

PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN INDIKATOR PETERNAKAN DENGAN MENGGUNAKAN ALGORITMA K-MEANS DAN K-MEDOIDS

Stefanus Arman^{1*}, Kris Suryowati², Maria Titah Jatipaningrum³

^{1,2,3}Program Studi Statistika, Fakultas Sains dan Teknologi Informasi, Universitas AKPRIND Indonesia

Email: stfnsarman@gmail.com,

*corresponding author

Abstract: Livestock in Indonesia, it is one of the sectors that really helps in economic growth and contributes to fulfilling nutrition. Livestock is divided into large livestock and small livestock, which are included in the large livestock group, namely cows (dairy/beef), buffalo and horses, while the small livestock group consists of pigs, goats, sheep, rabbits and poultry, namely ducks, chickens and quail. To support livestock productivity in Indonesia, data is needed that describes the distribution of existing livestock. In this research, grouping of provinces in Indonesia was carried out using the K-Means and K-Medoids Clustering algorithms to partition existing data in the form of one or more clusters based on Euclidean distance. The results of K-Means Clustering grouping obtained cluster 1 with 3 members 3 Province, and cluster 2 has members 1 Province. Meanwhile, the K-Medoids Clustering grouping obtained cluster 1 with 30 provinces, cluster 2 with 1 province, and cluster 3 with 3 provinces. The best method comparison result is the K-Means algorithm Clustering with $k = 2$, with a Silhouette Index (SI) value of 0,78 and the Davies-Bouldin Index (DBI) value is 0,21. Cluster profiling results for the K-Medoids Clustering method are divided into 2 categories namely high and while the high category is in cluster 2 and the medium category is in the cluster 1.

Keywords: Livestock, K-Means, K-Medoids, Silhouette Index, Davies-Bouldin Index.

Abstrak: Peternakan di Indonesia merupakan salah satu sektor yang sangat membantu dalam pertumbuhan ekonomi serta berkontribusi dalam pemenuhan gizi. Peternakan terbagi menjadi ternak besar dan ternak kecil, yang termasuk kedalam kelompok ternak besar yaitu sapi (perah/potong), kerbau dan kuda, sedangkan kelompok ternak kecil terdiri dari babi, kambing, domba, kelinci serta ternak unggas yakni itik, ayam dan burung puyuh. Dalam menunjang produktivitas peternakan di Indonesia diperlukan data yang menggambarkan persebaran ternak yang ada. Dalam penelitian ini pengelompokan Provinsi di Indonesia dilakukan menggunakan algoritma K-Means dan K-Medoids untuk mempartisi data yang ada dalam bentuk satu atau lebih *cluster* berdasarkan jarak *Euclidean*. Hasil pengelompokan K-Means Clustering diperoleh *cluster* 1 beranggotakan 33 Provinsi, dan *cluster* 2 beranggotakan 1 Provinsi. Sedangkan pengelompokan K-Medoids Clustering diperoleh *cluster* 1 beranggotakan 30 Provinsi, *cluster* 2 beranggotakan 1 Provinsi, dan *cluster* 3 beranggotakan 3 Provinsi. Hasil perbandingan metode terbaik adalah algoritma K-Means Clustering dengan $k = 2$ dengan nilai *Silhouette Index* (SI) sebesar 0,78 dan nilai *Davies-Bouldin Index* (DBI) sebesar 0,21. Hasil *profiling cluster* untuk metode K-Medoids Clustering dibagi menjadi 2 kategori yaitu tinggi dan sedang dimana untuk kategori tinggi berada pada *cluster* 2 dan kategori sedang berada pada *cluster* 1.

Kata kunci: Peternakan, K-Means, K-Medoids, Silhouette Index, Davies-Bouldin Index.

1. Pendahuluan

Peternakan merupakan kegiatan memelihara hewan ternak untuk dibudidayakan dan mendapatkan keuntungan dari kegiatan tersebut (Rasyaf, 2002). Peternakan merupakan salah satu subsektor pertanian yang menjadi salah satu sumber pertumbuhan ekonomi yang cukup potensial dalam upaya pembangunan ekonomi.

Peternakan terbagi menjadi ternak besar dan ternak kecil, yang termasuk kedalam kelompok ternak besar yaitu sapi (perah/potong), kerbau dan kuda, sedangkan kelompok ternak kecil terdiri dari babi, kambing, domba, kelinci serta ternak unggas yakni itik, ayam dan burung puyuh (Putri, 2019). Subsektor peternakan memiliki nilai strategis khususnya dalam pemenuhan

protein hewani bagi masyarakat yang dapat diperoleh dari komoditas utamanya seperti daging, telur dan susu yang sangat berperan dalam rangka pemenuhan kecukupan gizi.

Dalam menunjang produktivitas peternakan di Indonesia diperlukan data yang menggambarkan persebaran ternak yang ada. Salah satu upaya dalam menunjang produktivitas dari sektor peternakan yaitu dengan mengelompokkan kedalam kelompok-kelompok berdasarkan kesamaan karakteristik yang dimiliki dari setiap wilayah atau provinsi di Indonesia. Salah satu cara untuk melakukan pengelompokan ini yaitu dengan analisis *cluster*. Objek yang mempunyai kemiripan akan dikumpulkan menjadi *cluster* yang sama dan objek dengan karakteristik berbeda akan dipisahkan ke *cluster* yang berbeda (Hariyanto dan Shita, 2018).

Penelitian ini menerapkan metode algoritma *K-Means* dan *K-Medoids* untuk mengelompokkan provinsi di Indonesia berdasarkan indikator peternakan. Pada penelitian ini data data peternakan yang digunakan adalah data ternak besar. Secara umum penelitian ini bertujuan untuk mengetahui persebaran ternak besar di setiap provinsi di Indonesia dan mengetahui hasil pengelompokan dari kedua metode yang digunakan serta melakukan perbandingan untuk menentukan keakuratan dari kedua metode tersebut. Dari penelitian ini diharapkan mampu memberikan gambaran persebaran ternak besar pada setiap provinsi di Indonesia, agar provinsi dengan tingkat persebaran ternak besar yang masih sedikit dapat diperhatikan oleh pemerintah.

2. Metode

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari *website* resmi Badan Pusat Statistik (BPS). Data yang diambil merupakan data ternak besar untuk 34 provinsi di Indonesia tahun 2022.

a. Analisis Deskriptif

Statistik deskriptif merupakan ilmu dalam statistika yang mempelajari pengumpulan data, penyajian data dan peringkasan data, sehingga data-data tersebut menjadi sebuah informasi yang mudah dipahami dan berguna. Hasil statistika deskriptif hanya berhubungan dengan kegiatan menguraikan atau memberikan keterangan-keterangan secara umum sehingga tidak banyak kesimpulan yang dapat diambil.

b. Analisis Cluster

Analisis cluster merupakan pengelompokan objek atau kasus menjadi kelompok-kelompok yang lebih kecil dimana setiap kelompok berisi objek yang mirip satu sama lain (Supranto, 2004). Dalam analisis cluster terdapat dua asumsi yang harus dipenuhi yaitu:

1) Sampel Representatif

Sampel representatif artinya seberapa besar sampel mewakili populasi. Sampel ini dipilih sedemikian rupa sehingga memiliki karakteristik yang merupakan representasi yang baik dari seluruh populasi.

2) Tidak Terjadi Multikolinearitas

Uji multikolinearitas adalah uji yang dilakukan untuk memastikan apakah di dalam sebuah model regresi ada interkorelasi atau kolinearitas antar variabel bebas. Interkorelasi adalah hubungan yang linear atau hubungan yang kuat antara satu variabel bebas atau variabel prediktor dengan variabel prediktor lainnya di dalam sebuah model regresi. Penanggulangan pada data yang terdapat multikolinearitas dapat dilakukan dengan analisis komponen utama *Principal Component Analysis* (PCA). *Principal Component Analysis* (PCA) dilakukan dengan transformasi secara linear pada variabel asal yang saling berkorelasi yang mana hasilnya adalah komponen utama atau *principal component* (Alwi dan Hasrul, 2018).

c. Jarak Euclidean

Pendekatan ukuran jarak digunakan dalam menjelaskan atau mengukur kemiripan atau kesamaan antar objek. Menurut (Ulinuha dan veriani, 2020), jarak *Euclidean* adalah besar suatu jarak pada garis lurus yang menghubungkan antar objek. Berikut merupakan persamaan untuk perhitungan jarak *Euclidean*.

$$d_{i,j} = \sqrt{\sum_{r=1}^p (x_{ir} - x_{jr})^2} \tag{1}$$

keterangan:

- $d_{i,j}$: Jarak antara data i dengan data j , dengan $i, j = 1, 2, \dots, n$
- x_{ir} : Data ke- i pada variabel ke- r
- x_{jr} : Data ke- j pada variabel ke- r , dengan $r = 1, 2, \dots, p$
- n : Jumlah data
- p : Jumlah Variabel

d. Metode Silhouette

Metode ini digunakan dalam menentukan jumlah *cluster* optimal. Metode ini menggunakan nilai *silhouette coefficient* sebagai penentu k optimalnya. Tahapan perhitungan *silhouette coefficient* adalah sebagai berikut (Anggara, 2016):

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \tag{2}$$

dengan:

- a_i : Rata-rata jarak observasi ke- i dengan observasi lain dalam satu *cluster*
- b_i : Rata-rata jarak observasi ke- i dengan observasi pada *cluster* terdekat

e. Metode Elbow

Metode *elbow* merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah *cluster* terbaik dengan cara melihat persentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik (Rahman, 2017). Pada metode *elbow* nilai *cluster* terbaik yang akan diambil dari nilai *Sum of Square Error* (SSE) yang mengalami penurunan yang signifikan. Berikut merupakan perhitungan nilai SSE:

$$SSE = \sum_{i=1}^n (x_i - \bar{x})^2 \tag{3}$$

keterangan :

- SSE : *Sum Square Error*
- n : Banyaknya data
- x_i : Data pengamatan ke- i , $i = 1, 2, \dots, n$
- \bar{x} : Nilai rata-rata

f. Metode K-Means

Algoritma *K-Means* merupakan salah satu algoritma dengan partitional, karena *K-Means* didasarkan pada penentuan jumlah awal kelompok dengan mendefinisikan nilai centroid awalnya (Madhulatha, 2012). *K-Means* merupakan salah satu metode analisis cluster non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok berdasarkan karakteristik. Adapun langkah-langkah untuk *K-Means Clustering* adalah sebagai berikut (Rahman, 2017):

- a. Menentukan jumlah *cluster* yang akan dibentuk.
- b. Membangkitkan k *centroid* (titik pusat *cluster*) secara acak diantara objek-objek yang ada.
- c. Menghitung jarak setiap data terhadap masing-masing *centroid* menggunakan jarak *Euclidean* dengan Persamaan (1).
- d. Mengelompokkan data berdasarkan jarak terdekat antara data dengan *centroid*.
- e. Menghitung pusat *cluster* baru dengan cara menghitung rata-rata dari *cluster* yang bersangkutan.
- f. Ulangi Langkah 3 dan 5 hingga anggota *cluster* tidak ada yang berubah atau berpindah.

g. Metode *K-Medoids*

K-Medoids merupakan algoritma yang digunakan untuk menemukan medoids di dalam sebuah kelompok yang merupakan titik pusat dari suatu kelompok. Algoritma *K-Medoids* menggunakan metode pengelompokan partisi untuk mengelompokkan sekumpulan n objek menjadi sejumlah k pengelompokan. Menurut (Brinty, 2019), tahapan *K-Medoids Clustering* adalah sebagai berikut:

- Tentukan jumlah *cluster* (k) yang diinginkan.
- Secara acak pilih medoid awal sebanyak k dari n data.
- Menghitung jarak masing-masing objek ke medoid sementara, kemudian tandai jarak terdekat objek ke medoid dan hitung totalnya.
- Lakukan iterasi medoid.
- Hitung total simpangan (S) yang merupakan selisih total cost baru-total cost lama..
- Ulangi langkah 3 sampai 5 dan hentikan jika sudah tidak terjadi perubahan anggota medoids.

h. *Silhouette Index*

Silhouette Index (SI) digunakan untuk mengetahui seberapa baik suatu objek ditempatkan dalam suatu *cluster* yang terbentuk dapat diketahui dengan melihat nilai *silhouette index* maka dilakukan validasi (Trayasiwi, 2015). Rentang nilai SI adalah -1 hingga +1, hasil *cluster* yang baik ditunjukkan dengan nilai *silhouette* yang mendekati rentang nilai 1. berikut merupakan perhitungan *silhouette index*:

$$SI = \frac{\sum_{i=1}^n SC_i^j}{n} \quad (4)$$

dimana:

SC_i^j : *Silhouette coefficient*
 n : Jumlah data

i. *Davies-Bouldin Index*

Menurut (Kartikasari, 2021), *Davies-Bouldin Index* (DBI) adalah rasio dari jumlah jarak intra-cluster dan jarak inter-cluster harus rendah. Oleh karena itu, nilai DBI yang lebih rendah diperlukan agar menunjukkan hasil pengelompokan yang baik.

$$DBI = \frac{1}{k} \sum_{e=1}^k D_e \quad (5)$$

dengan:

$D_e: \max_{f \neq e} R_{ef}$, $ef = 1, 2, \dots, K$

j. *Profiling Cluster*

Profiling cluster digunakan untuk melihat nilai rata-rata dari anggota masing-masing variabel yang ada pada setiap pengelompokan yang kemudian akan didapatkan karakteristik setiap *cluster*. *Profiling cluster* dihitung menggunakan rata-rata setiap *cluster* pada setiap variabel (Azuri dkk., 2016).

3. Hasil dan Pembahasan

a. Analisis Deskriptif

Berdasarkan studi kasus ternak besar di Indonesia tahun 2022 memiliki karakteristik yang disajikan dalam analisis deskriptif. Pada variabel kerbau, jumlah ternak kerbau tertinggi sebesar 172.850 ekor berada pada provinsi Nusa Tenggara Timur. Jumlah ternak kerbau terendah sebesar 0 ekor berada pada provinsi Sulawesi Utara, Gorontalo dan Papua Barat. Jumlah ternak kuda tertinggi sebesar 157.755 ekor berada pada provinsi Sulawesi Selatan. Jumlah ternak kuda terendah sebesar 1 ekor berada pada provinsi Kepulauan Riau. Jumlah ternak sapi perah tertinggi sebesar 314.385 ekor berada di provinsi Jawa Timur. Jumlah ternak sapi perah terendah sebesar 0 ekor berada di provinsi Bali, Kalimantan Tengah, Gorontalo, Sulawesi Barat,

Maluku, Maluku Utara, Papua Barat dan Papua. Jumlah ternak sapi potong tertinggi sebesar 5.070.24 ekor berada di provinsi Jawa Timur. Jumlah ternak sapi potong terendah sebesar 1.751 ekor berada di provinsi DKI Jakarta.

b. Asumsi Cluster

1) Sampel Representatif

Pada penelitian ini tidak dilakukan uji asumsi sampel representatif, karena pada asumsi dilakukan jika penelitian menggunakan data sampel yang mewakili populasi, sedangkan pada penelitian ini data yang digunakan adalah data populasi

2) Tidak Terjadi Multikolinearitas

Adanya hubungan linear antara variabel dapat diketahui dari nilai koefisien korelasi antar variabel. Nilai koefisien *Pearson* yang melebihi 0,8 menandakan bahwa terjadi multikolinearitas (Alwi dkk., 2018). Berdasarkan hasil analisis menggunakan *software R-Studio* pada Lampiran 3 didapatkan nilai korelasi antar variabel yang disajikan pada Tabel 1:

Tabel 1 Nilai Korelasi Antar Variabel

Variabel	Kerbau	Sapi perah	Sapi potong	Kuda
Kerbau	1.0000	0.6575	0.0627	0.3168
Sapi perah	0.6575	1.0000	-0.0123	0.4073
Sapi potong	0.0627	-0.0123	1.0000	0.8514
Kuda	0.3168	0.3073	0.8514	1.0000

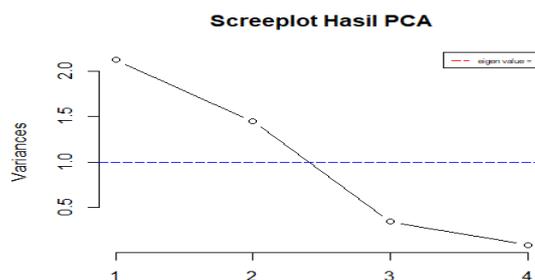
Berdasarkan Tabel 1, diketahui terjadi multikolinearitas dikarenakan nilai koefisien variabel kuda dan sapi potong yakni $0,8514 > 0,8$, sehingga perlu diatasi agar dapat dilanjutkan ke proses berikutnya.

Principal Component Analysis (PCA) merupakan salah satu solusi untuk mengatasi adanya multikolinearitas (Alwi dan Hasrul, 2018). *Principal Component Analysis* (PCA) akan memilih variabel-variabel baru yang disebut *principal components*. Berikut merupakan hasil PCA pada Lampiran 5 yang disajikan pada Tabel 2:

Tabel 2 Hasil PCA

	PC1	PC2	PC3	PC4
Kerbau	-0,45	0,52	-0,73	-0,05
Kuda	-0,43	0,56	0,68	-0,22
Sapi perah	-0,49	-0,56	-0,06	-0,67
Sapi potong	-0,61	-0,32	0,11	0,71

Nilai standar deviasi PCA juga dapat dilihat menggunakan *screeplot eigenvalue* dengan batasan tinggi = 1. Hasil *screeplot* dapat dilihat pada Gambar 1 dan Lampiran 6 berikut:



Gambar 1 *Screeplot Eigenvalue*

Berdasarkan hasil pada Gambar 1 didapatkan dua faktor dengan batasannya adalah 1. Dengan demikian diperoleh nilai PCA baru untuk setiap provinsi. Nilai PCA baru dapat dilihat pada Tabel 3 dan Lampiran 7:

Tabel 3 Data Baru

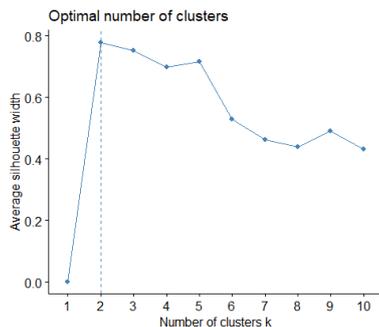
Provinsi	PC1	PC2
ACEH	-0,532	0,986
SUMATERA UTARA	-0,681	0,521
SUMATERA BARAT	-0,152	0,587
RIAU	0,511	0,078
⋮	⋮	⋮
PAPUA BARAT	0,926	-0,236
PAPUA	0,846	-0,202

c. Jarak Euclidean

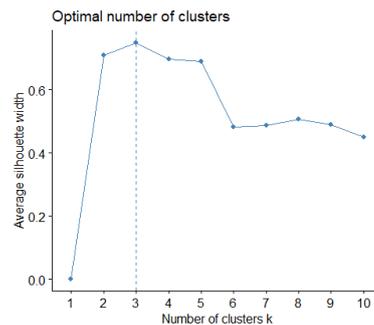
Berdasarkan data ternak besar pada setiap provinsi di Indonesia, maka akan dihitung jarak terdekat menggunakan nilai *Euclidean* untuk mengukur kesamaan antar objek. Berdasarkan hasil output *software R-Studio* diperoleh bahwa jarak terdekat sebesar 0,004 antara Provinsi DKI Jakarta dan Provinsi Kep. Riau, sedangkan jarak terjauh sebesar 7,964 antar Provinsi Jawa Timur dan Provinsi Sulawesi Selatan.

d. Metode Silhouette

Metode ini digunakan dalam menentukan jumlah *cluster* optimal. Berdasarkan hasil output *software R-Studio*, berikut merupakan grafik metode *silhouette* untuk menentukan *cluster* optimal untuk metode *K-Means* dan *K-Medoids* yang disajikan pada Gambar 2 dan 3:



Gambar 3 Grafik Silhouette untuk K-Means

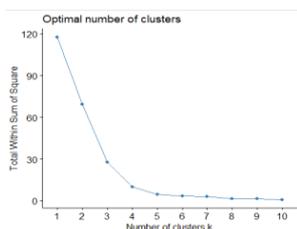


Gambar 2 Grafik Silhouette untuk K-Medoids

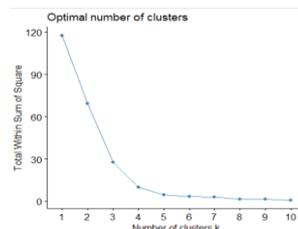
Penentuan nilai k dengan metode *silhouette* dengan pendekatan rata-rata nilai tinggi atau garis putus-putus pada grafik. Berdasarkan Gambar 2 diketahui nilai k optimal untuk metode *K-Means* adalah 2. Dari Gambar 3 diketahui nilai k optimal untuk metode *K-Medoids* adalah 3.

e. Metode Elbow

Penentuan nilai k dengan metode *elbow* dengan grafik yang membentuk siku. Berdasarkan hasil output *software R-Studio* pada Lampiran 10, berikut adalah grafik metode *Elbow* untuk menentukan jumlah *cluster* optimal untuk metode *K-Means* dan *K-Medoids* yang disajikan pada Gambar 4 dan Gambar 5:



Gambar 5 Grafik Elbow untuk K-Means



Gambar 4 Grafik Elbow untuk K-Medoids

Dari Gambar 4 nilai k optimal dari metode K -Means adalah 3, dan nilai k optimal untuk metode K -Medoids adalah 3.

f. Metode K -Means

Berdasarkan data penelitian, berikut adalah perhitungan algoritma K -Means clustering dengan bantuan software R -Studio dan $Microsoft Excel$ dan dikelompokan berdasarkan jarak terdekat disajikan pada Tabel 4:

Tabel 4 Pengelompokan Data Berdasarkan Jarak Terdekat

C_1	Data: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34
C_2	Data: 15

Pada Tabel 4 dapat diketahui bahwa $cluster$ 1 terdiri dari 33 anggota $cluster$ yaitu Provinsi Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua. Sedangkan pada $cluster$ 2 terdiri dari 1 anggota $cluster$ yaitu Provinsi Jawa Timur.

g. Metode K -Medoids

Berdasarkan data penelitian, berikut adalah perhitungan algoritma K -Means clustering dengan bantuan software R -Studio dan $Microsoft Excel$ dan dikelompokan berdasarkan jarak terdekat disajikan pada Tabel 5:

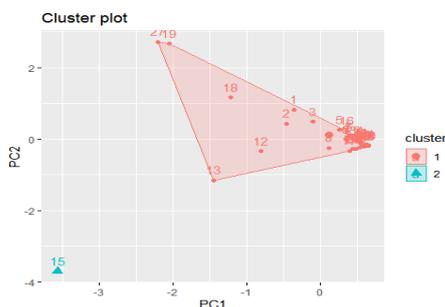
Tabel 5 Pengelompokan Data Berdasarkan Jarak Terdekat

C_1	Data: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17, 20, 21, 22, 23, 24, 25, 26, 28, 29, 30, 31, 32, 33, 34
C_2	Data: 15
C_3	Data: 18, 19, 27

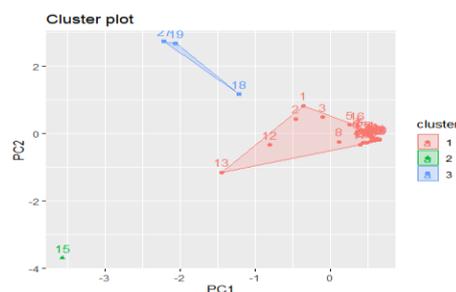
Pada Tabel 5 dapat diketahui bahwa $cluster$ 1 terdiri dari 30 anggota $cluster$ yaitu Provinsi Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Banten, Bali, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua. Pada $cluster$ 2 terdiri dari 1 anggota $cluster$ yaitu Provinsi Jawa Timur. Sedangkan pada $cluster$ 3 terdiri 3 anggota $cluster$ yaitu Provinsi Nusa Tenggara Barat, Nusa Tenggara Timur, Sulawesi Selatan.

h. Plot K -Means dan K -Medoids

Pola sebaran data setiap anggota yang terdapat pada setiap $cluster$ dapat dilihat pada plot di Gambar 6 dan Gambar 7.



Gambar 6 Plot K-Means



Gambar 6 Plot K-Medoids

Pada Gambar 6, warna merah menunjukkan *cluster* 1 dengan jumlah 33 anggota dan warna biru menunjukkan *cluster* 2 dengan jumlah 1 anggota. Pada Gambar 7, warna merah menunjukkan *cluster* 1 dengan jumlah 30 anggota, warna hijau menunjukkan *cluster* 2 dengan jumlah 1 anggota dan warna biru menunjukkan *cluster* 3 dengan jumlah 3 anggota.

i. Penentuan Metode Terbaik

Berdasarkan proses *clustering* dari kedua metode *K-Means* dan *K-Medoids*, kemudian dilakukan perbandingan keakuratan metode mana yang terbaik untuk mengelompokkan provinsi di Indonesia berdasarkan indikator peternakan tahun 2022. Perbandingan ini dilakukan dengan melihat nilai *silhouette index* dan *Davies-Bouldin index* dari kedua metode. Berikut adalah hasil perhitungan *silhouette index* dan *Davies-Bouldin index* berdasarkan output *software R-Studio* pada Tabel 6:

Tabel 6 Nilai *Silhouette Index* dan *Davies-Bouldin Index*

	Metode	
	<i>K-Means</i>	<i>K-Medoids</i>
	$k = 2$	$k = 3$
<i>Silhouette Index</i>	0,78	0,75
<i>Davies-Bouldin Index</i>	0,21	0,35

Berdasarkan Tabel 6 diketahui nilai *silhouette index* pada metode *K-Means* lebih besar dibandingkan perolehan pada metode *K-Medoids*. Hal ini menunjukkan bahwa metode *K-Means* lebih baik dibanding metode *K-Medoids* karena memiliki nilai *silhouette index* lebih besar dengan nilai sebesar 0,78 dengan jumlah *cluster* sebanyak 2 *cluster*. Berdasarkan nilai *Davies-Bouldin index* metode *K-Means* memiliki nilai yang lebih kecil dibanding metode *K-Medoids*, dengan nilai sebesar 0,21 dan jumlah *cluster* sebanyak 3 *cluster*. Berdasarkan perbandingan nilai *silhouette index* dan *Davies-Bouldin index* dapat disimpulkan bahwa metode *K-Means* merupakan metode terbaik dalam mengelompokkan Provinsi di Indonesia berdasarkan indikator peternakan tahun 2022.

j. Profiling Cluster

Profiling cluster dilakukan guna melihat karakteristik *cluster* yang terbentuk. Berikut adalah *profiling cluster* pengelompokan Provinsi di Indonesia berdasarkan indikator peternakan tahun 2022 yang diperoleh dengan *software R-Studio* pada Lampiran 19 yang disajikan pada Tabel 7 berikut:

Tabel 7 *Profiling Cluster*

	Kerbau	Kuda	Sapi perah	Sapi potong
C_1	34.803	11.620	8.440	410.300
C_2	21.703	10.894	314.385	5.070.240

Dari Tabel 7 diketahui kedua *cluster* tersebut dapat dibagi menjadi wilayah dengan persebaran ternak besar yang tinggi dan sedang. Provinsi atau wilayah dengan rata-rata persebaran ternak sapi perah dan sapi potong yang tinggi terdapat pada *cluster* 2, hal ini dikarenakan rata-rata pada *cluster* 2 lebih tinggi dibanding *cluster* 1. Provinsi dengan persebaran ternak kerbau dan kuda yang tinggi terdapat pada *cluster* 1, hal ini ditandai dengan rata-rata pada *cluster* 1 lebih tinggi dari *cluster* 2.

4. Kesimpulan

Berdasarkan hasil analisis dari kedua metode validasi yang digunakan, metode terbaik dalam mengelompokkan provinsi di Indonesia berdasarkan indikator peternakan tahun 2022 adalah metode *K-Means* dengan banyaknya *cluster* adalah 2. *Cluster* 1 memiliki 33 anggota yaitu Provinsi Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan,

Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua. Sedangkan pada *cluster 2* terdiri dari 1 anggota *cluster* yaitu Provinsi Jawa Timur.

Daftar Pustaka

- Alwi, W., & Hasrul, M., 2018, Analisis Klaster Untuk Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat, *Jurnal MSA*, Vol. 6, 35-42.
- Azuri, F. D., Zulhanif, & Pontoh, S. R., 2016, Pengelompokan Kabupaten/Kota Di Pulau Jawa Berdasarkan Pembangunan Manusia Berbasis Gender Menggunakan Bisceting K-Means, *Prosiding Seminar Nasional MIPA*, Hal 78-83.
- Brinty, T., 2019, *K-Medoids/Partitioning Around Medoids (PAM) - NonHierarchical Clustering with R*, Retrieved from <https://medium.com/@story.of.stats/k-medoids-partitioning-around-medoids-pam-non-hierarchical-clustering-with-r-9d0af590bbc0>.
- Hariyanto, M., & Shita R.T., 2018, *Clustering Pada Data Mining Untuk Mengetahui Potensi Penyebaran Penyakit DBD Menggunakan Metode Algoritma K-Means dan Metode Perhitungan Jarak Euclidean Distance*, *Jurnal SKANIKA*, Vol. 1, No. 1.
- Madhulatha, T.S., 2012, *An Overview on Clustering Methods*. *IOSR Journal of Engineering*, II (4), pp.719-725
- Putri, Pungkas Setya, 2019, Usaha Peternakan, <http://cybex.pertanian.go.id/mobile/artikel/79191/USAHA-PETERNAKAN/>
- Rahman, A. T., 2017, Coal trade data clustering using K-means (case study Pt. Global Bangkit Utama), *Jurnal Teknologi dan Informasi*, Volume 6, Hal 24-31.
- Rasyaf, M., 2002, Bahan Makanan Unggas di Indonesia, Cetakan IX, Jakarta: Kanisius.
- Trayasiwi, G. P., 2015, *Penerapan Metode Klastering Dengan Algoritma K-Means Untuk Prediksi Kelulusan Mahasiswa Pada Program Studi Teknik Informatika Strata Satu*, Semarang: Universitas Dian Nuswantoro.
- Ulinuha, N., & Veriani, R., 2020, Analisis Cluster dalam Pengelompokan Provinsi di Indonesia Berdasarkan Variabel Penyakit Menular Menggunakan Metode Complete Linkage, Average Linkage dan Ward. *Jurnal Nasional Informatika dan Teknologi Jaringan*, Vol. 5.