

IMPLEMENTASI ALGORITMA *K-MEDOIDS* DAN *K-ERROR* UNTUK PENGELOMPOKAN KABUPATEN/KOTA DI PROVINSI JAWA TENGAH BERDASARKAN JUMLAH PRODUKSI PETERNAKAN TAHUN 2020

Fahrur Rozzi Iskak^{1*}, Iut Tri Utami², Triastuti Wuryandari³

^{1,2,3} Departemen Statistika FSM Universitas Diponegoro

*email: rozzi120221@gmail.com

DOI: 10.14710/j.gauss.11.3.366-376

Article Info:

Received: 2022-06-15

Accepted: 2022-08-20

Available Online: : 2023-01-03

Keywords:

Livestock; Cluster Analysis; *K-Error Method*; *K-Medoids*; Standard deviation ratio; WB index

Abstract: The livestock sub-sector is one of the sub-sectors that contribute to the national economy and can significantly absorb labour so that it can be relied upon in efforts to improve the national economy. One of the steps used to increase livestock production in each region in Central Java Province is regional mapping. Cluster analysis is one of the regional mapping methods that can increase livestock production by grouping regencies/cities with characteristics of the same level of livestock production based on the type of livestock production. The *k-error* and *k-medoids* method is a non-hierarchical cluster analysis method, where the *k-error* is a method developed to overcome the problem of data measurement errors in classical cluster analysis, while the *k-medoids* is a method used to overcome the problem of outliers contained in the data. The validity test of the standard deviation ratio and the WB Index was used to determine the quality of the clustering results. The small validity value of the standard deviation ratio and the WB Index shows the best results of clustering and selecting method. Based on the results of the clustering, the optimal cluster was obtained at $k=7$ using the *k-medoids* algorithm, where the validation value of the standard deviation ratio=0.773 and WB Index=0.531.

1. PENDAHULUAN

Subsektor peternakan merupakan salah satu subsektor yang memberikan kontribusi pada perekonomian nasional serta mampu menyerap tenaga kerja secara signifikan, sehingga dapat diandalkan dalam upaya perbaikan perekonomian nasional. Pemerintah daerah dan para peternak harus berupaya bersama dalam rangka meningkatkan produksi baik kuantitas maupun kualitas produksi peternakan, karena konsumsi masyarakat terhadap daging dan produk peternakan lain meningkat setiap tahunnya. Kondisi ini menyebabkan pemerintah harus membuat kebijakan yang dapat meningkatkan dan menguntungkan para peternak dan dapat meningkatkan produksi sektor peternakan serta menjamin ketersediaan pangan. Salah satu upaya yang dapat dilakukan pemerintah dalam rangka meningkatkan dan memelihara produksi di sektor peternakan yaitu dengan mengelompokkan wilayah kabupaten/kota ke dalam kelompok-kelompok berdasarkan kesamaan karakteristik yang dimiliki. Salah satu cara untuk pengelompokan ini dapat dilakukan dengan menggunakan analisis kluster.

Analisis kluster adalah teknik multivariat yang bertujuan untuk mengelompokkan objek berdasarkan karakteristik yang dimiliki oleh objek tersebut. Pada analisis kluster terdapat masalah yang berpengaruh terhadap hasil pengklasteran yang dicapai, salah satunya adalah data yang mengandung pencilan (*outlier*). Algoritma dari analisis kluster yang dapat mengatasi masalah tersebut adalah algoritma *k-medoids*, dengan menggunakan algoritma *k-medoids* objek-objek bisa dikelompokkan pada suatu data yang mengandung pencilan. Pada metode analisis kluster tradisional, diasumsikan tidak ada kesalahan pengukuran yang berhubungan dengan data. Untuk mengatasi hal tersebut, maka dikembangkan algoritma

klustering yang tidak mengabaikan kesalahan pengukuran pada data. Algoritma yang dikembangkan adalah algoritma klustering berdasarkan kesalahan (*error based klustering*), *h-error* dan *k-error*, yang merupakan generalisasi dari algoritma klustering hierarki *ward* dan *k-means*.

Penelitian ini membahas tentang pengelompokan dengan metode *k-medoids* dan *k-error* dengan variabel jenis produksi ternak berdasarkan Kabupaten/Kota di Provinsi Jawa Tengah Tahun 2020. Penentuan metode terbaik diantara metode *k-medoids* dan *k-error* dapat dilakukan dengan uji validasi. Dalam penelitian ini menggunakan validasi rasio simpangan baku dan WB Indeks, dimana metode dengan nilai validasi terkecil merupakan metode pengklasteran yang terbaik.

2. TINJAUAN PUSTAKA

Analisis kluster merupakan teknik variabel ganda yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan kemiripan karakteristik yang dimilikinya. Karakteristik objek-objek dalam suatu kluster memiliki tingkat kemiripan yang tinggi, sedangkan karakteristik antar objek pada suatu kluster dengan kluster lain memiliki tingkat kemiripan yang rendah. Dengan kata lain, keragaman dalam suatu kluster minimum sedangkan antar keragaman antar kluster maksimum (Sumertajaya, 2011).

Asumsi yang harus terpenuhi sebelum melakukan analisis kluster ada dua, yaitu sampel yang mewakili populasi atau sampel representatif dan tidak terjadi multikolinieritas.

a. Sampel representatif

Uji KMO (Kaiser Mayer Olkin) digunakan untuk uji sampel representatif. Sampel yang representatif atau sampel mewakili populasi adalah sampel yang dapat dikatakan mewakili populasi yang ada.

Uji hipotesis untuk melihat apakah sampel dapat mewakili populasi yang ada adalah sebagai berikut :

Hipotesis

H_0 : Sampel representatif atau sampel mewakili populasi

H_1 : Sampel tidak representatif atau sampel tidak mewakili populasi

Menurut Widarjono (2010), Rumus KMO adalah sebagai berikut (Widarjono 2010):

$$KMO = \frac{\sum_{k=1}^p \sum_{l=1, l \neq k}^p r_{X_k X_l}^2}{\sum_{k=1}^p \sum_{l \neq k}^p r_{X_k X_l}^2 + \sum_{k=1}^p \sum_{l \neq k}^p \rho_{X_k X_l, m}^2}$$

dengan,

$$r_{X_k X_l} = \frac{\frac{1}{n} \sum_{i=1}^n (X_{ik} - \bar{X}_k)(X_{il} - \bar{X}_l)}{\sqrt{\frac{\sum_{i=1}^n (X_{ik} - \bar{X}_k)^2}{n}} \sqrt{\frac{\sum_{i=1}^n (X_{il} - \bar{X}_l)^2}{n}}}$$

$$\rho_{X_k X_l, X_m} = \frac{r_{X_k X_l} - r_{X_k X_m} r_{X_l X_m}}{\sqrt{(1 - r_{X_k X_m}^2)(1 - r_{X_l X_m}^2)}}$$

dengan

p : banyaknya variabel

n : banyaknya objek.

$r_{X_k X_l}$: korelasi antara variabel X_k dan X_l .

\bar{X}_k : rata – rata variabel X_k .

\bar{X}_l : rata – rata variabel X_l

$\rho_{X_k X_l, X_m}$: korelasi parsial antara variabel x_j dan x_l dengan menjaga agar x_m konstan.

Menurut Widarjono (2010), jika nilai KMO yang diperoleh dari data antara 0,5 sampai 1 maka dapat disimpulkan bahwa sampel mewakili populasi atau sampel representatif, sehingga bisa dilanjutkan ke tahap analisis berikutnya.

b. Deteksi Multikolinieritas

Salah satu cara identifikasi adanya multikolinieritas adalah menghitung nilai *Variance Inflation Factor* (VIF) (Gujarati 2009). Uji hipotesis untuk melihat ada tidaknya multikolinieritas dalam data adalah sebagai berikut:

Hipotesis

H_0 : Data tidak terjadi multikolinieritas

H_1 : Data terjadi multikolinieritas (Ada hubungan linier antar variabel)

Rumus Nilai *Variance Inflation Factor* (VIF) adalah sebagai berikut:

$$VIF_i = \frac{1}{(1 - R_i^2)}$$

Dengan,

R_i^2 : koefisien determinasi yang diperoleh bila nilai variabel ke-i yang diregresikan dengan variabel lainnya

i : variabel, dengan $i = 1, 2, \dots, p$

VIF_i : nilai *Variance Inflation Factor* variabel ke - i

Jika nilai $VIF \leq 10$ maka bisa disimpulkan bahwa tingkat multikolinieritas tidak tinggi atau masih bisa ditolerir. Salah satu cara yang dapat dilakukan bila terjadi multikolinieritas adalah dengan mengeluarkan variabel yang berkorelasi dalam model. Selain itu, multikolinieritas juga bisa diperbaiki dengan Analisis Komponen Utama yang biasanya juga disebut *PCA (Principal Components Analysis)*.

Metode *k-error* merupakan salah satu metode analisis kluster yang memasukkan informasi kesalahan terkait data. Metode *k-error* digunakan saat diketahui jumlah kluster sejak awal. Metode *k-error* bersifat iteratif dengan tahap awal yaitu inisialisasi partisi data dengan algoritma sebagai berikut (Kumar dan Patel 2005) :

1. Menentukan banyaknya kluster sebanyak *k*-kluster.
2. Hitung pusat kluster untuk masing-masing kluster dengan menggunakan rumus sebagai berikut :

$$\underline{\hat{\theta}}_k = \left(\sum_{i \in R_k} S_i^{-1} \right)^{-1} \left(\sum_{i \in R_k} S_i^{-1} \underline{x}_i \right)$$

dengan,

$R_k = \{i | x_i \in C_k\}$ dengan $i = 1,2,3,\dots,n$;

n = banyaknya objek pengamatan

k = jumlah kluster

C_k = kluster ke-*k*

\underline{x}_i = vektor objek pengamatan ke-i

S_i = matriks ragam peragam pada objek pengamatan ke-i

$$S_i = \begin{bmatrix} S_{xi1}^2 & S_{i1,i2} & \dots & S_{i1,ip} \\ S_{i2,i1} & S_{xi2}^2 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ S_{ip,i1} & S_{ip,i2} & \dots & S_{xip}^2 \end{bmatrix}$$

S_{xij}^2 = ragam objek pengamatan ke-i dari variabel ke-j

S_{ij_1,ij_2} = peragam antara objek pengamatan i variabel ke-j₁, dan objek pengamatan ke-i pada variabel ke-j₂, dengan $j_1 \neq j_2$.

3. Hitung kembali jarak setiap objek pengamatan ke masing-masing pusat klaster, masukkan objek pengamatan ke dalam klaster dengan jarak terdekat dengan menggunakan rumus persamaan :

$$d_{ik} = (\underline{x}_i - \widehat{\underline{\theta}}_k)^t S_i^{-1} (\underline{x}_i - \widehat{\underline{\theta}}_k)$$

dengan,

d_{ik} = jarak x_i dari pusat klaster ke- k

x_i = nilai observasi ke- i

$\widehat{\underline{\theta}}_k$ = penduga maksimum likelihood pusat klaster ke- k

4. Ulangi langkah 1 sampai dengan 3, sampai tidak ada lagi perubahan keanggotaan klaster.

Partitioning Around Medoids (PAM) atau *k-medoids* merupakan algoritma klustering yang hampir sama dengan algoritma *k-means*. Perbedaan dari kedua algoritma ini adalah algoritma *k-medoids* menggunakan objek perwakilan (*medoid*) sebagai pusat klaster (*centroid*) untuk setiap klaster, sedangkan *k-means* menggunakan nilai rata-rata (*mean*) sebagai pusat klaster (*centroid*). Algoritma *k-medoids* merupakan metode partisi klustering yang mengelompokkan sekumpulan n objek menjadi sejumlah k -klaster. Klaster dibangun dengan menghitung kedekatan yang dimiliki antara *medoids* dengan objek *non-medoids*.

Kelebihan dari algoritma *k-medoids* dapat mengatasi kelemahan pada pada algoritma *k-means* yang sensitif terhadap *noise* dan *outlier* atau pencilan karena *k-means* menggunakan nilai rata-rata (*mean*) sebagai pusat klasternya, selain itu metode *k-medoids* adalah metode yang dapat mengelompokkan data dengan *outlier* (pencilan) atau tanpa pencilan serta metode *k-medoids* juga dapat digunakan untuk data dengan variabel dominan kontinu maupun diskrit.

Langkah-langkah dari proses analisis klaster menggunakan algoritma *k-medoids* adalah sebagai berikut :

1. Menentukan k sebagai banyaknya klaster yang ingin dibentuk (tergantung banyaknya data yang digunakan).
 2. Menentukan *medoids* atau pusat klaster secara acak
 3. Menghitung jarak objek *non-medoid* dengan *medoid* pada tiap klaster dan menempatkan tiap objek *non-medoid* tersebut ke *medoid* terdekat, kemudian hitung total jaraknya
 4. Memilih kandidat *medoids* baru secara acak dari objek *non-medoid* pada masing-masing klaster
 5. Menghitung jarak setiap objek *non-medoid* dengan *medoid* baru dan menempatkan tiap objek *non-medoid* tersebut ke kandidat *medoid* terdekat, kemudian hitung total jaraknya
 6. Menghitung selisih total jarak ($S_{\text{total jarak}}$), dengan $S_{\text{total jarak}} = \text{total jarak pada kandidat medoid baru} - \text{total jarak pada medoid lama}$
 7. Jika diperoleh nilai $S_{\text{total jarak}} < 0$, maka kandidat *medoid* baru tersebut menjadi *medoid* baru dan jika diperoleh $S_{\text{total jarak}} > 0$ iterasi berhenti
- Kembali ke langkah (4) sampai (7) sampai tidak terjadi perubahan *medoid* atau $S_{\text{total jarak}} > 0$

Menurut Kalkstein *et al.*(1987), jarak dari klaster yang terbentuk dapat dijadikan acuan dalam melakukan uji validitas klaster. Terdapat dua ukuran yang digunakan dalam melakukan validitas klaster, yaitu sebagai berikut :

1. Jarak di dalam klaster yang dapat diukur dengan menggunakan rata-rata simpangan baku dalam klaster dengan rumus sebagai berikut :

$$S_w = \frac{1}{g} \sum_{k=1}^g \hat{\sigma}_{jk}$$

dengan,

S_w = rata-rata simpangan baku dalam klaster

$\hat{\sigma}_{jk}$ = simpangan baku dalam klaster ke-k variabel ke-j

2. Jarak antar klaster yang dapat diukur dengan menggunakan rata-rata simpangan baku antar klaster dengan rumus sebagai berikut :

$$S_b = \left[\frac{1}{g-1} \sum_{k=1}^g (\bar{x}_{jk} - \bar{\bar{x}}_j)^2 \right]^{1/2}$$

dengan,

\bar{x}_{jk} = nilai rata-rata variabel ke-j pada klaster ke-k

$\bar{\bar{x}}_j$ = nilai rata-rata variabel ke-j untuk seluruh klaster

$$R = \frac{1}{p} \sum_{i=1}^p \frac{S_w}{S_b}$$

dengan,

R = rasio S_w terhadap S_b

p = banyaknya variabel

Nilai rasio simpangan baku atau rasio antara S_w dan S_b dapat dijadikan acuan dalam menentukan jumlah klaster optimum. Nilai rasio simpangan baku yang kecil menunjukkan jumlah klaster yang optimum.

Menurut Zhao (2012), keragaman klaster dapat digunakan untuk melakukan validitas pengklasteran, yaitu keragaman dalam klaster dan keragaman antar klaster. Beberapa definisi keragaman sebagai berikut :

1. Jumlah kuadrat dalam klaster (*Sum of Square Within Cluster*) yang dinotasikan dengan SS_w . Jumlah kuadrat dalam klaster merupakan ukuran kedekatan antar objek dalam klaster. Nilai SS_w dapat dihitung dengan rumus sebagai berikut :

$$SS_w = \sum_{k=1}^g \sum_{i=1}^{n_k} \sum_{j=1}^p (x_{ijk} - \bar{x}_{kj})^2$$

dengan,

x_{ijk} : nilai objek ke-i pada variabel ke-j pada klaster ke-k ($k=1,2,\dots,g$)

\bar{x}_{kj} : rata-rata nilai variabel ke-j pada klaster ke-k ($k=1,2,\dots,g$)

2. Jumlah kuadrat antar klaster (*Sum of Square Between Cluster*) yang dinotasikan dengan SS_B . Nilai SS_B menunjukkan ukuran keterpisahan antar klaster. Nilai SS_B dapat dihitung dengan rumus sebagai berikut :

$$SS_B = \sum_{k=1}^g n_k \sum_{j=1}^p (\bar{x}_{kj} - \bar{\bar{x}}_j)^2$$

dengan,

\bar{x}_{kj} : rata-rata nilai variabel ke-j pada klaster ke-k ($k=1,2,\dots,g$)

$\bar{\bar{x}}_j$: rata-rata nilai variabel ke-j dari seluruh objek

n_k : banyaknya objek pada klaster ke-k

Nilai SS_w dan SS_B tersebut dapat digunakan untuk menentukan jumlah klaster yang optimum yaitu menggunakan WB indeks. Nilai WB indeks yang kecil mengindikasikan solusi jumlah klaster optimum (Zhao, 2012).

$$WB = k \times \frac{SS_w}{SS_B}$$

3. METODOLOGI PENELITIAN

Data yang digunakan dalam penelitian ini adalah data produksi ternak berdasarkan jenis ternak tiap kabupaten/kota di Provinsi Jawa Tengah tahun 2020. Data yang digunakan merupakan data sekunder yang diperoleh dari laman web BPS (Badan Pusat Statistik) Provinsi Jawa Tengah yang termuat dalam buku 'Provinsi Jawa Tengah Dalam Angka 2021'. Adapun Provinsi Jawa Tengah terdapat 35 kabupaten/kota.

Variabel yang digunakan dalam penelitian ini adalah :

- X1 : Jumlah produksi daging sapi
- X2 : Jumlah produksi daging kambing
- X3 : Jumlah produksi daging domba
- X4 : Jumlah produksi telur ayam kampung
- X5 : Jumlah produksi telur ayam petelur
- X6 : Jumlah produksi telur itik
- X7 : Jumlah produksi susu sapi
- X8 : Jumlah produksi kulit sapi/kerbau
- X9 : Jumlah produksi kulit kambing/domba

Seluruh variabel yang digunakan memiliki satuan ton.

Pada penelitian ini digunakan analisis kluster dengan menggunakan algoritma kluster *k-medoids* dan *k-error*. Data penelitian ini diolah dengan menggunakan software Microsoft Excel dan R i386 4.1.0. Tahapan analisis data yang digunakan adalah sebagai berikut :

1. Deteksi outlier/pencilan dari data.
2. Menghitung matriks ragam peragam dari data karena metode *k-error* membutuhkan matriks ragam peragam.
3. Uji asumsi analisis kluster yaitu :
 - a. Uji asumsi sampel representatif
 - b. Uji asumsi non-multikolinieritas
4. Menentukan banyaknya kluster yang akan dibuat (*k*), nilai *k* yang digunakan dalam penelitian ini yaitu $k = 2,3,4,5,6,7$
5. Melakukan analisis kluster
 - a. Algoritma metode *k-error*, tahapan yang dilakukan adalah sebagai berikut :
 1. Mempartisi objek pengamatan menjadi *k*-kelompok secara acak.
 2. Menghitung pusat kluster *k-error* untuk masing-masing kluster
 3. Menghitung jarak kuadrat setiap objek ke masing-masing pusat kluster.
 4. Menentukan keanggotaan kluster baru dari setiap objek yang memiliki jarak terdekat dengan pusat kluster.
 5. Mengulangi langkah 2 hingga 4 sampai tidak ada lagi objek pengamatan yang berpindah kluster atau perubahan keanggotaan kluster.
 - b. Algoritma metode *k-medoids*, tahapan yang dilakukan adalah sebagai berikut :
 1. Memilih *k* objek untuk menjadi O_c , dengan O_c adalah objek yang menjadi pusat kluster atau *medoid* di kluster ke-*c* dan $c = 1,2, \dots, k$
 2. Menghitung kemiripan antara objek *medoid* dengan objek *non-medoids* menggunakan jarak Euclidean.
 3. Menentukan anggota dari masing-masing kluster berdasarkan jarak terdekat dengan O_c atau *medoid*, dengan jarak terdekat = $\min \{d(x_i, c_1), d(x_i, c_2), \dots, d(x_i, c_k)\}$
 4. Menghitung total jarak objek *non-medoids* dengan objek O_c / total cost lama

5. Secara acak memilih objek Orandom, dengan Orandom adalah sebuah objek non- *medoids* untuk menggantikan Oc awal.
 6. Menghitung jarak objek *non-medoids* dengan objek Orandom menggunakan jarak Euclidean.
 7. Menentukan keanggotaan klaster baru dari setiap objek yang memiliki jarak terdekat dengan Orandom.
 8. Menghitung total jarak objek *non-medoids* terdekat dengan objek Orandom / total cost baru.
 9. Menghitung selisih total jarak ($S_{\text{total jarak}}$). Jika diperoleh nilai $S_{\text{total jarak}} < 0$, maka kandidat *medoid* baru tersebut menjadi *medoid* baru dan jika diperoleh $S_{\text{total jarak}} > 0$ iterasi berhenti.
 10. Kembali ke langkah (4) sampai (6) sampai tidak terjadi perubahan *medoid* atau $S_{\text{total jarak}} > 0$.
6. Melakukan uji validasi klaster menggunakan rasio simpangan baku dan WB Indeks.
 7. Melakukan ulasan perbandingan hasil dari metode yang digunakan yaitu *k-error* dan *k-medoids*.
 8. Interpretasi dan profilisasi karakteristik wilayah yang dipilih dari hasil pengklasteran terbaik.

4. ANALISIS DAN PEMBAHASAN

Pendeteksian pencilan atau *outlier* pada penelitian ini menggunakan metode jarak kuadrat mahalnobis. Pengamatan ke- i teridentifikasi pencilan jika, $d_{MD}^2(i) > \chi_{p,1-\alpha}^2$. Nilai p merupakan banyaknya variabel yang diteliti yaitu sebanyak 9 variabel dan nilai α yang digunakan sebesar 5%, sehingga diperoleh nilai $\chi_{p=9,(1-0.05)}^2$ sebesar 16,91898. Berdasarkan hasil pendeteksian pencilan dengan membandingkan hasil jarak kuadrat mahalnobis tiap objek, diketahui bahwa terdapat 5 kabupaten yang merupakan pencilan, yaitu Kabupaten Kebumen, Kabupaten Boyolali, Kabupaten Grobogan, Kabupaten Semarang dan Kabupaten Kendal. Pada kasus ini data pencilan tidak dihilangkan dan tetap disertakan dalam tahap analisis berikutnya karena ingin diketahui metode mana yang paling robust atau tidak terpengaruh terhadap pencilan dan diharapkan data pencilan nantinya dapat membuat klaster atau kelompok tersendiri.

Langkah pertama dalam proses pengklasteran yaitu dengan melakukan uji asumsi, dimana dalam analisis klaster terdapat dua asumsi yang harus terpenuhi yaitu:

a. Data Mewakili Populasi

Pengujian sampel representatif dapat dilakukan dengan menggunakan uji KMO (Kaiser Mayer Olkin), data atau sampel diharapkan dapat mewakili dari keseluruhan populasi. Berdasarkan uji KMO diperoleh nilai KMO sebesar 0,57. Sampel dikatakan mewakili populasi atau sampel representatif apabila diperoleh nilai KMO berkisar antara 0,5 sampai dengan 1. Berdasarkan nilai pengujian KMO tersebut, dimana nilai KMO berada diantara 0,5 sampai 1, sehingga dapat disimpulkan bahwa sampel dikatakan mewakili populasi atau sampel representatif.

b. Uji Multikolinearitas

Berdasarkan hasil pengujian nilai VIF, diperoleh nilai VIF sebagai berikut :

Tabel 1. Nilai VIF dari 9 Variabel

Variabel	Nilai VIF
Jumlah produksi daging sapi	89,965
Jumlah produksi daging kambing	5,563
Jumlah produksi daging domba	5,207

Jumlah produksi telur ayam kampung	1,717
Jumlah produksi telur ayam petelur	1,278
Jumlah produksi telur itik	1,331
Jumlah produksi susu sapi	7,461
Jumlah produksi kulit sapi/kerbau	110,330
Jumlah produksi kulit kambing/domba	11,489

Jika nilai VIF > 10 maka pada data yang diuji terjadi multikolinearitas antar variabel. Berdasarkan pengujian nilai VIF, diperoleh 6 variabel dengan nilai VIF ≤ 10, maka ke-6 variabel tersebut tidak terjadi multikolinearitas. Ada 3 variabel yang mempunyai nilai VIF > 10 yaitu variabel jumlah produksi daging sapi, variabel jumlah produksi kulit sapi/kerbau dan variabel jumlah produksi kulit kambing/domba, sehingga dapat disimpulkan bahwa variabel tersebut terdapat multikolinearitas. Uji asumsi multikolinearitas belum terpenuhi, karena masih terdapat variabel yang terjadi multikolinearitas, sehingga perlu dilakukan penanganan multikolinearitas dengan menggunakan analisis komponen utama.

Berdasarkan data yang telah diujikan pada uji asumsi, masih terdapat variabel yang terjadi multikolinearitas, maka solusi untuk menangani masalah tersebut adalah dengan menggunakan analisis komponen utama. Penentuan jumlah komponen utama dibentuk melalui kriteria berdasarkan nilai *eigenvalue*, hasil output ditunjukkan pada Tabel 2.

Tabel 2. Nilai *Eigen*, Proporsi Varian, dan Proporsi Kumulatif

Komponen Utama	Nilai <i>Eigen</i>	Proporsi Varian	Proporsi Kumulatif
1	1146073983,98535	0,703870000	0,7038700
2	238340252,78420	0,146378500	0,8502485
3	159308611,02149	0,097840590	0,9480891
4	74441383,99509	0,045718740	0,9938079
5	9313010,33668	0,005719656	0,9995275
6	664188,34131	0,000407916	0,9999354
7	58759,65960	0,000036088	0,9999750
8	37614,55097	0,000023101	0,9999946
9	8754,99426	0,000005377	1

Penentuan jumlah komponen yang akan dipilih menggunakan nilai *eigen* dengan ketentuan dimana nilai *eigen* tiap variabel yang dipilih adalah yang mempunyai nilai lebih dari satu ($\lambda_i \geq 1$). Berdasarkan Tabel 2 di atas dapat diketahui nilai *eigen* dari tiap komponen dan ditunjukkan bahwa semua komponen atau 9 komponen tersebut mempunyai nilai yang lebih besar dari 1 ($\lambda_i > 1$), sehingga 9 komponen utama dipilih dengan proporsi komponen masing-masing ditunjukkan pada Tabel 2. Adapun proporsi kumulatif dari 9 komponen utama sebesar 1, artinya 9 variabel yang diujikan harus dijelaskan oleh semua komponen utama atau 9 komponen utama.

Berdasarkan hasil analisis pengklasteran menggunakan algoritma *k-error* dan *k-medoids* dengan uji validitas rasio simpangan baku dan WB Indeks untuk $k = 2,3,4,5,6$ dan 7, selanjutnya akan dilakukan perbandingan nilai validasi. Perbandingan nilai validasi digunakan untuk mendapatkan pengklasteran yang terbaik, berikut tabel perbandingan hasil validasi :

Tabel 3. Perbandingan Hasil Pengklasteran

Klaster	Algoritma	Rasio Simpangan Baku	WB Indeks
2	<i>k-error</i>	4,157	13,093
	<i>k-medoids</i>	17,037	1,691
3	<i>k-error</i>	4,138	54,877
	<i>k-medoids</i>	6,127	1,750
4	<i>k-error</i>	1,810	15,929
	<i>k-medoids</i>	1,362	1,285
5	<i>k-error</i>	1,339	21,216
	<i>k-medoids</i>	1,147	0,847
6	<i>k-error</i>	1,371	6,892
	<i>k-medoids</i>	1,011	0,668
7	<i>k-error</i>	1,135	19,661
	<i>k-medoids</i>	0,773	0,531
Rasio Simpangan Baku (minimum)		0,773	
WB Indeks (minimum)		0,531	

Berdasarkan Tabel 3, dapat diketahui bahwa pada pengelompokan kabupaten/kota di Provinsi Jawa Tengah dengan metode *k-error* dan *k-medoids* dengan dua validasi diperoleh pengklasteran terbaik yaitu :

- Untuk validasi Rasio Simpangan Baku, diperoleh nilai validasi minimum pada $k = 7$ yaitu sebesar 0,773
- Untuk validasi WB Indeks, diperoleh nilai validasi minimum pada $k = 7$ yaitu sebesar 0,531

Berdasarkan hasil pengklasteran menggunakan algoritma *k-error* dan *k-medoids* dengan dua validasi yaitu validasi Rasio Simpangan Baku dan WB Indeks dapat disimpulkan bahwa hasil pengklasteran terbaik pada data jumlah produksi hasil peternakan di Provinsi Jawa Tengah tahun 2020 adalah pengklasteran pada $k = 7$ karena memiliki nilai indeks validasi yang optimum pada masing-masing validasi. Rata-rata tiap klaster atau bisa disebut dengan centroid diperlukan untuk mengetahui karakteristik pada klaster. Hasil untuk tiap nilai centroid adalah sebagai berikut :

Tabel 4. Nilai Centroid masing-masing Klaster untuk $k = 7$

X	K1	K2	K3	K4	K5	K6	K7
X1	1.222,631	1.798,407	1.081,420	1.327,985	14.081,220	1.360,726	908,950
X2	175,944	432,372	259,140	798,420	198,380	643,098	147,420
X3	70,031	144,318	5,460	677,598	63,440	385,388	27,040
X4	537,499	1.451,838	3.004,470	1.557,803	166,440	1.620,386	606,050
X5	3.433,651	10.743,403	685,940	7.009,820	23.948,950	6.533,630	46.739,810
X6	631,396	1.144,963	76.218,700	1.012,310	993,900	2.365,108	1.549,480
X7	975,788	5.613,252	0	441,400	51.388,020	3,572	0
X8	6.373,868	10.020,258	6.047,900	7.708,750	70.823	7.346,480	4.601
X9	17.084,699	36.927,258	18.900,440	116.341,250	18.230	64.980,200	12.146

Keterangan:

- : Rata-rata Tertinggi
- : Rata-rata Terendah

Berikut adalah daerah dari hasil pengklasteran tujuh klaster :

Tabel 5. Hasil Pengklasteran Tujuh Klaster

Klaster	Anggota Klaster
1	Kab. Cilacap, Kab. Banjarnegara, Kab. Purworejo, Kab. Wonosobo, Kab. Klaten, Kab. Blora, Kab. Rembang, Kab. Kudus, Kab. Jepara, Kab. Temanggung, Kab. Batang, Kab. Pekalongan, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang, Kota Pekalongan
2	Kab. Banyumas, Kab. Purbalingga, Kab. Karanganyar, Kab. Pati, Kab. Semarang, Kota Tegal
3	Kab. Kebumen
4	Kab. Magelang, Kab. Sragen, Kab. Grobogan, Kab. Tegal
5	Kab. Boyolali
6	Kab. Sukoharjo, Kab. Wonogiri, Kab. Demak, Kab. Pemalang, Kab. Brebes
7	Kab. Kendal

Dari hasil 7 (tujuh) pengklasteran yang terbentuk, dapat menunjukkan bahwa wilayah pada klaster lima yaitu Kabupaten Boyolali merupakan daerah dengan hasil produksi daging sapi, susu sapi, dan kulit sapi/kerbau tertinggi. Pada variabel produksi daging kambing, daging domba dan kulit kambing/domba tertinggi ada pada klaster empat. Pada klaster tiga yaitu Kabupaten Kebumen merupakan daerah dengan hasil produksi telur ayam kampung dan produksi telur itik tertinggi. Produksi telur ayam petelur tertinggi ada pada klaster tujuh yaitu daerah Kabupaten Kendal.

Klaster tujuh yaitu wilayah Kabupaten Kendal merupakan klaster yang memiliki banyak variabel dengan penghasil produksi hasil ternak terendah yaitu produksi daging sapi, daging kambing, susu sapi, kulit sapi/kerbau dan kulit kambing/domba. Variabel produksi daging kambing, telur ayam petelur dan susu sapi terendah ada pada klaster tiga yaitu daerah Kabupaten Kebumen. Variabel produksi telur itik terendah ada pada klaster satu dan variabel produksi telur ayam kampung terendah ada pada klaster lima yaitu daerah Kabupaten Boyolali.

5. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah diuraikan pada bab sebelumnya, maka diperoleh kesimpulan sebagai berikut:

1. Pengklasteran menggunakan metode *k-error* dan *k-medoids* untuk $k = 2,3,4,5,6$ dan 7 diperoleh klaster yang optimal pada $k = 7$ dimana nilai Rasio Simpangan Baku = 0,773 dan WB Indeks = 0,531. Berdasarkan hasil pengklasteran pada 2 metode tersebut didapatkan bahwa metode yang terbaik yang digunakan pada data jumlah produksi ternak di Provinsi Jawa Tengah tahun 2020 adalah metode *k-medoids*.
2. Berdasarkan profilisasi hasil analisis klaster dengan 7 klaster, diketahui bahwa klaster 3 produksi hasil ternak yang menonjol adalah produksi telur ayam kampung dan telur itik. Untuk klaster 4 jenis produksi hasil ternak yang menonjol adalah produksi daging kambing, daging domba, dan kulit kambing/domba. Untuk klaster 5, jenis produksi hasil ternak yang menonjol adalah produksi daging sapi, susu sapi dan kulit sapi/kerbau. Untuk klaster 1,2,6 dan 7 banyak diisi oleh jenis produksi hasil ternak yang sedang dan rendah, pada klaster 3 juga terdapat jenis produksi hasil ternak yang merupakan produksi terendah daripada klaster lain yaitu produksi daging domba, telur ayam petelur dan susu sapi, sehingga pemerintah dan para peternak perlu berupaya dan menentukan kebijakan baik berupa penyuluhan, pembinaan, pemberian bantuan berupaya bibit unggul atau pakan ataupun kebijakan lain supaya bisa meningkatkan hasil produksi ternak didaerah tersebut.

- **Saran**

1. Penelitian selanjutnya dapat dilakukan dengan menggunakan uji validitas lain untuk membandingkan metode *k-medoids* dengan metode *k-error* seperti uji validitas *Davies-Bouldin Index*, *C-Index* atau uji validitas lain.
2. Pemerintah selaku pengambil kebijakan diharapkan dapat melihat potensi-potensi yang terdapat dalam suatu wilayah terkait hasil produksi peternakan serta dapat memberikan perhatian lebih untuk wilayah-wilayah yang mempunyai hasil produksi peternakan yang tinggi maupun rendah supaya dapat menambah nilai pendapatan Provinsi Jawa Tengah.

DAFTAR PUSTAKA

- Badan Pusat Statistika Jawa Tengah. 2021. *Provinsi Jawa Tengah dalam Angka 2021*. Semarang: Badan Pusat Statistika Jawa Tengah.
- Gujarati, D. 2009. *Dasar-dasar Ekonometrika Jilid 2*. Jakarta: Erlangga.
- Kalkstein LS, Tan G, dan Skindlov JA. 1987. *An Evaluation of Three Clustering Procedures for Use in Synoptic Climatological Classification*. *Journal of Climate and Applied Meteorology* Vol. 26 No. 2: Hal. 717-730.
- Kumar M, Patel NR. 2005. *Clustering Data With Measurement Errors*. New Jersey (US): Rucor Research Report of Rutgers University.
- Mattjik AA, Sumertajaya IM. 2011. *Sidik Peubah Ganda dengan Menggunakan SAS*. Bogor: IPB Press.
- Widarjono, A. 2010. *Analisis Statistika Multivariat Terapan Edisi Pertama*. Yogyakarta: UPP STIM YKPN.
- Zhao Q. 2012. *Clustering Validity in Clustering Methods*. Joensuu (FI): University of Eastern Finland.