

# Perbandingan Metode *K-Means*, *K-Medoids*, dan Hierarki untuk Penggerombolan Kabupaten/Kota di Sumatera Utara Berdasarkan Jenis Kekerasan terhadap Perempuan

Ardelle Albani<sup>1</sup>, Budi Susetyo<sup>1‡</sup>, Mohammad Masjkur<sup>1</sup>

<sup>1</sup>Department of Statistics, IPB University, Indonesia

‡corresponding author: budisu@apps.ipb.ac.id

Copyright © 2023 Ardelle Albani, Budi Susetyo, and Mohammad Masjkur. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

Violence is an act that is intentionally by a person or group with the aim of oppressing someone to suffer. The high rate of depression which leads to acts of violence against those around them, including women. This study aims to compare the clustering of districts/cities in North Sumatra based on the type of violence against women using the K-Means, K-Medoids, and Hierarchical Methods, and look at the characteristics of the regional distribution of the clustering results. This study uses data the number of victims of violence against women in 2021 from Sistem Informasi Online Perlindungan Perempuan dan Anak and Badan Pusat Statistik Sumatera which consists of 33 districts/cities and six variables. Evaluation of clustering results using the Davies Bouldin Index and Cophenetic Correlation. The optimal cluster analysis method obtained is K-Medoids with 4 clusters. Cluster 1 has the lowest average value for each type of violence. Group 2 has an average value of physical violence that is higher than the average. Group 3 has a higher average score for the types of psychological violence, trafficking, and neglect than the average. Cluster 4 had a highest average score for each type of violence except trafficking than the average.

**Keywords:** cophenetic correlation, Davies Bouldin Index, hierarchical clustering, non-hierarchical clustering, violence.

---

\*Received: Oct 2023; Reviewed: Nov 2023; Published: Des 2023

## 1. Pendahuluan

Berdasarkan Kamus Besar Bahasa Indonesia, kekerasan adalah tindakan individu atau sekelompok orang yang dapat mengakibatkan cedera atau matinya orang lain dan merusak fisik atau barang orang lain. Kekerasan adalah perbuatan yang dilakukan dengan sengaja oleh seseorang atau kelompok dengan tujuan menindas yang lemah agar terus menerus menderita. Faktor yang dapat menyebabkan terjadinya tindak kekerasan adalah ekonomi, budaya, perselingkuhan, dan kurangnya komunikasi antar keluarga (Johny 2011).

Adanya pandemi Covid-19 menyebabkan krisis ekonomi di Indonesia sehingga masyarakat lebih sulit untuk mencari nafkah. Kesulitan tersebut diakibatkan oleh bertambahnya Pemutusan Hubungan Kerja (PHK) di berbagai industri, meningkatnya persentase pengangguran, menurunnya jumlah pendapatan, bertambahnya beragam kebutuhan keluarga, dan implementasi kebijakan Pembatasan Sosial Berskala Besar (PSBB). Faktor-faktor tersebut mengakibatkan tingginya angka depresi dan putus asa hingga berakibat terjadinya tindak kekerasan terhadap orang-orang di sekitar, termasuk perempuan (Aristi et al. 2021).

Berdasarkan siaran PERS Kemen PPPA (2020), sebelum pandemi tercatat ada 1888 anak perempuan di Indonesia menjadi korban kekerasan, setelah pandemi angka ini meningkat menjadi 5242 anak perempuan. Dilansir dari DataIndonesia.id, Sumatera Utara menjadi provinsi dengan banyaknya kejadian Kekerasan dalam Rumah Tangga (KDRT) tertinggi di Indonesia tahun 2021 dengan mayoritas korban KDRT adalah perempuan (Rizaty 2021). Berdasarkan siaran PERS Komnas Perempuan (2022), data pengaduan ke Komnas Perempuan meningkat secara signifikan sebesar 80 persen dari tahun 2020 ke tahun 2021. Semakin meningkatnya tindak kekerasan, masyarakat harus lebih waspada dan pemerintah juga harus lebih memerhatikan dan memperketat sistem keamanan di wilayah tersebut. Hasil dari analisis penggerombolan (*clustering*) diharapkan dapat menjadi landasan untuk pembuatan kebijakan oleh pemerintah di Sumatera Utara.

Menurut Xu dan Wunch (2008), penggerombolan adalah suatu metode pengelompokan objek ke dalam subset tertentu sehingga karakteristik objek di dalam subset memiliki kemiripan yang tinggi. Menurut Kaufman dan Rousseeuw (2009), metode penggerombolan dapat dibagi menjadi dua jenis, yaitu berhierarki dan tak-berhierarki. Perbedaan dari kedua jenis tersebut adalah banyaknya gerombol pada metode berhierarki ditentukan setelah proses penggerombolan sedangkan banyaknya gerombol pada metode tak-berhierarki ditentukan sebelum proses penggerombolan. Contoh metode berhierarki adalah Pautan Tunggal (*Single Linkage*), Pautan Lengkap (*Complete Linkage*), Pautan Rataan (*Average Linkage*), Metode Centroid, dan Metode Ward. Prinsip perbedaan dari metode tersebut adalah perhitungan jarak antar gerombol. Contoh metode tak-berhierarki adalah *K-Means* dan *K-Medoids*. Prinsip perbedaan dari kedua metode tersebut adalah saat perhitungan nilai centroid yang baru.

Wijayanto dan Luthfi (2021) juga melakukan penelitian pada data indeks pembangunan manusia Indonesia menggunakan *K-Means*, *K-Medoids*, dan Metode Hierarki dan didapatkan metode yang terbaik untuk digunakan pada data tersebut adalah *K-Medoids*. Taram et al. (2019) melakukan penelitian penggerombolan provinsi di Indonesia berdasarkan tingkat kriminalitas menggunakan metode *K-Means* dan Metode Hierarki, didapatkan bahwa metode Pautan Tunggal dengan 3 gerombol lebih baik digunakan pada data tersebut. Saat menggunakan metode *K-Means* dan *K-Medoids*, peneliti menentukan sendiri banyaknya gerombol yang akan dibentuk. Hal

tersebut dapat berisiko karena banyaknya gerombol yang digunakan belum tentu yang paling optimal sehingga diperlukan evaluasi banyaknya gerombol untuk mengetahui apakah banyaknya gerombol yang digunakan sudah tepat. Contoh metode untuk mengevaluasi banyaknya gerombol yaitu *Davies Bouldin Index* (DBI) dan korelasi *Cophenetic*. Penelitian ini bertujuan membandingkan penggerombolan kabupaten/kota di Sumatera Utara berdasarkan jenis kekerasan terhadap perempuan menggunakan metode *K-Means*, *K-Medoids*, dan Metode Hierarki dengan metode evaluasi DBI dan korelasi *Cophenetic*, serta melihat karakteristik persebaran wilayah dari hasil penggerombolan yang diperoleh.

## 2. Metodologi

### 2.1 Bahan dan Data

Penelitian ini menggunakan data banyaknya korban kekerasan terhadap perempuan selama tahun 2021 dengan satuan per 100.000 jumlah penduduk yang terdiri atas 33 kabupaten/kota di Sumatera Utara. Data yang digunakan merupakan data sekunder yang bersumber dari Sistem Informasi Online Perlindungan Perempuan dan Anak (SIMFONI PPPA) pada [sigasumutprov.go.id](http://sigasumutprov.go.id) dan Badan Pusat Statistik (BPS) Sumatera Utara pada [sumut.bps.go.id](http://sumut.bps.go.id). Data ini terdiri dari enam peubah yang dapat dilihat pada Tabel 1.

Tabel 1 Peubah yang digunakan dalam penelitian

Kode	Peubah
$X_1$	Banyaknya korban kekerasan fisik
$X_2$	Banyaknya korban kekerasan psikis
$X_3$	Banyaknya korban kekerasan seksual
$X_4$	Banyaknya korban kekerasan perdagangan orang
$X_5$	Banyaknya korban kekerasan penelantaran
$X_6$	Banyaknya korban kekerasan lainnya

### 2.2 Analisis Komponen Utama

Salah satu syarat analisis penggerombolan adalah tidak adanya multikolinearitas dalam data, yaitu keadaan ketika terdapat hubungan linear antara beberapa atau semua peubah yang digunakan (Sharma 1996). Salah satu cara untuk mendeteksi multikolinearitas adalah dengan menghitung nilai *Variance Inflation Factor* (VIF). Jika nilai VIF lebih dari 10 maka dapat terdeteksi adanya multikolinearitas pada data yang digunakan (Jolliffe 2010). Multikolinearitas dapat diatasi salah satunya adalah dengan melakukan Analisis Komponen Utama. Sebelum melakukan penggerombolan, data yang terdeteksi multikolinearitas dilakukan Analisis Komponen Utama terlebih dahulu.

Menurut Jolliffe (2010), Analisis Komponen Utama atau *Principal Component Analysis* (PCA) adalah suatu metode untuk mengubah data dengan  $p$  peubah yang saling berkorelasi menjadi data  $k$  komponen utama yang saling tidak berkorelasi. Persamaan Komponen Utama adalah sebagai berikut (Mattijk dan Sumertajaya 2011).

$$\mathbf{y} = \mathbf{A} \mathbf{x} \quad (1)$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix}, \mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{p1} & \cdots & a_{pp} \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$$

Dengan  $\mathbf{y}$  adalah vektor komponen utama,  $\mathbf{A}$  adalah matriks dari vektor ciri, dan  $\mathbf{x}$  adalah vektor data asal.

## 2.3 Penggerombolan Berhierarki

Menurut Kaufman dan Rousseeuw (2009), penggerombolan berhierarki adalah metode pengelompokan dimana informasi banyaknya gerombol tidak diperlukan di awal proses penggerombolan. Contoh metode penggerombolan berhierarki adalah Pautan Tunggal (*Single Linkage*), Pautan Lengkap (*Complete Linkage*), Pautan Rataan (*Average Linkage*), Metode *Centroid*, dan Metode *Ward*.

### 2.3.1 Pautan Tunggal

Menurut Johnson dan Wichern (2007), Pautan Tunggal menggunakan jarak terdekat sebagai jarak antar gerombol. Tahapan dengan menggunakan algoritma Pautan Tunggal adalah sebagai berikut (Johnson dan Wichern 2007):

1. Asumsikan setiap objek adalah gerombol.
2. Hitung matriks jarak ( $D$ ) antar gerombol menggunakan persamaan ukuran Jarak Euclidean.
3. Tentukan dua gerombol yang mempunyai jarak paling minimum dan gabungkan, misal gerombol  $U$  dan gerombol  $V$  merupakan dua gerombol dengan jarak paling minimum dan hasil gabungannya menjadi gerombol  $(UV)$ .
4. Perbaharui label matriks jarak ( $D$ ) dengan menghapus baris dan kolom gerombol  $U$  dan gerombol  $V$  lalu menambah baris dan kolom gerombol  $(UV)$ .
5. Hitung jarak antara gerombol  $(UV)$  dengan gerombol lainnya ( $W$ ) menggunakan persamaan berikut.

$$d_{(UV)W} = \min\{d_{(UW)}, d_{(VW)}\} \quad (2)$$

Dengan  $d_{(UV)W}$  adalah jarak antara gerombol gabungan  $(UV)$  dengan gerombol lainnya ( $W$ ),  $d_{(UW)}$  adalah jarak antara gerombol  $U$  dengan gerombol  $W$ , dan  $d_{(VW)}$  adalah jarak antara gerombol  $V$  dengan gerombol  $W$ .

6. Ulangi dari langkah 3 hingga semua objek menjadi satu gerombol.

### 2.3.2 Pautan Lengkap

Menurut Johnson dan Wichern (2007), Pautan Lengkap menggunakan jarak terjauh sebagai jarak antar gerombol. Tahapan dengan menggunakan algoritma Pautan Lengkap sama seperti Pautan Tunggal tetapi untuk perhitungan jarak antara gerombol  $(UV)$  dengan gerombol lainnya ( $W$ ) menggunakan persamaan berikut.

$$d_{(UV)W} = \max\{d_{(UW)}, d_{(VW)}\} \quad (3)$$

Dengan  $d_{(U,V)W}$  adalah jarak antara gerombol gabungan  $(UV)$  dengan gerombol lainnya ( $W$ ),  $d_{(UW)}$  adalah jarak antara gerombol  $U$  dengan gerombol  $W$ , dan  $d_{(VW)}$  adalah jarak antara gerombol  $V$  dengan gerombol  $W$ .

### 2.3.3 Pautan Rataan

Menurut Johnson dan Wichern (2007), Pautan Rataan menggunakan rataan jarak seluruh objek suatu gerombol terhadap seluruh objek pada gerombol lainnya sebagai jarak antar gerombol. Tahapan dengan menggunakan algoritma Pautan Rataan sama seperti Pautan Tunggal tetapi untuk perhitungan jarak antara gerombol  $(UV)$  dengan gerombol lainnya ( $W$ ) menggunakan persamaan berikut.

$$d_{(UV)W} = \frac{\sum_{q=1}^{N_{UV}} \sum_{r=1}^{N_W} d_{qr}}{N_{(UV)} N_W} \quad (4)$$

Dengan  $d_{(UV)W}$  adalah jarak antara gerombol gabungan  $(UV)$  dengan gerombol lainnya ( $W$ ),  $d_{qr}$  adalah jarak antara objek ke- $q$  pada gerombol  $UV$  dengan objek

ke- $r$  pada gerombol  $W$ ,  $N_{(UV)}$  adalah banyaknya objek pada gerombol  $(UV)$ , dan  $N_W$  adalah banyaknya objek pada gerombol  $W$ .

### 2.3.4 Metode Centroid

Menurut Kauffman dan Rousseeuw (2009), Metode *Centroid* menggunakan jarak antar *centroid* sebagai jarak antar gerombol. *Centroid* merupakan rataan jarak antar semua objek dalam satu gerombol. Tahapan dengan menggunakan algoritma Metode Centroid sama seperti Pautan Tunggal tetapi untuk perhitungan jarak antara gerombol  $(UV)$  dengan gerombol lainnya  $(W)$  menggunakan persamaan berikut

$$d_{(UV)W} = \frac{N_U}{N_U + N_V} d_{(UW)} + \frac{N_V}{N_U + N_V} d_{(VW)} - \frac{N_U N_V}{(N_U + N_V)^2} d_{(UV)} \quad (5)$$

Dengan  $d_{(UV)W}$  adalah jarak antara gerombol gabungan  $(UV)$  dengan gerombol lainnya  $(W)$ ,  $d_{(UW)}$  adalah jarak antara gerombol  $U$  dengan gerombol  $W$ ,  $d_{(VW)}$  adalah jarak antara gerombol  $V$  dengan gerombol  $W$ ,  $d_{(UV)}$  adalah jarak antara gerombol  $U$  dengan gerombol  $V$ ,  $N_U$  adalah banyaknya objek pada gerombol  $U$ , dan  $N_V$  adalah banyaknya objek pada gerombol  $V$ .

### 2.3.5 Metode Ward

Menurut Johnson dan Wichern (2007), Metode *Ward* bertujuan untuk menghasilkan gerombol dengan keragaman dalam gerombol yang sekecil mungkin. Metode ini menggunakan Jarak Kuadrat Galat (JKG) atau *Sum Square Error* (SSE) sebagai jarak antar gerombol. Tahapan dengan menggunakan algoritma Metode Ward sama seperti Pautan Tunggal tetapi untuk perhitungan jarak antara gerombol  $(UV)$  dengan gerombol lainnya  $(W)$  menggunakan persamaan berikut

$$d_{(UV)W} = \frac{N_U + N_W}{N_U + N_V + N_W} d_{(UW)} + \frac{N_V + N_W}{N_U + N_V + N_W} d_{(VW)} - \frac{N_W}{N_U + N_V + N_W} d_{(UV)} \quad (6)$$

Dengan  $d_{(UV)W}$  adalah jarak antara gerombol gabungan  $(UV)$  dengan gerombol lainnya  $(W)$ ,  $d_{(UW)}$  adalah jarak antara gerombol  $U$  dengan gerombol  $W$ ,  $d_{(VW)}$  adalah jarak antara gerombol  $V$  dengan gerombol  $W$ ,  $d_{(UV)}$  adalah jarak antara gerombol  $U$  dengan gerombol  $V$ ,  $N_U$  adalah banyaknya objek pada gerombol  $U$ ,  $N_V$  adalah banyaknya objek pada gerombol  $V$ , dan  $N_W$  adalah banyaknya objek pada gerombol  $W$ .

## 2.4 Penggerombolan Tak-Berhierarki

Menurut Johnson dan Wichern (2007), Penggerombolan Tak-berhierarki adalah metode pengelompokan dimana informasi banyaknya gerombol diperlukan dan termasuk ke dalam tahapan penggerombolan. Setiap gerombol harus terdiri setidaknya satu objek dan setiap objek harus masuk tepat satu gerombol (Kauffman dan Rousseeuw 2009). Contoh metode Penggerombolan Tak-berhierarki adalah *K-Means* dan *K-Medoids*.

### 2.4.1 K-Means

Menurut Wu (2012), *K-Means* merupakan metode pengelompokan dimana masing-masing gerombol diwakilkan oleh rataan dari setiap objek dalam gerombol

tersebut. Tahapan dengan menggunakan algoritma *K-Means* adalah sebagai berikut (Wu 2012):

1. Tentukan banyaknya gerombol sebanyak  $k$ .
2. Pilih nilai *centroid* awal secara acak sebanyak  $k$ .
3. Hitung jarak antara titik *centroid* dengan titik tiap objek menggunakan persamaan ukuran Jarak Euclidean.
4. Tandai jarak terdekat objek ke *centroid* awal dan kelompokkan menjadi gerombol awal.
5. Tentukan nilai *centroid* baru dari masing-masing gerombol dengan menghitung rataan objek dalam gerombol.
6. Hitung jarak antara titik *centroid* baru dengan titik tiap objek menggunakan persamaan ukuran Jarak Euclidean.
7. Tandai jarak terdekat objek ke *centroid* baru dan kelompokkan menjadi gerombol baru.
8. Kembali ke langkah 5, lakukan pengulangan hingga nilai *centroid* yang dihasilkan tetap dan anggota gerombol tidak berpindah ke gerombol lain.

#### 2.4.2 *K-Medoids*

Menurut Kaufman dan Rousseeuw (2009), *K-Medoids* atau *Partitioning Around Medoids* (PAM) adalah metode pengelompokan dimana setiap gerombol diwakili oleh salah satu objek di dalam gerombol. Tahapan dengan menggunakan algoritma *K-Medoids* adalah sebagai berikut (Maimon dan Rokach 2005; Kaufman dan Rousseeuw 2009):

1. Tentukan banyaknya gerombol sebanyak  $k$ .
2. Pilih nilai *centroid* awal secara acak sebanyak  $k$ .
3. Hitung jarak antara titik *centroid* awal dengan titik tiap objek menggunakan persamaan ukuran Jarak Euclidean
4. Tandai jarak terdekat objek ke *centroid* awal dan kelompokkan menjadi gerombol awal.
5. Hitung total jarak terdekat ( $c$ ).
6. Pilih secara acak objek pada masing-masing gerombol awal sebagai kandidat *centroid* baru.
7. Hitung jarak antara titik *centroid* baru dengan titik tiap objek menggunakan persamaan ukuran Jarak Euclidean
8. Tandai jarak terdekat objek ke *centroid* baru dan kelompokkan menjadi gerombol baru.
9. Hitung total jarak terdekat yang terbaru ( $c'$ ).
10. Hitung total simpangan ( $S$ ) dengan persamaan berikut. Jika  $S$  kurang dari 0, ulangi langkah 6 sampai 10. Jika  $S$  lebih dari 0, iterasi berhenti.

$$S = c' - c \quad (7)$$

### 2.5 Prosedur Analisis

Prosedur analisis data yang dilakukan dalam penelitian ini adalah sebagai berikut.

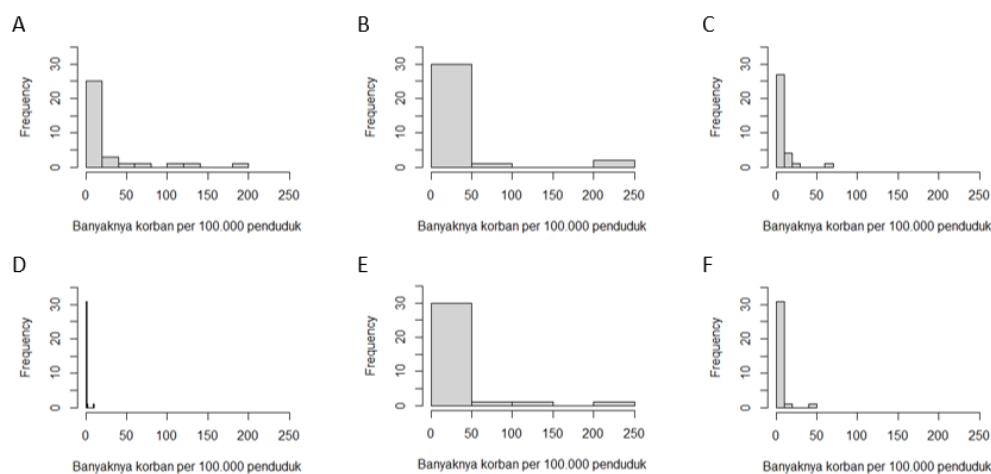
1. Melakukan eksplorasi data dengan menggunakan histogram dan statistik lima serangkai.
2. Melakukan pengujian multikolinearitas.
3. Jika ada multikolinearitas, melakukan Analisis Komponen Utama untuk mengatasinya.
4. Melakukan penggerombolan dengan metode *K-Means*, *K-Medoids*, Pautan Tunggal, Pautan Lengkap, Patuan Rataan, Metode *Centroid*, dan Metode *Ward*

- dengan banyaknya gerombol sebesar 2 hingga 10.
5. Mengevaluasi banyaknya gerombol yang optimal dengan metode DBI dan koefisien korelasi *Cophenetic*.
  6. Melakukan pemetaan persebaran wilayah anggota gerombol berdasarkan hasil penggerombolan yang diperoleh.
  7. Melakukan pemilihan metode terbaik.
  8. Melakukan interpretasi karakteristik masing-masing gerombol pada metode terbaik.

### 3. Hasil dan Pembahasan

#### 3.1 Eksplorasi Data

Gambar 1 menampilkan histogram sebaran masing-masing peubah. Berdasarkan histogram tersebut, terlihat sebaran data menjulur ke kanan dan terdapat kejadian ekstrem di kabupaten/kota tertentu. Kabupaten/kota di Sumatera Utara yang termasuk kejadian ekstrem pada data ini adalah Asahan pada peubah  $X_1$ , Simalungun pada peubah  $X_6$ , Deli Serdang pada peubah  $X_2, X_4$ , dan  $X_5$ , Langkat pada peubah  $X_1$  dan  $X_3$ , serta Medan pada peubah  $X_1, X_2, X_3, X_5$ , dan  $X_6$ .



Gambar 1 Histogram banyaknya korban per 100.000 penduduk. (A) Histogram peubah  $X_1$ . (B) Histogram peubah  $X_2$ . (C) Histogram peubah  $X_3$ . (D) Histogram peubah  $X_4$ . (E) Histogram peubah  $X_5$ . (F) Histogram peubah  $X_6$

#### 3.2 Pengujian Multikolinearitas

Multikolinearitas dapat terdeteksi dengan perhitungan nilai VIF. Jika nilai VIF lebih dari 10 maka data tersebut terindikasi adanya multikolinearitas. Tabel 2 menunjukkan nilai VIF pada data yang digunakan. Terlihat bahwa terdapat beberapa peubah yang memiliki nilai VIF lebih dari 10 sehingga terdeteksi adanya multikolinearitas pada data. Multikolinearitas dapat disebabkan oleh korelasi yang tinggi antar peubah. Oleh karena itu, diperlukan analisis tambahan, yaitu Analisis Komponen Utama. Analisis Komponen Utama bertujuan untuk mengatasi multikolinearitas dengan mereduksi dimensi data menjadi lebih kecil dengan peubah-peubah baru yang saling bebas.

Tabel 2 Nilai VIF tiap peubah

Peubah	VIF
$X_1$	5,316
$X_2$	17,021
$X_3$	11,108
$X_4$	14,073
$X_5$	17,678
$X_6$	19,413

### 3.3 Analisis Komponen Utama

Tabel 3 menunjukkan akar ciri dan proporsi keragaman yang dapat dijelaskan oleh setiap komponen utama. Semakin besar proporsi dari suatu komponen utama maka semakin baik komponen utama tersebut dalam menjelaskan peubah asalnya. Komponen utama tersebut terbentuk dari kombinasi linear vektor data asal dan matriks vektor ciri. Menurut Jolliffe (2010), banyaknya komponen utama yang digunakan dapat ditentukan berdasarkan proporsi kumulatif keragaman total sebesar 70 persen hingga 90 persen dan berdasarkan nilai akar ciri yang lebih dari 1. Berdasarkan Tabel 3, terlihat bahwa dengan 2 komponen utama yaitu PC1 dan PC2, nilai akar ciri sudah lebih dari 1 dan total keragaman yang dapat dijelaskan sebesar 94,200% sehingga data yang digunakan untuk analisis penggerombolan adalah data hasil analisis komponen utama dengan komponen utama sebesar 2.

Tabel 3 Nilai akar ciri dan proporsi keragaman yang dapat dijelaskan oleh tiap komponen utama

Komponen Utama	Nilai Akar Ciri	Proporsi (%)	Proporsi Kumulatif (%)
PC1	2,108	74,740	74,740
PC2	1,100	20,160	94,200
PC3	0,420	2,940	97,140
PC4	0,326	1,770	98,910
PC5	0,208	0,720	99,630
PC6	0,148	0,370	100,000

### 3.4 Hasil Penggerombolan Metode Tak-Berhierarki

Penelitian ini melakukan penggerombolan dengan metode *K-Means* dan *K-Medoids* menggunakan banyaknya gerombol sebesar 2 hingga 10. Penentuan banyaknya gerombol terbaik dilakukan dengan melihat nilai DBI. Semakin kecil nilai DBI menunjukkan bahwa hasil gerombol memiliki keragaman dalam gerombol yang rendah dan keragaman antar gerombol tinggi sehingga semakin baik banyaknya gerombol yang dihasilkan. Berdasarkan evaluasi nilai DBI, banyaknya gerombol yang optimal dengan metode *K-Means* adalah 10 gerombol dengan nilai DBI sebesar 0,502 dan dengan metode *K-Medoids* adalah 4 gerombol dengan nilai DBI sebesar 0,181.

Tabel 4 menunjukkan anggota gerombol hasil metode *K-Means*. Gerombol 1, 2, dan 8 terdiri dari 3 kabupaten/kota, gerombol 3, 5, dan 10 terdiri dari 1 kabupaten/kota, gerombol 4 terdiri dari 2 kabupaten/kota, gerombol 6 dan 9 terdiri dari 7 kabupaten/kota. Tabel 5 menunjukkan anggota gerombol hasil metode *K-Medoids*. Gerombol 1 terdiri dari 28 kabupaten/kota, gerombol 2 terdiri dari 3 kabupaten/kota, gerombol 3 terdiri dari 1 kabupaten/kota, dan gerombol 4 terdiri dari 1 kabupaten/kota. Masing-masing gerombol tersebut memiliki karakteristik masing-masing tiap peubah.

Tabel 4 Anggota gerombol hasil metode *K-Means*

<b>Gerombol</b>	<b>Kabupaten/Kota</b>
1	Humbang Hasundutan, Padang Sidempuan, Gunung Sitoli
2	Labuhanbatu, Dairi, Karo
3	Deli Serdang
4	Serdang Bedagai, Tanjung Balai
5	Labuhanbatu Utara
6	Nias, Tapanuli Utara, Toba, Batubara, Nias Utara, Nias Barat, Sibolga
7	Pakpak Bharat, Samosir, Padang Lawas, Labuhanbatu Utara, Labuhanbatu Selatan
8	Asahan, Simalungun, Langkat
9	Mandailing Natal, Tapanuli Selatan, Tapanuli Tengah, Nias Selatan, Pematang Siantar, Tebing Tinggi, Binjai
10	Medan

Tabel 5 Anggota gerombol hasil metode *K-Medoids*

<b>Gerombol</b>	<b>Kabupaten/Kota</b>
1	Nias, Mandailing Natal, Tapanuli Selatan, Tapanuli Tengah, Tapanuli Utara, Toba, Labuhanbatu, Dairi, Karo, Nias Selatan, Humbang Hasundutan, Pakpak Bharat, Samosir, Serdang Bedagai, Batubara, Padang Lawas, Padang Lawas Utara, Labuhanbatu Selatan, Labuhanbatu Utara, Nias Utara, Nias Barat, Sibolga, Tanjung Balai, Pematang Siantar, Tebing Tinggi, Padangsidimpuan, Gunungsitoli, Binjai
2	Asahan, Simalungun, Langkat
3	Deli Serdang
4	Medan

### 3.5 Hasil Penggerombolan Metode Berhierarki

Penelitian ini melakukan penggerombolan dengan beberapa metode berhierarki menggunakan banyaknya gerombol sebesar 2 hingga 10. Penentuan banyaknya gerombol terbaik dilakukan dengan melihat nilai DBI. Semakin kecil nilai DBI menunjukkan bahwa hasil gerombol memiliki keragaman dalam gerombol yang rendah dan keragaman antar gerombol tinggi sehingga semakin baik banyaknya gerombol yang dihasilkan. Berdasarkan evaluasi nilai DBI, hasil banyaknya gerombol terbaik cukup konstan dengan banyaknya gerombol yang optimal pada metode Pautan Tunggal, Pautan Lengkap, Pautan Rataan, dan Metode *Centroid* adalah 3 gerombol dengan nilai DBI sebesar 0,098. Sedangkan, banyaknya gerombol yang optimal pada metode *Ward* adalah 4 gerombol dengan nilai DBI sebesar 0,183. Pemilihan metode penggerombolan terbaik memerlukan evaluasi tambahan karena terdapat nilai DBI yang sama, salah satu metodenya adalah dengan membandingkan nilai koefisien korelasi *Cophenetic* setiap metode. Metode terbaik dapat dilihat ketika metode tersebut memiliki nilai koefisien korelasi *Cophenetic* tertinggi. Berdasarkan evaluasi nilai koefisien korelasi *Cophenetic* setiap metode, terlihat bahwa metode Pautan Rataan memiliki nilai koefisien korelasi *Cophenetic* tertinggi, yaitu sebesar 0,993.

Hasil penggerombolan metode Pautan Tunggal, Pautan Lengkap, Pautan Rataan, dan Metode *Centroid* memiliki hasil anggota yang sama walaupun memiliki nilai DBI yang sedikit berbeda. Tabel 6 menunjukkan anggota gerombol hasil keempat metode tersebut. Gerombol 1 terdiri dari 31 kabupaten/kota, gerombol 2 dan 3 masing-masing terdiri dari 1 kabupaten/kota. Masing-masing gerombol tersebut memiliki karakteristik masing-masing tiap jenis kekerasan. Sedangkan, hasil penggerombolan metode *Ward* memiliki hasil anggota yang sama dengan metode *K-Medoids*.

Tabel 6 Anggota gerombol hasil metode Pautan Tunggal, Pautan Lengkap, Pautan Rataan, dan metode *Centroid*

<b>Gerombol</b>	<b>Kabupaten/Kota</b>
1	Nias, Mandailing Natal, Tapanuli Selatan, Tapanuli Tengah, Tapanuli Utara, Toba, Labuhanbatu, Dairi, Karo, Nias Selatan, Humbang Hasundutan, Pakpak Bharat, Samosir, Serdang Bedagai, Batubara, Padang Lawas, Padang Lawas Utara, Labuhanbatu Selatan, Labuhanbatu Utara, Nias Utara, Nias Barat, Sibolga, Tanjung Balai, Pematang Siantar, Tebing Tinggi, Padangsidimpuan, Gunungsitoli, Binjai, Asahan, Simalungun, Langkat
2	Deli Serdang
3	Medan

### 3.6 Pemilihan Metode Terbaik

Tabel 7 menunjukkan nilai DBI banyaknya gerombol terbaik tiap metode. Berdasarkan evaluasi nilai DBI dan nilai koefisien korelasi *Cophenetic*, analisis gerombol Pautan Rataan dengan banyaknya gerombol sebesar 3 menghasilkan penggerombolan terbaik dengan nilai DBI yang terkecil sebesar 0,098 dan nilai koefisien korelasi *Cophenetic* yang tertinggi sebesar 0,993. Namun, berdasarkan anggota hasil penggerombolan, metode *K-Medoids* dengan banyaknya gerombol sebesar 4 merupakan metode terbaik karena data ekstrem, yaitu Asahan, Simalungun, dan Langkat membentuk gerombol tersendiri dan selisih nilai DBI dengan metode Pautan Rataan hanya 0,083 lebih tinggi. Berdasarkan Kaufman dan Rousseeuw (2009), jika terdapat data ekstrem, metode *K-Medoids* lebih baik digunakan karena nilai *centroid* didapatkan dari salah satu objek dalam gerombol, bukan berdasarkan rataannya sehingga metode tersebut lebih kekar terhadap data ekstrem. Oleh karena itu, metode terbaik dalam analisis penggombolan ini adalah *K-Medoids* dengan banyaknya gerombol sebesar 4.

Tabel 7 Nilai DBI banyaknya gerombol terbaik tiap metode

<b>Metode Penggerombolan</b>	<b>Banyaknya Gerombol Terbaik</b>	<b>DBI</b>
<i>K-Means</i>	10	0,502
<b><i>K-Medoids</i></b>	<b>4</b>	<b>0,181</b>
Pautan Tunggal	3	0,098
Pautan Lengkap	3	0,098
<b>Pautan Rataan</b>	<b>3</b>	<b>0,098</b>
Metode <i>Centroid</i>	3	0,101
Metode <i>Ward</i>	4	0,183

### 3.7 Interpretasi Hasil Metode Terbaik

Tabel 8 menunjukkan nilai rataan masing-masing peubah tiap gerombol. Berdasarkan hasil tersebut, dibandingkan dengan rataan tiap peubah, terlihat bahwa pada gerombol 1 memiliki nilai rataan yang paling rendah tiap peubah. Gerombol 2 memiliki nilai rataan pada peubah  $X_1$  yang lebih tinggi dibandingkan dengan rataannya. Gerombol 3 memiliki nilai rataan pada peubah  $X_2$ ,  $X_4$ , dan  $X_5$  yang lebih tinggi dibandingkan dengan rataannya. Gerombol 4 memiliki nilai rataan pada peubah  $X_1$ ,  $X_2$ ,  $X_3$ ,  $X_5$ , dan  $X_6$  yang lebih tinggi dibandingkan dengan rataannya. Gerombol 1 dapat dikelompokkan sebagai daerah dengan tingkat kekerasan rendah, gerombol 2 sebagai daerah dengan tingkat kekerasan sedang, gerombol 3 sebagai daerah dengan tingkat kekerasan tinggi, dan gerombol 4 sebagai daerah dengan tingkat kekerasan sangat tinggi.

Tabel 8 Rataan masing-masing peubah tiap gerombol metode *K-Medoids*

<b>Gerombol</b>	<b>Rataan</b>						<b>Rataan Tiap Gerombol</b>
	<i>X</i> <sub>1</sub>	<i>X</i> <sub>2</sub>	<i>X</i> <sub>3</sub>	<i>X</i> <sub>4</sub>	<i>X</i> <sub>5</sub>	<i>X</i> <sub>6</sub>	
1	6,357	3,342	2,434	0,068	1,879	0,540	2,437
2	100,180	44,601	16,080	0,000	41,066	9,301	35,205
3	77,147	212,154	19,287	9,643	212,154	9,643	90,005
4	185,349	247,131	61,783	0,000	135,922	49,426	113,269
<b>Rataan Tiap Peubah</b>	<b>92,258</b>	<b>126,807</b>	<b>24,896</b>	<b>2,428</b>	<b>97,755</b>	<b>17,228</b>	<b>60,229</b>

#### 4. Simpulan

Hasil penggerombolan kabupaten/kota di Sumatera Utara tahun 2021 berdasarkan jenis kekerasan terhadap perempuan menggunakan metode *K-Means*, *K-Medoids*, dan beberapa metode berhierarki memiliki hasil banyaknya gerombol yang berbeda. Metode terbaik dalam analisis penggerombolan ini adalah *K-Medoids* dengan banyaknya gerombol sebesar 4 karena data ekstrem membentuk gerombol tersendiri dan selisih nilai DBI dengan metode Pautan Rataan hanya 0,083. Masing-masing gerombol memiliki karakteristik yang berbeda tiap jenis kekerasan. Dibandingkan dengan rataan tiap peubah, terlihat bahwa pada gerombol 1 memiliki nilai rataan yang paling rendah tiap jenis kekerasan. Gerombol 2 memiliki nilai rataan pada jenis kekerasan fisik yang lebih tinggi dibandingkan dengan rataannya. Gerombol 3 memiliki nilai rataan pada jenis kekerasan psikis, perdagangan orang, dan penelantaran yang lebih tinggi dibandingkan dengan rataannya. Gerombol 4 memiliki nilai rataan yang paling tinggi pada setiap jenis kekerasan kecuali perdagangan orang dibandingkan dengan rataannya.

#### Daftar Pustaka

- Aristi N, Janitra PA, Prihandiri P. 2021. Fokus narasi kekerasan seksual pada portal berita daring selama pandemi covid-19. *Jurnal Kajian Komunikasi*. 9(1):121-137.
- [BPS] Badan Pusat Statistik. Jumlah Penduduk menurut Jenis Kelamin dan Kabupaten/Kota (Jiwa), 2020-2022. [diakses 2023 Mar 8]. <https://sumut.bps.go.id/indicator/12/65/1/jumlah-penduduk-menurut-jenis-kelamin-dan-kabupaten-kota.html>
- Johnson RA, Wichern DW. 2007. *Applied Multivariate Statistical Analysis*. New Jersey (US): Pearson Education, Inc.
- Johny RH. 2011. Tindak pidana kekerasan terhadap perempuan. *Jurnal Dinamika Hukum*. 11(2):214-229.
- Jolliffe IT. 2010. *Principal Component Analysis*. Ed ke-2. Aberdeen (UK): Springer.
- Kauffman L, Rousseeuw PJ. 2009. *Finding Groups in Data: An Introduction to Cluster Analysis*. New Jersey (US): John Wiley and Sons Inc.
- [KBBI] Kamus Besar Bahasa Indonesia. 2016. *Kamus Besar Bahasa Indonesia Daring*. [diakses 2023 Feb 3]. <https://kbbi.kemdikbud.go.id/>
- Kementerian Pemberdayaan Perempuan dan Perlindungan Anak. 2020. Siaran PERS: DPD RI Apresiasi Upaya Kemen PPA Lindungi Anak di Tengah Pandemi. Siaran Pers Nomor: B- 313 /Set/Rokum/MP 01/11/2020. [diakses 2023 Feb 3]. Tersedia dari : <https://www.kemenpppa.go.id/index.php/page/read/29/2962/dpd-riapresias-i-upaya-kemen-pppa-lindungi-anak-di-tengah-pandemi/hqkyl/i1022431.html>
- Komnas Perempuan. 2022. Siaran PERS: Peringatan Hari Perempuan Internasional

- 2022 dan Peluncuran Catatan Tahunan tentang Kekerasan Berbasis Gender terhadap Perempuan. [diakses 2023 Feb 3]. Tersedia dari : <https://komnasperempuan.go.id/siaran-pers-detail/peringatan-hari-perempuan-internasional-2022-dan-peluncuran-catatan-tahunan-tentang-kekerasan-berbasis-gender-terhadap-perempuan>
- Maimon O, Rokach L.2005. *Data Mining and Knowledge Discovery*. Tel Aviv (IL): Springer Science+Business Media, Inc.
- Mattjik AA, Sumertajaya IM. 2011. *Sidik Peubah Ganda dengan Menggunakan SAS*. Bogor (ID): Institut Pertanian Bogor.
- Rizaty MA. 2021. Indonesia catat 7.435 kasus KDRT pada 2021, terbanyak di Sumut. [diakses 2023 Feb 3]. Tersedia dari :<https://dataindonesia.id/Ragam/detail/indonesia-catat-7435-kasus-kdrt-pada2021-terbanyakdisumut#>.
- Sharma S. 1996. *Applied Multivariate Techniques*. Canada (US): John Wiley & Sons, Inc.
- [SIMFONI PPPA] Sistem Informasi Online Perlindungan Perempuan dan Anak. 2021. Banyaknya Korban Kekerasan Terhadap Perempuan Menurut Jenis Kekerasan Dan Kabupaten/Kota Di Provinsi Sumatera Utara 2021. [diakses 2023 Feb 3]. <http://siga.sumutprov.go.id/show/datasorted/non-sipd/122?periode=10>
- Taram NG, Sukarsa IK, Srinadi IG. 2019. Pengelompokan tingkat kriminalitas dengan metode agglomerative dan k-means serta peubah pencirinya. *Jurnal Matematika*. 8(2):102-111.
- Xu R, Wunsch DC. 2008. *Clustering*. Piscataway (NJ): Wiley-IEEE Press.
- Wijayanto AW, Luthfi E. 2021. Analisis perbandingan metode hierarchical, k-means, dan k-medoids clustering dalam pengelompokan indeks pembangunan manusia Indonesia. *Jurnal Inovasi*. 17(4):761-773.
- Wu J. 2012. *Advances in K-Means clustering*. Beijing (CN): Springer-Verlag Berlin Heidelberg.