Jurnal Fraction, Vol. 4 No.2, Desember 2024, Hal. 49-57
E-ISSN: **2830-2028**; P-ISSN: **2962-1402**
Dipublikasi: Desember 2024

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 Disease in Indonesia)

Angelina[1], Lisa Harsyiah[2], Nur Asmita Purnamasari[3, a]
[1]Universitas Mataram [Email: angelinayoo07@gmail.com]
[2]Universitas Mataram [Email:lisa_harsyiah@unram.ac.id]
[3]Universitas Mataram [Email: asmitapurnamasari@unram.ac.id]
[a]*asmitapurnamasari@unram.ac.id*

ABSTRAK

Tujuan dari penelitian ini adalah untuk menentukan hasil pengelompokan persebaran Penyakit Mulut dan Kuku (PMK) di Indonesia dengan menggunakan dua metode berbeda yaitu *average linkage* dan *k-means*. Penelitian ini bertujuan untuk mengetahui metode terbaik dalam mengelompokkan persebaran PMK di Indonesia. Selain itu, penelitian ini juga bertujuan untuk menentukan metode terbaik dalam mengelompokkan persebaran PMK di Indonesia di antara kedua metode yang digunakan. Berdasarkan hasil validasi *cluster*, jumlah *cluster* optimal yang terbentuk pada metode *average linkage* sebanyak 4 *cluster*, sedangkan pada metode *k-means* sebanyak 3 *cluster*. Hasil pengelompokan dengan metode *average linkage* lebih baik dibandingkan dengan hasil pengklasifikasian dengan metode *k-means*, karena nilai rasio simpangan baku pada metode *average linkage* lebih kecil yaitu sebesar 0,035 dibandingkan dengan metode *k-means* sebesar 0,258. Oleh karena itu, disimpulkan bahwa metode *average linkage* lebih baik dibandingkan metode *k-means* dalam mengelompokkan persebaran PMK di Indonesia.

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

ABSTRACT

The purpose of this study was to analyze the spread of Foot and Mouth Disease (FMD in Indonesia by using two different methods: average linkage and k-means. In addition, this study also aimed to determine the most effective method of classifying the distribution of FMD in Indonesia between the two methods used. The results of cluster validation showed that the optimal number of clusters formed in the average linkage method was 4, while in the k-means method, there were 3 clusters. The grouping with the average linkage method was better than the results of classifying with the k-means method, as the standard deviation ratio in the average linkage method was smaller at 0,035, compared to 0,258 in the k-means method. Therefore, it was concluded that the average linkage method was better than the k-means method in classifying the distribution of FMD in Indonesia.

Keywords: *Average Linkage, Cluster Analysis, FMD, K-Means*

1. PENDAHULUAN

Penyakit Mulut dan Kuku (PMK) adalah suatu penyakit yang disebabkan oleh infeksi virus yang bersifat sangat menular dan akut pada hewan berkuku belah/genap [1]. Pada tanggal 27 April 2022, kasus PMK pertama untuk tahun 2022 terjadi di Kabupaten Gresik pada 402 sapi potong [2]. Kasus ini merupakan kasus pertama sejak Indonesia dinyatakan sebagai negara bebas PMK pada tahun 1986. Penyakit ini memiliki potensi penyebaran yang cepat dan dapat menjangkit populasi hewan rentan di Indonesia. Selain dengan melakukan vaksinasi terhadap hewan ternak, strategi utama yang seharusnya dilakukan adalah dengan menerapkan sistem wilayah (*zoning*) dan *stamping out*, sehingga daerah yang tidak terinfeksi tetap dapat melakukan perdagangan dan dapat dipertahankan bebas PMK [1]. Oleh karena itu, perlu dilakukan pengelompokan provinsi di Indonesia ke dalam suatu kelompok/*cluster* yang memiliki karakteristik yang mirip. Hal ini bertujuan untuk memberi informasi yang berkaitan dengan penyebaran PMK pada provinsi-provinsi di Indonesia.

Metode yang dapat digunakan untuk mengelompokkan provinsi di Indonesia adalah analisis *cluster*. Analisis *cluster* merupakan teknik multivariat yang memiliki tujuan utama untuk mengelompokkan objek-objek berdasarkan kemiripan karakteristik yang dimilikinya [3]. Data dalam penelitian ini merupakan *unlabed data*, di mana data tidak memiliki label atau keterangan yang menjadi variabel terikat (Y). Oleh karena itu, data yang digunakan sesuai untuk dianalisis menggunakan analisis *cluster*.

Pada analisis *cluster* terdapat metode pengelompokan nonhierarki dan hierarki. Proses pengelompokan metode hierarki lebih terstruktur dan bertahap, serta tidak perlu dilakukan penentuan jumlah *cluster* terlebih dahulu, namun memiliki proses perhitungan yang cukup lambat. Di samping itu, proses pengelompokan metode nonhierarki memiliki perhitungan yang cepat dan dapat digunakan untuk data yang besar, namum perlu menentukan jumlah *cluster* terlebih dahulu serta sensitif terhadap nilai pusat *cluster* awal. Berdasarkan hal tersebut, maka dalam penelitian ini dilakukan perbandingan metode dalam analisis *cluster* hierarki dan nonhierarki.

Menurut penelitian Septianingsih yang menggunakan metode *agglomerative hierarchical clustering* dalam memetakan kabupaten/kota di Provinsi Jawa Timur berdasarkan tingkat kasus penyakit, diperoleh hasil bahwa metode a*verage linkage* adalah metode terbaik dalam melakukan penelitian tersebut berdasarkan nilai koefisien korelasi *cophenetic* tertinggi [4]. Oleh karena itu, pada analisis *cluster* hierarki digunakan metode *average linkage.* Adapun menurut penelitian Putri & Dwidayati yang membandingkan metode *fuzzy c-means* dan *k-means* untuk mengelompokkan daerah penyebaran *Covid-19* Indonesia, diperoleh hasil kinerja metode *k-means* lebih baik dibanding metode *fuzzy c-means* [5]. Hal ini dikarenakan nilai indeks Davies-Bouldin metode *k-means* yang lebih mendekati nilai nol. Oleh karena itu, pada analisis *cluster* nonhierarki digunakan metode *k-means*. Berdasarkan uraian tersebut, maka dilakukan penelitian dengan tujuan, yaitu: (1) Menentukan hasil pengelompokan persebaran PMK di Indonesia menggunakan metode *average linkage*, (2) Menentukan hasil pengelompokan persebaran PMK di Indonesia menggunakan metode *k-means*, dan (3) Menentukan metode terbaik di antara metode *average linkage* dan *metode k-means* dalam mengelompokkan persebaran PMK di Indonesia.

2. METODE PENELITIAN

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh melalui situs resmi Siaga PMK *Crisis Center* (situs: https://siagapmk.crisis-center.id). Data berupa informasi mengenai 27 provinsi di Indonesia yang terdampak PMK pada bulan Maret 2023. Penelitian ini menggunakan enam variabel, 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). Penentuan variabel didasarkan pada penelitian terdahulu yang dilakukan oleh Putri & Dwidayanti (2021) dan Septianingsih (2022), yang menggunakan variabel terkait dalam analisis persebaran penyakit dan zonasi daerah yang terdampak. Penelitian ini terdiri dari beberapa tahapan yang dijabarkan sebagai berikut:

2.1. Uji Asumsi Analisis *Cluster*

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. Sampel dikatakan layak untuk dianalisis jika nilai KMO $\geq$ 0,5, dengan rumus sebagai berikut [6]:

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

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

b. Uji Multikolinearitas

Selanjutnya, dilakukan uji multikolinearitas dengan nilai VIF. Multikolinearitas merupakan hubungan linear yang terdapat di antara variabel bebas [6]. Menurut Widarjono dalam Nabillah dan Nugraha, untuk melakukan uji multikoliniearitas dapat dilihat dari nilai *Variance Inflation Factor* (VIF) dengan rumus sebagai berikut [6]:

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

di mana, ${R_k}^2$ merupakan koefisien determinasi antara variabel ke-k dengan variabel lainnya. Apabila nilai VIF yang diperoleh $\geq$ 10, maka dapat disimpulkan adanya multikolinearitas [6]. Jika

terdapat multikolinearitas pada data, tahapan akan dilanjutkan dengan transformasi data menggunakan analisis komponen utama.

2.2. Metode Analisis Komponen Utama

Komponen utama merupakan kombinasi linear dari variabel yang diamati, informasi yang terkandung pada komponen utama merupakan gabungan dari semua variabel dengan bobot tertentu [3]. Menurut Haumahu & Lewaherilla, kriteria dalam pemilihan komponen utama pada analisis komponen utama adalah proporsi kumulatif keragaman data asal yang dijelaskan oleh sebanyak m komponen utama lebih dari 80% [7]. Selain itu dapat juga dilakukan dengan memilih komponen utama yang memiliki nilai eigen yang lebih dari satu [7].

2.3. Metode *Average Linkage*

Metode *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 [4]. Metode ini bertujuan untuk meminimalkan rataan jarak seluruh pasangan observasi dan dua *cluster* yang digabungkan, serta *cluster* yang terbentuk cenderung memiliki ragam yang kecil [8]. Menurut Supiati, metode ini relatif terbaik di antara metode hierarki lainnya, namun di sisi lain memiliki waktu komputasi yang paling lama [9].

2.4. Metode *K-Means*

Metode *k-means* merupakan salah satau metode analisis *cluster* nonhierarki yang mengelompokkan n buah objek ke dalam k kelompok/*cluster* berdasarkan jaraknya dengan pusat *cluster* [5]. 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* [10].

2.5. Validasi *Cluster*

Validasi *cluster* merupakan langkah pengelompokan yang memiliki tujuan untuk memeriksa keoptimalan hasil analisis *cluster* secara kumulatif dan objektif berdasarkan metode statistika guna menghasilkan jumlah *cluster* optimum [4]. Validasi *cluster* dapat dilakukan dengan menggunakan beberapa indeks, seperti:

a. Indeks Dunn

Septianingsih mendefinisikan indeks Dunn sebagai rasio jarak terkecil objek pada *cluster* berbeda dengan jarak terbesar antar objek pada *cluster* yang sama [4]. Kovacs dkk. dalam Cahyoningtyas menjelaskan bahwa, semakin besar nilai indeks Dunn yang diperoleh, maka diindikasikan banyak *cluster* yang terbentuk semakin baik [11].

b. Indeks Davies-Bouldin

Prinsip pendekatan pengukuran Indeks Davies-Bouldin adalah memaksimalkan jarak antar *cluster* serta meminimalkan jarak dalam *cluster* [12]. 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) [13].

c. Koefisien *Silhouette*

Koefisien *silhouette* mengukur tingkat homogenitas anggota dalam suatu *cluster* dan tingkat heterogenitas anggota antar *cluster* [4]. Hasil validasi *cluster* yang terbentuk dikatakan baik, jika nilai koefisien *silhouette* mendekati 1 dan sebaiknya jika nilai yang diperoleh mendekati -1 [4].

2.6. Pemilihan Metode Terbaik

Hasil analisis *cluster* dapat dikatakan optimal jika objek dalam satu *cluster* bersifat homogen dan sebaliknya, bersifat heterogen antar *cluster* [11]. Menurut Bunkers dkk. dalam Laeli, metode dengan kinerja terbaik dapat diketahui menggunakan rata-rata simpangan baku dalam *cluster* (s_w) dan simpangan baku antar *cluster* (s_b) [6]. Rumus simpangan baku dalam *cluster* dinyatakan dalam persamaan sebagai berikut [14]:

$$S_w = \frac{1}{c}\sum_{m=1}^{C} S_m \tag{3}$$

di mana:

c : Banyak *cluster* yang terbentuk

Jurnal Fraction, Vol. 4 No.2, Desember 2024, Hal. 49-57
E-ISSN: **2830-2028**; P-ISSN: **2962-1402**
Dipublikasi: Desember 2024

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

Adapun simpangan baku antar *cluster* dinyatakan dalam persamaan sebagai berikut [14]:

$$S_b = \sqrt{\frac{1}{c}\sum_{m=1}^{C}(\bar{x}_m - \bar{X})^2} \quad (4)$$

di mana:

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

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

Nilai rasio simpangan baku dapat ditentukan dengan rumus pada persamaan berikut [14]:

$$rasio\ simpangan\ baku = \frac{S_w}{S_b} \quad (5)$$

Metode yang memiliki nilai rasio simpangan baku paling kecil merupakan metode terbaik [15].

3. HASIL DAN PEMBAHASAN

Sebelum dilakukan proses analisis data, data perlu distandarisasi terlebih dahulu karena memiliki satuan yang berbeda-beda dan diperoleh variabel hasil standarisasi berupa variabel Z_1, Z_2, Z_3, Z_4, Z_5, dan Z_6. Proses kemudian dilanjutkan dengan pengujian asumsi analisis *cluster*.

3.1. Pengujian Asumsi Analisis *Cluster*

Tahapan dimulai dengan melakukan uji KMO dan uji multikolinearitas. Untuk menentukan nilai KMO digunakan persamaan (1), dengan hasil sebagai berikut:

$$KMO = \frac{0{,}691^2 + 0{,}683^2 + 0{,}640^2 + \cdots + 0{,}893^2}{(0{,}691^2 + \cdots + 0{,}893^2) + (0{,}429^2 + 0{,}676^2 + \cdots + 0{,}738^2)} = 0{,}661$$

Berdasarkan hasil perhitungan diperoleh nilai KMO sebesar 0,661 yang lebih besar dari 0,5. Oleh karena itu, dapat disimpulkan bahwa data yang digunakan layak untuk dianalisis lebih lanjut.

Tahapan dilanjutkan dengan uji multikolinearitas menggunakan nilai VIF. Hasil dari uji multikolinearitas disajikan dalam Tabel 1 sebagai berikut:

Tabel 1 Hasil Uji Multikolinearitas

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

Berdasarkan tabel 1, variabel-variabel penelitian memiliki nilai $VIF > 10$. Dengan demikian, dapat disimpulkan terdapat multikolinearitas dalam data yang digunakan, sehingga perlu dilakukan transformasi data.

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

Nilai eigen dalam proses transformasi data penelitian ini diperoleh berdasarkan matriks korelasi ($\boldsymbol{R}$). Berdasarkan tabel 2, diperoleh dua komponen utama yang 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 disajikan dalam Tabel 3 sebagai berikut:

Jurnal Fraction, Vol. 4 No.2, Desember 2024, Hal. 49-57
E-ISSN: **2830-2028**; P-ISSN: **2962-1402**
Dipublikasi: Desember 2024

Tabel 3 Rotasi *Factor Loading*

Variabel		Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
Komponen	1	0,442	0,909	0,929	0,942	0,150	0,502
	2	0,864	0,291	0,337	0,287	0,979	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 Z_2, Z_3, dan Z_4 berupa variabel vaksinasi, sakit, sembuh. Adapunn variabel-variabel yang masuk ke dalam KU_2 adalah Z_1, Z_5, dan 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 $(\boldsymbol{f})$ dan diperoleh matriks *factor scores* sebagai berikut:

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

Setelah dilakukan transformasi data proses analisis dapat dilanjutkan dengan analisis *cluster*.

3.2. Hasil Pengelompokan Persebaran PMK di Indonesia dengan Metode *Average Linkage*

Ukuran jarak yang digunakan dalam analisis ini adalah jarak Euclidean. Sebelum dilanjutkan dengan proses pengelompokan, jumlah *cluster* yang ditetapkan dalam analisis ini yaitu 2 hingga 5 *cluster*. Hal ini dikarenakan jika terlalu banyak jumlah *cluster* yang terbentuk, maka akan meyulitkan proses interpetasi *cluster* [16]. Hasil perhitungan validasi *cluster* ditampilkan dalam Tabel 4 berikut:

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

Jumlah *Cluster*	Indeks Dunn (*DI*)	Indeks Davies-Bouldin (*DBI*)	Koefisien *Silhoutte* (*SC*)
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

Pada tabel 4 dapat diketahui nilai *DI* terbesar sebesar 1,41, sehingga jumlah *cluster* optimal berdasarkan indeks Dunn adalah 4 *cluster*. Berdasarkan indeks Davies-Bouldin diperoleh jumlah *cluster* optimal adalah sebanyak 4 *cluster* karena memiliki nilai *DBI* terkecil, yaitu sebesar 0,10. Di samping itu, hasil berbeda ditunjukkan oleh koefisien *silhouette* di mana diperoleh jumlah *cluster* optimal adalah sebanyak 3 *cluster* dengan nilai *SC* sebesar 0,81. Hal ini dikarenakan nilai *SC* paling mendekati 1 terdapat pada jumlah *cluster* sebanyak 3 *cluster*.

Proses validasi dilanjutkan dengan menentukan indeks gabungan dikarenakan terdapat perbedaan hasil keputusan jumlah *cluster* optimal oleh ketiga indeks yang digunakan. Indeks gabungan merupakan 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 [11]. Berikut hasil perhitungan indeks gabungan yang disajikan dalam Tabel 5.

Tabel 5 Indeks Gabungan Metode *Average Linkage*

Jumlah *Cluster*	*DI*	*Rank*	*DBI*	*Rank*	*SC*	*Rank*	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 adalah sebanyak 4 *cluster* karena memiliki nilai indeks gabungan paling kecil. Oleh karena itu, dapat disimpulkan bahwa jumlah *cluster* optimal yang terbentuk dalam mengelompokkan persebaran PMK di Indonesia menggunakan metode *average linkage* adalah sebanyak 4 *cluster*.

Adapun proses pengelompokan metode *average linkage* dimulai dengan menentukan jarak terdekat antar seluruh objek. Pada pengelompokan metode *average linkage* terdapat 27 *cluster* awal sebanyak

Jurnal Fraction, Vol. 4 No.2, Desember 2024, Hal. 49-57
E-ISSN: **2830-2028**; P-ISSN: **2962-1402**
Dipublikasi: Desember 2024

objek yang dikelompokkan. Setelah proses aglomerasi dilakukan, dapat diketahui setiap provinsi masuk ke dalam *cluster* mana saja. Hasil pengelompokan metode *average linkage* ditunjukkan pada Tabel 6.

Tabel 6 Hasil Pengelompokan Metode *Average Linkage*

C	Jumlah Anggota	Provinsi
1	24	Aceh, Bali, Bangka Belitung, Banten, Bengkulu, DIY, DKI Jakarta, Jambi, Jawa Tengah, Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Kepulauan Riau, Lampung, Riau, Sulawesi Barat, Sulawesi Selatan, Sulawesi Tengah, Sumatera Tenggara, Sumatera Barat, Sumatera Selatan, dan Sumatera Utara
2	1	Nusa Tenggara Barat
3	1	Jawa Timur
4	1	Jawa Barat

3.3. Hasil Pengelompokan Persebaran PMK di Indonesia dengan Metode *K-Means*

Metode *k-means* termasuk ke dalam analisis *cluster* nonhierarki, sehingga dalam metode ini jumlah *cluster* harus ditetukan terlebih dahulu. Dalam analisis ini, jumlah *cluster* yang digunakan, yaitu 2 hingga 5 *cluster*. Adapun hasil validasi *cluster* metode *k-means* disajikan dalam Tabel 7 sebagai berikut:

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

Jumlah *Cluster*	Indeks Dunn (DI)	Indeks Davies-Bouldin (DBI)	Koefisien *Silhoutte* (SC)
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

Pada Tabel 7 dapat diketahui nilai DI terbesar adalah 0,84, sehingga jumlah *cluster* optimal berdasarkan indeks Dunn adalah 3 *cluster*. Berdasarkan indeks Davies-Bouldin diperoleh jumlah *cluster* optimal sebanyak 3 *cluster* karena memiliki nilai DBI terkecil, yaitu sebesar 0,40. Begitu pula berdasarkan koefisien *silhouette* diperoleh jumlah *cluster* optimal sebanyak 3 *cluster* dengan nilai SC sebesar 0,81. Hal ini dikarenakan nilai SC paling mendekati 1 terdapat pada jumlah *cluster* sebanyak 3 *cluster*. Oleh karena itu, berdasarkan ketiga indeks diperoleh jumlah *cluster* optimal yang terbentuk dalam mengelompokkan persebaran PMK di Indonesia menggunakan metode *k-means* adalah 3 *cluster*.

Pada proses 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 menggunakan 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	Aceh, Bali, Bangka Belitung, Banten, Bengkulu, DIY, DKI Jakarta, Jambi, Jawa Tengah, Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Kepulauan Riau, Lampung, Riau, Sulawesi Barat, Sulawesi Selatan, Sulawesi Tengah, Sumatera Tenggara, Sumatera Barat, Sumatera Selatan, Sumatera Utara
2	1	Jawa Barat
3	1	Nusa Tenggara Barat dan Jawa Timur

3.4. Pemilihan Metode Terbaik

Pemilihan metode terbaik berguna untuk mengetahui seberapa baik kinerja kedua metode dalam mengelompokkan persebaran PMK di Indonesia. Suatu *cluster* dikatakan baik jika memiliki keheterogenan yang tinggi antar *cluster* dan kehomogenan tinggi dalam *cluster*. Kehomogenan dalam *cluster*

Jurnal Fraction, Vol. 4 No.2, Desember 2024, Hal. 49-57
E-ISSN: **2830-2028**; P-ISSN: **2962-1402**
Dipublikasi: Desember 2024

dapat diketahui melalui nilai simpangan baku dalam *cluster* (S_w), sedangkan keheterogenan antar *cluster* dapat diketahui melalui nilai simpangan baku antar *cluster* (S_b). Kemudian diperoleh nilai rasio simpangan baku untuk masing-masing hasil pengelompokan yang dirangkum dalam Tabel 9 berikut:

Tabel 9 Nilai Rasio Simpangan Baku

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

Berdasarkan tabel 9, didapatkan bahwa metode terbaik dalam mengelompokkan persebaran PMK di Indonesia adalah metode *average linkage* dengan nilai rasio simpangan baku yang lebih kecil dibandingkan metode *k-means*. Masing-masing *cluster* memiliki karakteristik yang berbeda berdasarkan keenam indikator yang mengukur persebaran PMK di Indonesia, dimana keenam indikator ini direduksi dalam dua komponen utama, yaitu KU_1 dan KU_2. Karakteristik persebaran PMK di Indonesia pada masing-masing *cluster* dapat diketahui melalui nilai rata-rata setiap komponen utama. Nilai rata-rata komponen utama masing-masing *cluster* pada metode *average linkage* adalah sebagai berikut:

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

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

Berdasarkan tabel 10, dapat dilakukan interpetasi untuk hasil pengelompokan persebaran PMK di Indonesia menggunakan metode *average linkage*. *Cluster* 1 beranggotakan 24 provinsi seperti tercantum dalam tabel 6. *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.

4. KESIMPULAN DAN SARAN

Berdasarkan analisis yang telah dilakukan, dapat ditarik kesimpulan sebagai berikut: (1) Hasil pengelompokan persebaran 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 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 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.

5. UCAPAN TERIMAKASIH

Terima kasih kepada dosen pembimbing yang telah membimbing dan memfasilitasi kegiatan penelitian ini, serta terima kasih juga kepada pihak-pihak yang telah membantu dalam proses penelitian.

DAFTAR PUSTAKA

[1] Direktorat Kesehatan Hewan. (2022). *Pedoman Kesiagaan Darurat Veteriner Indonesia, Seri: Penyakit Mulut dan Kuku*. Kementerian Pertanian.

[2] 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/.

[3] Sandag, A. M. V., Mongi, C. E., & Mananohas, M. L. (2020). Pengelompokan Sekolah Menegah Atas (SMA) di Kabupaten Minahasa Tenggara Berdasarkan Stadar 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.

[4] 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.

[5] 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.

[6] 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.

[7] Haumahu, G. & Lewaherilla, N. (2020). Penerapan Analisis Komponen Utama dalam Mereduksi Faktor-Faktor Penyebab Diare di Provinsi Maluku. *Map Journal*, 2(1), 41-46. https://doi.org/10.15548/map.v2i1.1638.

[8] 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.

[9] Supiati. (2010). *Aplikasi Analisis Cluster dalam Mengelompokkan Kecamatan Berdasarkan Faktor Penyebab Penduduk Buta Aksara di Kabupaten Jeneponto*. Universitas Islam Negeri Alauddin.

[10] 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.

[11] Cahyoningtyas, R. A. (2019). *Metode Ward dan Average Linkage Clustering untuk Segmentasi Objek Wisata di Malag Raya*. Universitas Brawijaya.

[12] 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.

[13] 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.

[14] 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.

[15] Laeli S. (2014). *Analisis Cluster dengan Average Linkage Method dan Ward's Method untuk Data Responden*. Universitas Negeri Yogyakarta.

[16] 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.