan

rasio (Talakua et al., 2017). Selain itu, data yang digunakan dalam penelitian merupakan *unlabeled data*, di mana data tidak memiliki label atau keterangan yang menjadi variabel terikat (*Y*), maka data yang digunakan dalam penelitian ini sesuai untuk dianalisis menggunakan analisis *cluster*.

Pada analisis *cluster* terdapat metode penggerombolan hierarki dan nonhierarki. Pada proses pengelompokan, metode hierarki lebih terstruktur dan bertahap, serta tidak perlu menentukan jumlah *cluster* terlebih dahulu, namun memiliki proses perhitungan yang cukup lambat. Di samping itu, proses pengelompokan pada metode nonhierarki memiliki proses perhitungan yang cepat dan dapat digunakan untuk data yang besar, namun perlu menentukan jumlah *cluster* terlebih dahulu serta sensitif terhadap nilai pusat *cluster* awal. Oleh karena itu, dalam penelitian ini membandingkan metode dalam analisis *cluster* hierarki dan nonhierarki.

Menurut penelitian yang telah dilakukan Septianingsih (2022) diperoleh hasil bahwa metode *average linkage* merupakan metode terbaik dalam melakukan penelitian tersebut. Hal ini didasarkan dari nilai koefisien korelasi *cophenetic* metode *average linkage* yang paling baik dibandingkan dengan metode lainnya. Oleh karena itu, pada analisis *cluster* hierarki digunakan metode *average linkage*.

Menurut penelitian yang telah dilakukan Putri & Dwidayati (2021) diperoleh hasil metode *k-means* lebih baik dibandingkan dengan metode *fuzzy c-means* dalam penelitian tersebut. Hal ini dikarenakan metode *k-means* memperoleh nilai indeks Davies-Bouldin yang lebih mendekati nilai nol. Oleh karena itu, pada analisis *cluster* nonhierarki digunakan metode *k-means*. Berdasarkan uraian tersebut, maka dilakukan penelitian dengan tujuan untuk menentukan hasil pengelompokan persebaran kasus PMK di Indonesia menggunakan analisis *cluster* dengan metode *average linkage* dan metode *k-means*. Selain itu, penelitian ini bertujuan untuk menentukan metode terbaik di antara metode *average linkage* dan metode *k-means* dalam mengelompokkan persebaran kasus PMK di Indonesia.

## **MATERI DAN METODE**

Data yang digunakan adalah data sekunder. Data dalam penelitian ini diperoleh melalui situs resmi Siaga PMK *Crisis Center* (situs: <https://siagapmk.crisis-center.id>) (Kementan RI, 2022). Data berupa berupa informasi mengenai 27 provinsi di Indonesia

yang terdampak PMK pada bulan Maret 2023. Dalam penelitian ini, terdapat enam variabel yang digunakan, yaitu data populasi hewan rentan PMK ( $X_1$ ), data vaksinasi ( $X_2$ ), data sakit ( $X_3$ ), data sembuh ( $X_4$ ), data potong bersyarat ( $X_5$ ), dan data jumlah kematian ( $X_6$ ).

Metode yang digunakan dalam penelitian ini adalah analisis *cluster*, yaitu metode *average linkage* dan *k-means*. Penelitian ini terdiri dari beberapa tahapan yang dijabarkan sebagai berikut:

1. Tahap awal adalah uji asumsi analisis cluster, yang melibatkan pengujian sampel representatif dan uji multikolinearitas.

- a. Uji Kecukupan Sampel menggunakan Uji KMO

Pertama, dilakukan Uji Kecukupan Sampel menggunakan Uji KMO untuk menilai apakah data layak untuk analisis lebih lanjut. dengan rumus sebagai berikut (Abidin, 2017):

$$KMO = \frac{\sum_k \sum_l r_{kl}^2}{\sum_k \sum_l r_{kl}^2 + \sum_k \sum_l a_{kl}^2}, \text{ untuk } k \neq l \text{ dan } k, l = 1, 2, \dots, p \quad (1)$$

di mana:

$p$  : banyaknya variabel

$r_{kl}$ : koefisien korelasi Pearson antara variabel ke- $k$  dan ke- $l$

$a_{kl}$ : koefisien korelasi parsial antara variabel ke- $k$  dan ke- $l$

Sampel dikatakan layak untuk dianalisis jika nilai  $\geq 0,5$ .

- b. Uji Multikolinearitas

Selanjutnya, dilakukan uji multikolinearitas dengan nilai *VIF*. Multikolinearitas adalah hubungan linear yang ada di antara variabel bebas (Nabillah & Nugraha, 2017). Menurut (Sartika W, 2020) Sartika *et al.* (2020) data yang digunakan pada proses *cluster* harus terbebas dari multikolinearitas karena pada analisis *cluster* setiap variabel diberi bobot yang sama dalam perhitungan jarak Euclidean. Untuk melakukan uji multikoliniearitas dapat dilihat dari nilai *Variance Inflation Factor (VIF)* dengan rumus sebagai berikut ((Widarjono, 2010) dalam (Nabillah & Nugraha, 2017)):

$$VIF_k = \frac{1}{(1 - R_k^2)} \quad (2)$$

di mana:

$R_k^2$ : koefisien determinasi antara variabel ke- $k$  dengan variabel lainnya

Jika nilai *VIF* yang diperoleh  $\geq 10$ , maka dapat disimpulkan adanya multikolinearitas (Nabillah & Nugraha, 2017). Jika terdapat multikolinearitas pada data, akan dilanjutkan dengan transformasi menggunakan analisis komponen utama. Namun jika tidak terjadi multikolinearitas, maka tahapan langsung dilanjutkan dengan proses analisis *cluster*.

## 2. Metode Analisis Komponen Utama

Komponen utama merupakan kombinasi linear dari peubah yang diamati, informasi yang terkandung pada komponen utama merupakan gabungan dari semua peubah dengan bobot tertentu (Sandag et al., 2020). Nugroho (2008) menjelaskan tentang sifat komponen utama, antara lain (1) hasil komponen utama yang diperoleh bersifat saling orthogonal dan tidak berkorelasi, (2) sebagian besar keragaman cenderung berkumpul pada komponen utama pertama dan sedikit keragaman dari peubah asal terkumpul pada komponen utama urutan terakhir.

## 3. Metode *Average Linkage*

*Average linkage* merupakan metode pengelompokan yang menggunakan konsep jarak antara dua *cluster* sebagai jarak rata-rata antara semua objek data dalam satu *cluster* dengan seluruh objek pada *cluster* lainnya (Septianingsih, 2022). Tujuan dari metode ini adalah untuk meminimumkan rata-rata jarak semua pasangan pengamatan dan dua *cluster* yang digabungkan serta cenderung membentuk *cluster* dengan ragam kecil (Irwan, 2014). Menurut Supiati (2010), metode ini relatif terbaik di antara metode hierarki lainnya, namun di sisi lain memiliki waktu komputasi yang paling lama.

## 4. Metode *K-Means*

Metode *k-means* merupakan suatu algoritma yang relatif sederhana untuk mengklasifikasikan atau mengelompokkan sejumlah besar objek dengan atribut tertentu ke dalam kelompok-kelompok sebanyak *k* (Muhammad & Akhsani, 2016). Menurut Kasim et al. (2021), metode ini melakukan pengelompokan berdasarkan jarak setiap data dengan titik pusat *cluster* atau *centroid*, sehingga kemiripan data dalam satu *cluster* diminimalkan dan kemiripan data antar *cluster* dimaksimalkan. Tujuan dari pengelompokan metode ini adalah untuk meminimalkan fungsi objektif yang diset dalam proses pengelompokan, di mana pada umumnya berusaha

meminimalkan variasi di dalam suatu *cluster* dan memaksimalkan variasi antar *cluster* (Sangga, 2018).

## 5. Validasi *Cluster*

Validasi *cluster* adalah tahap pengelompokan yang bertujuan untuk mengevaluasi keoptimalan hasil analisis *cluster* secara kumulatif dan objektif berdasarkan metode statistika untuk menghasilkan jumlah *cluster* optimum (Septianingsih, 2022). Terdapat beberapa indeks yang dapat digunakan dalam melakukan validasi *cluster*, di antaranya sebagai berikut:

### a) Indeks Dunn

Septianingsih (2022) mendefinisikan indeks Dunn sebagai rasio jarak terkecil objek pada *cluster* berbeda dengan jarak terbesar antar objek pada *cluster* yang sama. Rentang nilai indeks Dunn, yaitu berada pada nol (0) sampai tak hingga ( $\infty$ ) (Prabowo et al., 2020). Menurut Kovacs *et al.* (2005) dalam Cahyoningtyas (2019) semakin besar nilai indeks Dunn yang diperoleh, maka diindikasikan banyak *cluster* yang terbentuk semakin baik.

### b) Indeks Davies-Bouldin

Prinsip pendekatan pengukuran Indeks Davies-Bouldin adalah memaksimalkan jarak antar *cluster* serta meminimalkan jarak dalam cluster (Gie & Jollyta, 2020). Rentang indeks Davies-Bouldin berada pada nol (0) sampai tak hingga ( $\infty$ ) (Mekonnen et al., 2023) (Mekonnen *et al.*, 2023). Semakin kecil nilai *DBI* yang diperoleh maka semakin baik *cluster* yang terbentuk, namun nilai yang memadai untuk indeks Davies-Bouldin umumnya dianggap di bawah satu (1) (Mekonnen et al., 2023).

### c) Koefisien *Silhouette*

Koefisien *silhouette* mengukur tingkat homogenitas anggota dalam internal *cluster* dan tingkat heterogenitas anggota antar *cluster* (Septianingsih, 2022). Hasil validasi *cluster* yang terbentuk dikatakan baik, jika nilai koefisien *silhouette* mendekati 1 dan kondisi sebaiknya jika nilai yang diperoleh mendekati -1 (Septianingsih, 2022).

## 6. Pemilihan Metode Terbaik

Menurut Cahyoningtyas (2019), hasil analisis *cluster* yang optimal adalah apabila objek dalam satu *cluster* memiliki sifat yang homogen dan sebaliknya,

bersifat heterogen antar *cluster*. Menurut Bunkers *et al.* (1996) dalam Laeli (2014), untuk mengetahui metode yang memiliki kinerja terbaik dapat menggunakan rata-rata simpangan baku dalam *cluster* ( $s_w$ ) dan simpangan baku antar *cluster* ( $s_b$ ).

Rumus simpangan baku dalam cluster dinyatakan dalam persamaan sebagai berikut (Veriani, 2020)(Veriani, 2020):

$$S_w = \frac{1}{c} \sum_{m=1}^c S_m \quad (7)$$

di mana:

$c$  : Banyak cluster yang terbentuk

$S_m$  : Simpangan baku *cluster* ke- $m$ , di mana  $m = 1, 2, \dots, c$

Adapun simpangan baku antar cluster dinyatakan dalam persamaan sebagai berikut (Veriani, 2020):

$$S_b = \sqrt{\frac{1}{c} \sum_{m=1}^c (\bar{x}_m - \bar{X})^2} \quad (8)$$

di mana:

$c$  : Banyak cluster yang terbentuk

$\bar{x}_m$  : Rata-rata *cluster* ke- $m$ , di mana  $m = 1, 2, 3, \dots, c$

$\bar{X}$  : Rata-rata seluruh *cluster*

Nilai rasio simpangan baku dapat ditentukan dengan rumus pada persamaan berikut (Veriani, 2020):

$$\text{rasio simpangan baku} = \frac{S_w}{S_b} \quad (9)$$

Metode yang memiliki nilai rasio simpangan baku terkecil merupakan metode terbaik (Laeli, 2014). Tahap terakhir adalah interpretasi hasil analisis *cluster*, di mana *cluster-cluster* yang terbentuk diberi nama atau keterangan yang tepat. Proses ini memungkinkan pemahaman yang lebih baik terhadap persebaran kasus PMK di Indonesia berdasarkan karakteristik yang diidentifikasi melalui analisis *cluster*.

## HASIL DAN DISKUSI

Penelitian ini menggunakan analisis *cluster* dalam mengelompokkan persebaran kasus Penyakit Mulut dan Kuku (PMK) di Indonesia. Metode analisis *cluster* yang digunakan adalah metode *average linkage* dan *k-means*. Sebelum dilakukan proses

analisis data, data perlu distandarisasi terlebih dahulu karena memiliki satuan yang berbeda-beda. Berikut hasil dan pembahasan dari penelitian mengenai perbandingan metode average linkage dan k-means dalam mengelompokkan persebaran penyakit mulut dan kuku di Indonesia:

### 1. Pengujian Asumsi Analisis *Cluster*

Pada tahap ini, dilakukan pengujian asumsi analisis *cluster* dengan melakukan uji *KMO* dan uji multikolinearitas. Untuk menentukan nilai *KMO* digunakan persamaan (1), dengan hasil sebagai berikut:

$$\begin{aligned}
 KMO &= \frac{\sum_k \sum_l r_{kl}^2}{\sum_k \sum_l r_{kl}^2 + \sum_k \sum_l a_{kl}^2}, \text{ untuk } k \neq l \text{ dan } k, l = 1, 2, \dots, p \\
 &= \frac{0,691^2 + 0,683^2 + 0,640^2 + \dots + 0,893^2}{(0,691^2 + \dots + 0,893^2) + (0,429^2 + 0,676^2 + \dots + 0,738^2)} \\
 &= \frac{8,516}{8,516 + 4,373} \\
 &= 0,661
 \end{aligned}$$

Berdasarkan hasil di atas diperoleh nilai *KMO* sebesar 0,661 yang lebih besar dari 0,5, sehingga dapat disimpulkan bahwa data yang digunakan layak untuk analisis lebih lanjut.

Selanjutnya dilakukan Uji multikolinearitas menggunakan nilai *VIF*. Uji ini digunakan untuk menguji adanya variabel bebas yang memiliki hubungan dengan variabel bebas lainnya. Hasil dari Uji multikolinearitas dapat dilihat pada Tabel 1.

Tabel 1. Hasil Uji Multikolinearitas

Variabel	<i>VIF</i>
$Z_1$	28,25
$Z_2$	10,11
$Z_3$	2679,73
$Z_4$	2316,80
$Z_5$	21,65
$Z_6$	21,46

Berdasarkan tabel 1, variabel-variabel yang digunakan dalam penelitian memiliki nilai *VIF* > 10. Oleh karena itu, dapat disimpulkan terdapat multikolinearitas dalam data yang digunakan, sehingga perlu dilakukan transformasi data. Hal ini juga sesuai dengan nilai koefisien korelasi sederhana dari beberapa variabel yang memiliki korelasi yang kuat.



Transformasi data dilakukan dengan menggunakan metode analisis komponen utama. Pada tahapan ini diperoleh nilai proporsi variasi yang dijelaskan oleh setiap komponen utama seperti pada tabel berikut:

Tabel 2. Nilai Proporsi Variasi

Komponen	Nilai Eigen	Proporsi	Kumulatif
1	4,676	0,779	0,779
2	1,063	0,177	0,957
3	0,176	0,029	0,986
4	0,060	0,010	0,996
5	0,025	0,004	0,99995
6	0,0002	0,00003	1,000

Berdasarkan tabel 2, diperoleh dua komponen utama yang sesuai dengan kriteria, yaitu komponen utama 1 ( $KU_1$ ) dan komponen utama 2 ( $KU_2$ ), di mana nilai eigen kedua komponen utama lebih dari satu. Selain itu, proporsi kumulatif variasi kedua komponen utama sudah lebih dari 80% yakni sebesar 95,7%. Dengan demikian, komponen yang terbentuk mampu menjelaskan sebesar 95,7% dari keragaman total item-item penelitian.

Untuk memudahkan penentuan setiap variabel masuk ke dalam komponen utama yang terbentuk, dapat dilakukan dengan menentukan matriks rotasi *factor loading*. Hasil rotasi ditampilkan pada Tabel 3 sebagai berikut:

Tabel 3. Rotasi *Factor Loading*

Variabel	Komponen	
	1	2
$Z_1$	0,442	0,864
$Z_2$	0,909	0,291
$Z_3$	0,929	0,337
$Z_4$	0,942	0,287
$Z_5$	0,150	0,979
$Z_6$	0,502	0,840

Tabel 3 menunjukkan perbedaan nilai *loading* yang lebih signifikan antara komponen satu dengan yang lain. Nilai *loading* yang tinggi hanya terdapat pada satu komponen saja, sehingga interpretasi lebih mudah dilakukan dan keanggotaan setiap variabel dapat ditentukan dengan jelas masuk pada komponen tertentu. Berdasarkan tabel 3 dapat disimpulkan bahwa variabel-variabel yang masuk ke dalam  $KU_1$  adalah variabel  $Z_2$ , variabel  $Z_3$ , dan variabel  $Z_4$  berupa variabel vaksinasi, sakit, sembuh. Untuk variabel-variabel yang masuk ke dalam  $KU_2$  adalah

variabel  $Z_1$ , variabel  $Z_5$ , dan variabel  $Z_6$  berupa variabel populasi, potong bersyarat, dan kematian.

Selanjutnya perlu ditentukan nilai *factor scores* karena analisis dilanjutkan dengan analisis *cluster*. *Factor scores* dinotasikan sebagai matriks ( $f$ ) dan diperoleh matriks *factor scores* sebagai berikut:

$$f = \begin{bmatrix} 0,127 & -0,358 & 0,001 & \dots & 0,427 \\ -0,232 & -0,257 & -0,328 & \dots & -0,422 \end{bmatrix}_{(2 \times 27)}$$

Setelah melakukan transformasi data dengan analisis komponen utama, pengolahan data dilanjutkan dengan proses analisis *cluster*.

## 2. Hasil pengelompokan persebaran kasus PMK di Indonesia dengan metode *average linkage*

Metode *average linkage* termasuk ke dalam metode analisis *cluster* hierarki aglomeratif karena dilakukan secara bertingkat dan bertahap. Dalam analisis ini, ukuran jarak yang digunakan adalah jarak Euclidean. Sebelum dilanjutkan dengan proses pengelompokan, jumlah *cluster* yang ditetapkan dalam analisis ini yaitu 2 hingga 5 *cluster*. Hal ini dikarenakan apabila jumlah *cluster* yang terbentuk terlalu banyak, akan menyulitkan interpretasi *cluster* yang terbentuk (Mubarak et al., 2015). Kemudian ditentukan jumlah *cluster* optimal dengan melakukan validasi *cluster*. Adapun proses validasi *cluster* dilakukan dengan menggunakan nilai indeks Dunn, indeks Davies-Bouldin, dan koefisien *silhouette* dan diperoleh hasil yang disajikan dalam Tabel 4 berikut:

Tabel 4 Hasil Validasi *Cluster* Metode *Average Linkage*

Jumlah <i>Cluster</i>	Indeks Dunn ( <i>DI</i> )	Indeks Davies-Bouldin ( <i>DBI</i> )	Koefisien <i>Silhouette</i> ( <i>SC</i> )
2	0,93	0,22	0,80
3	0,84	0,40	0,81
4	1,41	0,10	0,77
5	1,13	0,11	0,66

Berdasarkan tabel 4 dapat diketahui bahwa berdasarkan indeks Dunn jumlah *cluster* optimal yang terbentuk adalah 4 *cluster* karena memiliki nilai *DI* terbesar, yakni sebesar 1,41. Berdasarkan indeks Davies-Bouldin, jumlah *cluster* optimal yang terbentuk adalah sebanyak 4 *cluster* karena memiliki nilai *DBI* terkecil, yaitu sebesar 0,10. Berdasarkan koefisien *silhouette*, jumlah *cluster* optimal yang terbentuk adalah sebanyak 3 *cluster* dengan nilai *SC* sebesar 0,81. Hal ini

dikarenakan pada jumlah *cluster* sebanyak 3 *cluster* terbentuk memiliki nilai *SC* paling mendekati 1, sehingga dapat disimpulkan struktur yang dihasilkan sangat kuat.

Setelah diketahui hasil dari ketiga indeks, selanjutnya dilanjutkan dengan menentukan indeks gabungan dikarenakan terdapat perbedaan hasil keputusan dari ketiga indeks. Indeks gabungan adalah suatu cara dalam menentukan jumlah *cluster* optimal dengan melakukan perhitungan jumlah peringkat dari masing-masing indeks yang digunakan sesuai kriterianya dan pemberian peringkat dimulai dari 1, 2, dan seterusnya (Cahyoningtyas, 2019). Berikut merupakan hasil perhitungan indeks gabungan untuk metode *average linkage* yang disajikan dalam Tabel 5.

Tabel 5. Indeks Gabungan Metode *Average Linkage*

Jumlah Cluster	DI	Rank DI	DBI	Rank DBI	SC	Rank SC	Indeks Gabungan
2	0,93	3	0,22	3	0,80	2	8
3	0,84	4	0,40	4	0,81	1	9
4	1,41	1	0,10	1	0,77	3	5
5	1,13	2	0,11	2	0,66	4	8

Berdasarkan hasil indeks gabungan pada tabel 5, diketahui bahwa jumlah *cluster* optimal yang terbentuk adalah sebanyak 4 *cluster*. Hal ini dikarenakan pada jumlah *cluster* sebanyak 4 *cluster* terbentuk memiliki nilai indeks gabungan terkecil. Oleh karena itu, diperoleh hasil jumlah *cluster* optimal yang terbentuk adalah sebanyak 4 *cluster* dalam mengelompokkan persebaran kasus PMK di Indonesia.

Adapun proses pengelompokan metode *average linkage* dimulai dengan menentukan jarak terdekat antar seluruh objek. Pada proses pengelompokan metode *average linkage* terdapat 27 *cluster* awal sebanyak objek yang dikelompokkan. Setelah proses aglomerasi dilakukan dapat diketahui setiap provinsi masuk ke dalam *cluster* mana saja. Hasil pengelompokan menggunakan metode *average linkage* dengan jumlah *cluster* yang terbentuk sebanyak 4 *cluster* ditunjukkan pada Tabel 6.

Tabel 6. Hasil Pengelompokan Metode *Average Linkage*

C	Jumlah Anggota	Provinsi
1	24	Sumatera Utara, Sumatera Selatan, Sumatera Barat, Sumatera Tenggara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Barat, Riau, Lampung, Kepulauan Riau, Kalimantan Utara, Kalimantan Timur,

		Kalimantan Tengah, Kalimantan Selatan, Kalimantan Barat, Jawa Tengah, Jambi, DKI Jakarta, DIY, Bengkulu, Banten, Bangka Belitung, Bali, dan Aceh.
2	1	Nusa Tenggara Barat
3	1	Jawa Timur
4	1	Jawa Barat

### 3. Hasil pengelompokan persebaran kasus PMK di Indonesia dengan metode *k-means*

Metode *k-means* termasuk ke dalam analisis cluster nonhierarki, sehingga dalam metode ini jumlah cluster harus ditentukan terlebih dahulu. Dalam analisis ini, jumlah *cluster* yang digunakan, yaitu 2 hingga 5 *cluster*. Hal ini dikarenakan apabila jumlah *cluster* yang terbentuk terlalu banyak, akan menyulitkan interpretasi *cluster* yang terbentuk (Mubarak, Murtiadi, & Sulistiyono, 2015). Adapun perolehan nilai indeks Dunn, indeks Davies-Bouldin, dan koefisien *silhouette* untuk metode *k-means* disajikan dalam Tabel 7 sebagai berikut:

Tabel 7. Hasil Validasi *Cluster* Metode *K-Means*

Jumlah <i>Cluster</i>	Indeks Dunn ( <i>DI</i> )	Indeks Davies-Bouldin ( <i>DBI</i> )	Koefisien <i>Silhouette</i> ( <i>SC</i> )
2	0,45	1,02	0,79
3	0,84	0,40	0,81
4	0,06	0,66	0,58
5	0,07	0,52	0,52

Berdasarkan tabel 7, dapat diketahui bahwa berdasarkan indeks Dunn jumlah *cluster* optimal yang terbentuk adalah 3 *cluster* karena memiliki nilai *DI* terbesar, yakni sebesar 0,84. Berdasarkan indeks Davies-Bouldin, jumlah *cluster* optimal yang terbentuk adalah sebanyak 3 *cluster* karena memiliki nilai *DBI* terkecil, yaitu sebesar 0,40. Berdasarkan koefisien *silhouette*, jumlah *cluster* optimal yang terbentuk adalah sebanyak 3 *cluster* dengan nilai *SC* sebesar 0,81. Hal ini dikarenakan pada jumlah *cluster* sebanyak 3 *cluster* terbentuk memiliki nilai *SC* paling mendekati 1, sehingga dapat disimpulkan struktur yang dihasilkan sangat kuat. Oleh karena itu, berdasarkan ketiga indeks, diperoleh jumlah *cluster* optimal yang terbentuk dalam mengelompokkan persebaran kasus PMK di Indonesia adalah 3 *cluster*.

Metode *k-means* mengelompokkan objek ke dalam *k cluster* yang memiliki karakteristik yang mirip. Pada pengelompokan dengan metode *k-means* ditentukan

titik pusat *cluster/centroid* awal. *Centroid* awal dalam penelitian ini, yaitu  $C_1$  adalah Provinsi Sumatera Utara,  $C_2$  adalah Provinsi Jawa Barat, dan  $C_3$  adalah Provinsi Aceh. Adapun hasil pengelompokan persebaran kasus PMK di Indonesia dengan metode *k-means* dengan jumlah *cluster* sebanyak 3 *cluster* adalah sebagaimana Tabel 8 berikut:

Tabel 8. Hasil Pengelompokan Metode *K-Means*

$C$	Jumlah Anggota	Provinsi
1	24	Sumatera Utara, Sumatera Selatan, Sumatera Barat, Sumatera Tenggara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Barat, Riau, Lampung, Kepulauan Riau, Kalimantan Utara, Kalimantan Timur, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Barat, Jawa Tengah, Jambi, DKI Jakarta, DIY, Bengkulu, Banten, Bangka Belitung, Bali, dan Aceh.
2	1	Jawa Barat
3	1	Nusa Tenggara Barat dan Jawa Timur

#### 4. Pemilihan Metode Terbaik

Pemilihan metode terbaik berguna untuk mengetahui seberapa baik kinerja metode *average linkage* dan metode *k-means* dalam mengelompokkan persebaran kasus PMK di Indonesia. Suatu *cluster* dikatakan baik apabila memiliki kehomogenan tinggi dalam *cluster* dan keheterogenan yang tinggi antar *cluster*. Kehomogenan dalam *cluster* dapat diketahui dengan menghitung nilai simpangan baku dalam *cluster* ( $S_w$ ), sedangkan keheterogenan antar *cluster* dapat diketahui dengan menghitung nilai simpangan baku antar *cluster* ( $S_b$ ). Kemudian diperoleh nilai rasio simpangan baku untuk masing-masing hasil pengelompokan yang dirangkum dalam Tabel 9 sebagai berikut:

Tabel 9. Nilai Rasio Simpangan Baku

Metode	$S_w$	$S_b$	Rasio Simpangan Baku
<i>Average Linkage</i>	0,057	1,635	0,035
<i>K-Means</i>	0,378	1,466	0,258

Metode terbaik dalam mengelompokkan persebaran kasus PMK di Indonesia memiliki hasil rasio simpangan baku yang minimum. Berdasarkan tabel 9, didapatkan bahwa metode terbaik dalam mengelompokkan persebaran kasus PMK di Indonesia adalah metode *average linkage* dengan nilai rasio simpangan baku terkecil.

Jumlah *cluster* optimal pada metode *average linkage* diperoleh hasil 4 *cluster* terbentuk. Setiap *cluster* memiliki karakteristik yang berbeda berdasarkan keenam indikator yang mengukur persebaran kasus PMK di Indonesia, dimana keenam indikator ini direduksi dalam dua komponen utama, yaitu  $KU_1$  dan  $KU_2$ . Karakteristik persebaran PMK di Indonesia pada setiap *cluster* dapat diketahui melalui nilai rata-rata masing-masing komponen utama. Nilai rata-rata komponen utama setiap *cluster* pada metode *average linkage* adalah sebagai berikut:

Tabel 10. Nilai Rata-Rata Komponen Utama pada Metode *Average Linkage*

<i>Cluster</i>	Banyak Anggota	$KU_1$	$KU_2$
<i>Cluster 1</i>	24	-0,238	-0,193
<i>Cluster 2</i>	1	2,529	-1,118
<i>Cluster 3</i>	1	3,926	1,113
<i>Cluster 4</i>	1	-0,743	4,643

Berdasarkan tabel 10, dapat dilakukan interpretasi untuk hasil pengelompokan persebaran kasus PMK di Indonesia dengan menggunakan metode *average linkage*. *Cluster 1* beranggotakan 24 provinsi, yaitu Sumatera Utara, Sumatera Selatan, Sumatera Barat, Sumatera Tenggara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Barat, Riau, Lampung, Kepulauan Riau, Kalimantan Utara, Kalimantan Timur, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Barat, Jawa Tengah, Jambi, DKI Jakarta, DIY, Bengkulu, Banten, Bangka Belitung, Bali, dan Aceh. *Cluster* ini memiliki karakteristik provinsi yang memiliki nilai kedua paling rendah untuk  $KU_1$  dan nilai kedua paling rendah untuk  $KU_2$  di antara *cluster* lainnya. Oleh karena itu, *cluster 1* memiliki tingkat vaksinasi, jumlah hewan sakit, dan jumlah hewan sembuh yang sedang, dimana ketiga variabel ini termuat dalam  $KU_1$ . Selain itu, *cluster 1* memiliki tingkat populasi, jumlah hewan potong bersyarat, dan jumlah kematian yang sedang, dimana ketiga variabel ini termuat pada  $KU_2$ .

*Cluster 2* beranggotakan 1 provinsi, yaitu Provinsi Nusa Tenggara Barat. *Cluster* ini memiliki karakteristik provinsi yang memiliki nilai kedua paling tinggi untuk  $KU_1$  dan nilai paling rendah untuk  $KU_2$  di antara *cluster* lainnya. Oleh karena itu, *cluster 2* memiliki tingkat vaksinasi, jumlah hewan sakit, dan jumlah hewan sembuh yang tinggi dimana ketiga variabel ini termuat dalam  $KU_1$ . Selain itu, *cluster 2* memiliki tingkat populasi, jumlah hewan potong bersyarat, dan jumlah kematian yang rendah, dimana ketiga variabel ini termuat pada  $KU_2$ .

*Cluster 3* beranggotakan 1 provinsi, yaitu Provinsi Timur. *Cluster* ini memiliki karakteristik provinsi yang memiliki nilai paling tinggi untuk  $KU_1$  dan memiliki nilai kedua paling tinggi untuk  $KU_2$  di antara *cluster* lainnya. Oleh karena itu, *cluster 3* memiliki tingkat vaksinasi, jumlah hewan sakit, dan jumlah hewan sembuh sangat tinggi dimana ketiga variabel ini termuat dalam  $KU_1$ . Selain itu, *cluster 3* memiliki tingkat populasi, jumlah hewan potong bersyarat, dan jumlah kematian yang tinggi, dimana ketiga variabel ini termuat pada  $KU_2$ .

*Cluster 4* beranggotakan 1 provinsi, yaitu Provinsi Jawa Barat. *Cluster* ini memiliki karakteristik provinsi yang memiliki nilai paling rendah untuk  $KU_1$  dan memiliki nilai paling tinggi untuk  $KU_2$  di antara *cluster* lainnya. Oleh karena itu, *cluster 4* memiliki tingkat vaksinasi, jumlah hewan sakit, dan jumlah hewan sembuh yang rendah dimana ketiga variabel ini termuat dalam  $KU_1$ . Selain itu, *cluster 4* memiliki tingkat populasi, jumlah hewan potong bersyarat, dan jumlah kematian yang sangat tinggi, dimana ketiga variabel ini termuat pada  $KU_2$ .

## KESIMPULAN

Berdasarkan hasil analisis yang telah dilakukan, dapat diambil kesimpulan dalam penelitian ini adalah sebagai berikut: (1) Hasil pengelompokan persebaran kasus PMK di Indonesia dengan metode *average linkage* diperoleh 4 *cluster*. *Cluster 1* beranggotakan 24 provinsi., *cluster 2* beranggotakan 1 provinsi, *cluster 3* beranggotakan 1 provinsi, dan *cluster 4* beranggotakan 1 provinsi. (2) Hasil pengelompokan persebaran kasus PMK di Indonesia dengan metode *k-means* diperoleh 3 *cluster*. *Cluster 1* beranggotakan 24 provinsi, *cluster 2* beranggotakan 1 provinsi, dan *cluster 3* beranggotakan 2 provinsi. (3) Metode terbaik dalam mengelompokkan persebaran kasus PMK di Indonesia adalah metode *average linkage* dengan nilai rasio simpangan baku sebesar 0,035, sedangkan nilai rasio simpangan baku metode *k-means* adalah sebesar 0,258.

## UCAPAN TERIMA KASIH

Peneliti mengucapkan terima kasih kepada Lisa Hasyiah, S.Pd., M.Si. dan Nur Asmita Purnamasari, S.Si., M.Si. atas arahan serta masukannya, sehingga proses penelitian bisa diselesaikan dan dapat dipublikasikan.

## DAFTAR PUSTAKA

- Abidin, Z. (2017). *Pengelompokan Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Kemiskinan dengan Menggunakan Analisis Cluster Hierarki*. Institut Teknologi Sepuluh Nopember.
- Cahyoningtyas, R. A. (2019). *Metode Ward dan Average Linkage Clustering untuk Segmentasi Objek Wisata di Malag Raya*. Universitas Brawijaya.
- Center for Indonesian Veterinary Analytical Studies. (2022, May 9). *PMK Datang Kembali*. Retrieved October 19, 2022. Civas.Net: <https://Civas.Net/2022/05/09/Pmk-Datang-Kembali/>.
- Direktorat Kesehatan Hewan. (2022). *Pedoman Kesiagaan Darurat Veteriner Indonesia, Seri: Penyakit Mulut dan Kuku*. Kementerian Pertanian.
- Gie, W., & Jollyta, D. (2020). Perbandingan Euclidean dan Manhattan Untuk Optimasi Cluster Menggunakan *Davies Bouldin Index*: Status Covid-19 Wilayah Riau. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 2, 187–191. <http://dx.doi.org/10.30645/senaris.v2i0.160>
- Govender, P., & Sivakumar, V. (2020). *Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019)*. *Atmospheric Pollution Research*, 11(1), 40–56. <https://doi.org/10.1016/j.apr.2019.09.009>
- Irwan. (2014). Aplikasi Analisis Cluster dalam Mengelompokkan Kecamatan Berdasarkan Faktor Penyebab Penduduk Buta Aksara di Kabupaten Jeneponto. *Jurnal Teknosains*, 8(3), 273–289. <https://doi.org/10.24252/teknosains.v8i3.1833>
- Kasim, R. J., Bahri, S., & Amir, S. (2021). Implementasi Metode *K-Means* Untuk Clustering Data Penduduk Miskin Dengan *Systematic Random Sampling*. *Prosiding SISFOTEK*, 5(1), 95–101. <https://seminar.iaii.or.id/index.php/SISFOTEK/article/view/265>
- Kementerian Pertanian Indonesia. (2022, September 28). *Siaga PMK*. Kementerian Pertanian Republik Indonesia: <https://Siagapmk.Crisis-Center.Id>.
- Laeli S. (2014). *Analisis Cluster dengan Average Linkage Method dan Ward's Method untuk Data Respon*. Universitas Negeri Yogyakarta.
- Mekonnen, A. A., Sipos, T., & Krizsik, N. (2023). *Identifying Hazardous Crash Locations Using Empirical Bayes and Spatial Autocorrelation*. *ISPRS International Journal of Geo-Information*, 12(3), 85. <https://doi.org/10.3390/ijgi12030085>
- Mubarak, R., Murtiadi, S., & Sulistiyono, H. (2015). Pengembangan Kriteria Analisis Risiko Bagi Developer Perumahan Di Kota Mataram. *Jurnal Sains Teknologi & Lingkungan*, 1(1). <https://doi.org/10.29303/jstl.v1i1.9>
- Muhammad, M., & Akhsani, L. (2016). Kemampuan Komunikasi Matematis Dengan Metode *K-Means Clustering* Melalui Model *Problem Based Learning*. *PYTHAGORAS: Jurnal Program Studi Pendidikan Matematika*, 5(2). <https://doi.org/10.33373/pythagoras.v5i2.464>
- Nabillah, I. P. W., & Nugraha, J. (2017). Analisis Cluster Tingkat Kualitas Udara Ambien Jalan Raya di DIY 2015. *Prosiding SI MaNIS (Seminar Nasional Integrasi Matematika Dan Nilai-Nilai Islami)*, 1(1), 178–187. <http://conferences.uin-malang.ac.id/index.php/SIMANIS/article/view/60>
- Nugroho, S. (2008). *Statistika Multivariat Terapan*. In *Bengkulu: UNIB Press Bengkulu*. UNIB Press.
- Prabowo, R. A., Nisa, K., Faisol, A., & Setiawan, E. (2020). Simulasi Pemilihan Metode Analisis Cluster Hierarki *Agglomerative* Terbaik antara *Average Linkage* dan *Ward* pada Data yang Mengandung Multikolinearitas. *Jurnal Siger Matematika*, 1(2), 49–55. <https://doi.org/10.23960/jsm.v1i2.2497>
- Putri, A. L. R., & Dwidayati, N. (2021). Analisa perbandingan *k-means* dan *fuzzy c-means* dalam pengelompokan daerah penyebaran COVID-19 Indonesia. *Unnes Journal of Mathematics*, 50–55. <http://doi.org/10.15294/UJM.V10I2.50433>
- Sandag, A. M. V., Mongi, C. E., & Mananohas, M. L. (2020). Pengelompokan Sekolah Menengah Atas (SMA) di Kabupaten Minahasa Tenggara Berdasarkan Standar Kompetensi Lulusan Tahun 2018 Menggunakan Analisis Gerombol. *D'CARTESIAN: Jurnal Matematika Dan Aplikasi*, 9(2), 113–119. <https://doi.org/10.35799/dc.9.2.2020.28755>
- Sangga, V. A. P. (2018). *Perbandingan algoritma K-Means dan algoritma K-Medoids dalam pengelompokan komoditas peternakan di provinsi Jawa Tengah tahun 2015*. Universitas Islam Indonesia.
- Sartika W. (2020). *Sistem Informasi untuk Mengevaluasi Hasil Program dari Intervensi Gizi Spesifik pada Kasus Stunting dengan Metode K-Means Clustering*. Universitas Diponegoro.
- Septianingsih, A. (2022). Pemetaan Kabupaten Kota Di Provinsi Jawa Timur Berdasarkan Tingkat Kasus Penyakit Menggunakan Pendekatan *Agglomeratif Hierarchical Clustering*. *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, 3(2), 367–386. <https://doi.org/10.46306/lb.v3i2.139>



- Supiati. (2010). *Aplikasi Analisis Cluster dalam Mengelompokkan Kecamatan Berdasarkan Faktor Penyebab Penduduk Buta Aksara di Kabupaten Jeneponto*. Universitas Islam Negeri Alauddin.
- Talakua, M. W., Leleury, Z. A., & Taluta, A. W. (2017). Analisis *cluster* dengan menggunakan metode *k-means* untuk pengelompokkan Kabupaten/Kota di provinsi maluku berdasarkan indikator indeks pembangunan manusia tahun 2014. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 11(2), 119–128. <https://doi.org/10.30598/barekengvol11iss2pp119-128>
- Ulinuha, N., & Sholihah, S. A. (2021). Analisis *Cluster* Untuk Pemetaan Data Kasus *Covid-19* Di Indonesia Menggunakan *K-Means*. *Jurnal MSA (Matematika Dan Statistika Serta Aplikasinya)*, 9(2), 27–31. <https://doi.org/10.24252/msa.v9i2.19478>
- Veriani, R. (2020). *Analisis Cluster Dalam Pengelompokkan Provinsi di Indonesia Berdasarkan Variabel Penyakit Menular Menggunakan Metode Complete Linkage, Average Linkage Dan Ward*. Universitas Islam Negeri Sunan Ampel.
- Widyadhana, D., Hastuti, R. B., Kharisudin, I., & Fauzi, F. (2021). Perbandingan analisis klaster *k-means* dan *average linkage* untuk pengklasteran kemiskinan di provinsi jawa tengah. *PRISMA, Prosiding Seminar Nasional Matematika*, 4, 584–594. <https://journal.unnes.ac.id/sju/index.php/prisma/article/view/45032>