

PENGELOMPOKAN TINGKAT KRIMINALITAS DENGAN METODE AGGLOMERATIVE DAN K-MEANS SERTA PEUBAH PENCIRINYA

Nyoman Gde Prajnawiweka Ratmasa Taram^{1§}, I Komang Gde Sukarsa², I Gusti Ayu Made Srinadi³

¹Program studi Matematika, Fakultas MIPA – Universitas Udayana [Email: prajna.wiweka@gmail.com]

²Program studi Matematika, Fakultas MIPA – Universitas Udayana [Email: gedesukarsa@unud.ac.id]

³Program studi Matematika, Fakultas MIPA – Universitas Udayana [Email: srinadi@unud.ac.id]

[§]Corresponding Author

ABSTRACT

The purpose of this research is to define the classification of crime rate based on the type of crime and its characterization variable. The data used in this research are secondary data in form of data on the number of crime rates based on the type of crime in 32 regional police in the province with 12 variables. This research is used Agglomerative and K-Means method along with Biplot Analysis. The result shows that there are three classification based upon Single Linkage as the best method which each characteristic of every classification is the crime of family abuse, crime of minor mistreatment and crime of corruption.

Keywords : Classification, Agglomerative, K-Means, Biplot Analysis, Single Linkage, Crime

1. PENDAHULUAN

Tingkat kriminal adalah skala atau gambaran profil yang dapat mengukur kriminalitas gerombol (masyarakat) di suatu daerah. Semakin tinggi angka kriminalitas menunjukkan semakin banyak tindak kejahatan pada masyarakat yang merupakan indikasi bahwa masyarakat merasa semakin tidak aman. Dalam analisis statistika, indikator-indikator tersebut dinyatakan sebagai variabel. Untuk meringkas data dengan peubah banyak, dapat digunakan analisis gerombol untuk mengelompokkan obyek-obyek pengamatan menjadi beberapa gerombol berdasarkan pengukuran peubah-peubah yang diamati.

Gerombol atau *cluster* dapat diartikan sebagai kelompok, pada dasarnya analisis gerombol akan menghasilkan sejumlah kelompok (gerombol). Analisis gerombol merupakan salah satu metode dalam analisis statistika peubah ganda yang digunakan untuk mengelompokkan obyek-obyek ke dalam suatu gerombol berdasarkan karakteristik yang dimiliki, sehingga obyek-obyek dalam satu kelompok memiliki ciri-ciri yang lebih homogen dibandingkan dengan obyek dalam gerombol

lain. Untuk mendapatkan gerombol yang sehomogen mungkin, maka yang digunakan sebagai dasar untuk mengelompokkan adalah kesamaan skor nilai yang dianalisis (Gudono, 2011).

Secara umum terdapat dua metode pengelompokan dalam analisis gerombol yaitu metode hierarki dan metode non-hierarki. Metode penggerombolan berhierarki dibedakan menjadi dua metode yaitu metode penggabungan (*agglomerative*) dan metode pemisah (*divisive*). Dalam metode *agglomerative* terdapat beberapa ukuran jarak antar gerombol, antara lain jarak minimum antar gerombol atau pautan tunggal (*single linkage*), jarak maksimum antar gerombol atau pautan lengkap (*complete linkage*) dan rata-rata dari semua jarak atau pautan rata-rata (*average linkage*) (Everitt, 2011).

Pengelompokan menggunakan metode *single linkage* dengan melihat jarak antar gerombol yang ada kemudian memilih jarak yang paling dekat atau jarak minimum. Menurut Johnson dan Wichern (2007) jarak minimum antara gerombol *i* dengan gerombol *w* dan gerombol *j* dengan gerombol *w* dapat ditulis sebagai

$$D_{(ij)w} = \min\{d_{iw}, d_{jw}\} \quad (1)$$

Pada metode *complete linkage*, jarak antar gerombol ditentukan oleh jarak terjauh atau jarak maksimum antara dua obyek dalam gerombol yang berbeda. Menurut Johnson dan Wichern (2007) jarak maksimum antara gerombol i dengan gerombol w dan gerombol j dengan gerombol w dapat ditulis sebagai:

$$D_{(ij)w} = \max\{d_{iw}, d_{jw}\} \quad (2)$$

dengan d_{iw} dan d_{jw} masing-masing adalah jarak terjauh antara gerombol i dengan gerombol w dan juga gerombol j dan gerombol w .

Pengelompokan dengan metode *average linkage* dengan menghitung jarak antara dua gerombol yang disebut sebagai jarak rata-rata antara semua anggota dalam satu gerombol dengan semua anggota gerombol lain. Jarak rata-rata antara gerombol (ij) dengan gerombol lain yaitu w dapat ditulis sebagai :

$$D_{(ij)w} = \frac{\sum_n \sum_k d_{nk}}{N_{(ij)}N_w} \quad (3)$$

dengan d_{nk} adalah jarak obyek n (pada gerombol i dan gerombol j dengan gerombol tersebut membentuk gerombol tunggal (ij), dengan obyek k pada gerombol w , sedangkan $N_{(ij)}$ dan N_w berturut-turut merupakan jumlah obyek dalam gerombol (ij) dan w

Metode pengelompokan non hierarki digunakan jika banyaknya kelompok yang akan dibentuk sudah diketahui sebelumnya. Metode non hierarki sering disebut sebagai metode *k-means*. Prosedur pada metode non hierarki dimulai dengan memilih pusat kelompok (*centroid*) kemudian menentukan jarak setiap objek dengan setiap *centroid*. Perhitungan jarak minimum antar objek dengan *centroid* menggunakan persamaan jarak *Euclidean*.

Tujuan penelitian ini adalah untuk mengetahui metode pengelompokan terbaik diantara metode *single linkage*, *complete linkage*, *average linkage* dan *k-means* serta perubahan penciri dari kelompok yang dihasilkan.

Metode pengelompokan yang baik mempunyai nilai simpangan baku dalam kelompok (s_w) yang minimum dan nilai simpangan baku antar kelompok (s_b) yang

maksimum (Barakbah & Kohei, 2004). Menurut Bunkers, et al (1996) rumus simpangan baku dalam kelompok adalah sebagai berikut :

$$s_w = \frac{1}{K} \sum_{k=1}^K s_k \quad (4)$$

dengan

K : Banyaknya kelompok yang terbentuk

s_k : Simpangan baku kelompok ke- k

Jika diberikan gerombol D_k , dimana $k = 1, 2, \dots, k$ dan setiap gerombol memiliki anggota x_i , dimana $i = 1, 2, \dots, n$ dan n adalah jumlah anggota dari setiap gerombol, dan \bar{x}_k adalah rata-rata dari gerombol k maka untuk mencari nilai s_k digunakan rumus sebagai berikut:

$$s_k = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_i - \bar{x}_k)^2} \quad (5)$$

Jika terdapat rata-rata variabel dalam setiap gerombol k (\bar{X}_k) maka komponen setiap gerombol berbeda dan simpangan baku antar kelompok (s_b) dapat dirumuskan berikut :

$$s_b = \left[\frac{1}{(K-1)} \sum_{k=1}^K (\bar{X}_k - \bar{X})^2 \right]^{1/2} \quad (7)$$

dengan

K = Banyak kerompok yang terbentuk

\bar{X}_k = Rataan kelompok ke- k

\bar{X} = Rataan keseluruhan kelompok

Menurut Barakbah dan Kohei (2004), pengelompokan yang baik memiliki nilai s_w minimum dan s_b maksimum dapat dirumuskan sebagai berikut :

$$s = \frac{s_w}{s_b} \times 100\% \quad (8)$$

Analisis biplot merupakan dua dimensi dari suatu matriks data X yang menampilkan titik untuk masing-masing vector pengamatan n bersama dengan titik untuk masing-masing p peubah (Rencher, 2002). Dengan demikian biplot dibangun dari suatu matriks data X , dengan masing-masing baris mewakili objek penelitian dan masing-masing kolom mewakili suatu peubah.

$${}_n X_p = \begin{bmatrix} X_{11} & \cdots & X_{1p} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{np} \end{bmatrix} \quad (9)$$

Matriks X memuat peubah-peubah yang akan diteliti sebanyak p dan objek penelitian sebanyak n .

Analisis Biplot didasarkan pada penguraian nilai singular. Suatu matriks $n \times p$ yang dapat diuraikan menjadi :

$$nX_p = U_{n \times r} L_{r \times r} A'_{r \times p} \quad (10)$$

Menurut Jolliffe (2002), terdapat dua nilai α yang digunakan untuk mendefinisikan $G = UL^\alpha$ dan $H' = L^{1-\alpha}A'$, yaitu $\alpha = 0$ dan $\alpha = 1$. Kedua nilai α tersebut berguna dalam interpretasi biplot.

a. Parameter $\alpha = 0$

Jika $\alpha = 0$, maka $G = UL^\alpha$ dan $H' = L^{1-\alpha}A'$, ini berarti bahwa

$$\begin{aligned} X'X &= (GH')'(GH') \\ &= HG'GH' \\ &= HU'UH' \\ &= HH' \end{aligned} \quad (11)$$

Secara keseluruhan pemilihan $\alpha = 0$, akan memberikan gambaran mengenai jarak antara kolom dan akan memberikan informasi mengenai keragaman peubah, hubungan antara peubah serta mendapatkan informasi mengenai pola objek.

b. $\alpha = 1$

Jika $\alpha = 1$, maka $G = UL^\alpha$ dan $H' = L^{1-\alpha}A'$, ini berarti bahwa

$$\begin{aligned} X'X &= (GH')'(GH') \\ &= GH'HG' \\ &= GA'AG' \\ &= GG' \end{aligned} \quad (12)$$

Secara keseluruhan, pemilihan $\alpha = 1$, akan memberikan gambaran tentang jarak antara pasangan baris, sehingga lebih memfokuskan dalam penggambaran kedekatan objek2 daripada keragaman dan hubungan peubah.

2. METODE PENELITIAN

Jenis data yang digunakan dalam penelitian ini adalah data kuantitatif, yaitu data yang berupa angka-angka dan dapat dihitung. Data yang dikumpulkan adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) berupa data jumlah kriminalitas menurut jenis kejahatan pada tahun 2016 yang terdiri dari 32 Polda di provinsi. Variabel yang digunakan dalam penelitian ini merupakan faktor-faktor yang turut mempengaruhi tingkat kriminalitas

meliputi jumlah kejahatan pembunuhan (X1), jumlah kejahatan penganiayaan berat (X2), jumlah kejahatan penganiayaan ringan (X3), jumlah kejahatan KDRT (X4), jumlah kejahatan perkosaan (X5), jumlah kejahatan pencabulan (X6), jumlah kejahatan penculikan (X7), jumlah kejahatan pencurian (X8), jumlah kejahatan narkoba dan psikotropika (X9), jumlah kejahatan penipuan (X10), jumlah kejahatan penggelapan (X11), dan jumlah kejahatan korupsi (X12).

Tahapan analisis yang digunakan dalam penelitian ini adalah sebagai berikut.

1. Penyiapan data

Pada tahap penyiapan data, hal yang perlu diperhatikan adalah satuan pengukuran untuk masing-masing variabel. Bila variabel-variabel yang diperoleh memiliki satuan pengukuran yang sama maka data tidak perlu distandarasi.

2. Melakukan analisis gerombol

Pada tahap ini meliputi :

a. Mengukur jarak kemiripan

Pada tahap ini dilakukan pengukuran kemiripan objek dari data jumlah tingkat kriminalitas menurut jenis kejahatan pada 32 Polda di provinsi menggunakan jarak *Euclidean* dengan menggunakan persamaan

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ki} - x_{kj})^2} \quad (13)$$

Semakin kecil jarak *Euclidean* maka semakin mirip kedua variabel.

b. Menentukan jumlah kelompok yang akan di bentuk.

Banyak kelompok yang dihasilkan akan ditentukan berdasarkan tingkat kriminalitasnya. Menurut Mustofa (2011) tingkat kriminalitas dibagi menjadi tiga, yaitu tingkat kriminalitas rendah, tingkat kriminalitas sedang, dan tingkat kriminalitas tinggi.

c. Metode pengelompokkan yang akan digunakan.

Metode yang akan digunakan dalam penelitian ini adalah metode *Agglomerative* dan metode *K-Means*. Dalam metode *Agglomerative* akan digunakan metode pautan yang terdiri dari

pautan tunggal (*Single Linkage*), pautan lengkap (*Complete Linkage*) dan pautan rata-rata (*Average Linkage*).

3. Pemilihan metode terbaik

Pada tahap ini akan dilakukan pemilihan metode terbaik. Hasil dari tahap ini akan dilanjutkan dengan metode biplot untuk mengetahui peubah pencirinya.

4. Melakukan analisis biplot

Pada tahap ini meliputi :

a. Penentuan parameter

Sesuai dengan tujuan penelitian, maka α yang dipilih adalah $\alpha = 0$, karena pemilihan $\alpha = 0$ memberikan kesesuaian optimal untuk data dan keragaman. Pada $\alpha = 0$, tampilan biplot mampu menggambarkan keragaman peubah, hubungan antar peubah serta mendapatkan informasi mengenai objek, sehingga sesuai dengan tujuan penelitian.

b. Penguraian nilai singular

Melalui penguraian nilai singular akan diperoleh:

- Vektor-vektor eigen yang memberikan informasi mengenai peubah-peubah yang digunakan.
- Nilai-nilai eigen yang berisi informasi tentang keragaman komponen-komponen Bersama dari masing-masing vektor eigen.
- Koordinat-koordinat dimensi peragaan biplot.

c. Pembuatan plot

Pada pembuatan plot ini akan menunjukkan presentase banyaknya informasi yang bias dijelaskan oleh biplot atau nilai kelayakan analisis biplot.

5. Interpretasi Hasil.

3. HASIL DAN PEMBAHASAN

3.1 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang diambil dari Badan Pusat Statistik (BPS) berupa data jumlah kriminalitas menurut jenis kejahatan pada tahun 2016. Peneliti menggunakan 32 provinsi tiap Polda sebagai objek amatan dengan jenis

kejahatan yang diamati sebanyak 12 amatan terangkum dalam lampiran 1. Satuan pengukuran yang digunakan untuk menyatakan jumlah dari setiap jenis kejahatan yang terjadi adalah sama maka tidak perlu dilakukan standarisasi data.

Jumlah kejahatan pembunuhan maksimum terjadi pada provinsi Sumatera Utara, sedangkan jumlah kejahatan pembunuhan minimum terjadi pada provinsi Papua Barat. Jumlah kejahatan pembunuhan berat maksimum terjadi pada Provinsi Sumatera Utara sedangkan jumlah kejahatan pembunuhan berat minimum terjadi pada Provinsi Bali.

3.2 Pengukuran Jarak *Euclidean*

Berdasarkan data yang diperoleh, dengan menggunakan persamaan jarak *Euclidean*, hasil perhitungan jarak *Euclidean* antara Provinsi Aceh dengan Provinsi Sumatera Utara adalah 5.929,24, sedangkan jarak *Euclidean* Provinsi Aceh dengan Provinsi Sumatera Barat adalah 1.064,17. Hal ini menunjukkan bahwa Provinsi Aceh memiliki karakteristik yang lebih mirip dengan Provinsi Sumatera Barat karena jarak *Euclidean* Provinsi Aceh lebih dekat dengan Provinsi Sumatera Barat dibandingkan dengan jarak *Euclidean* Provinsi Aceh ke Provinsi Sumatera Barat.

Demikian pula untuk penafsiran jarak *Euclidean* provinsi lainnya, semakin kecil jarak kedua provinsi maka akan semakin mirip karakteristik dari kedua objek tersebut. Berdasarkan perhitungan yang diperoleh, jarak terdekat adalah 228,17 yang merupakan jarak antara Provinsi Maluku Utara (nomor urut 30) dan Provinsi Papua Barat (nomor urut 31).

3.3 Pengelompokan dengan Metode *Single Linkage*

Proses penggabungan dengan metode pautan *Single Linkage* dengan melihat jarak antar gerombol yang ada kemudian memilih jarak minimum. Berdasarkan jarak penggabungan antara Provinsi Maluku Utara (nomor urut 30) dan Provinsi Papua Barat (nomor urut 31) maka diperoleh jarak antar gerombol (30,31) dengan gerombol lain menggunakan persamaan (2.2).

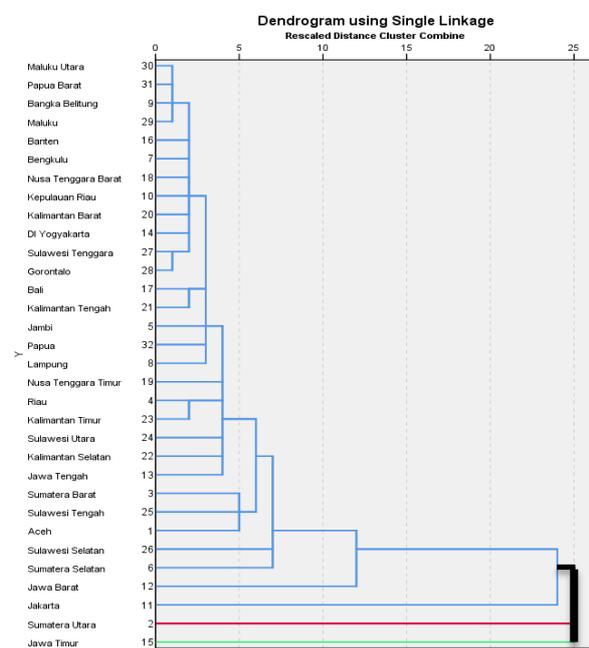
Sebagai contoh diberikan perhitungan antara gerombol (30,31) dengan gerombol 1,2 dan 3.

$$d_{(3031)1} = \min\{d_{301}, d_{3111}\} = \min\{2.624,32, 2.477,76\} = 2.477,76$$

$$d_{(3031)2} = \min\{d_{302}, d_{313}\} = \min\{7.983,40, 7.904,59\} = 7.904,59$$

$$d_{(3031)3} = \min\{d_{303}, d_{314}\} = \min\{2.971,25, 2.797,79\} = 2.797,76$$

Setelah beberapa provinsi bergabung selanjutnya dilakukan lagi perhitungan jarak sehingga terbentuk matriks jarak yang telah tereduksi. Demikian proses dilakukan sehingga seluruh provinsi membentuk tiga gerombol seperti diultrasikan dalam dendrogram pada gambar 1.



Gambar 1. Dendrogram *Single Linkage*

Berdasarkan gambar 1 dibentuk tabel 1 menyajikan hasil pengelompokan provinsi berdasarkan jenis kejahatan dengan metode *Single Linkage*.

Tabel 1. Pengelompokan Provinsi dengan Metode *Single Linkage*

Kelompok	Anggota Kelompok	
1	Aceh	Nusa Tenggara Barat
	Sumatera Barat	Nusa Tenggara Timur
	Riau	Kalimantan Barat
	Jambi	Kalimantan Tengah
	Sumatera Selatan	Kalimantan Selatan
	Bengkulu	Kalimantan Timur
	Lampung	Sulawesi Utara
	Bangka Belitung	Sulawesi Tengah
	Kepulauan Riau	Sulawesi Selatan
	Jakarta	Sulawesi Tenggara
	Jawa Barat	Gorontalo
	Jawa Tengah	Maluku
	DI Yogyakarta	Maluku Utara
	Banten	Papua Barat
	Bali	Papua
2	Sumatera Utara	
3	Jawa Timur	

Sumber : Data Sekunder (2018), diolah

3.4 Pengelompokan Dengan Metode *Complete Linkage*

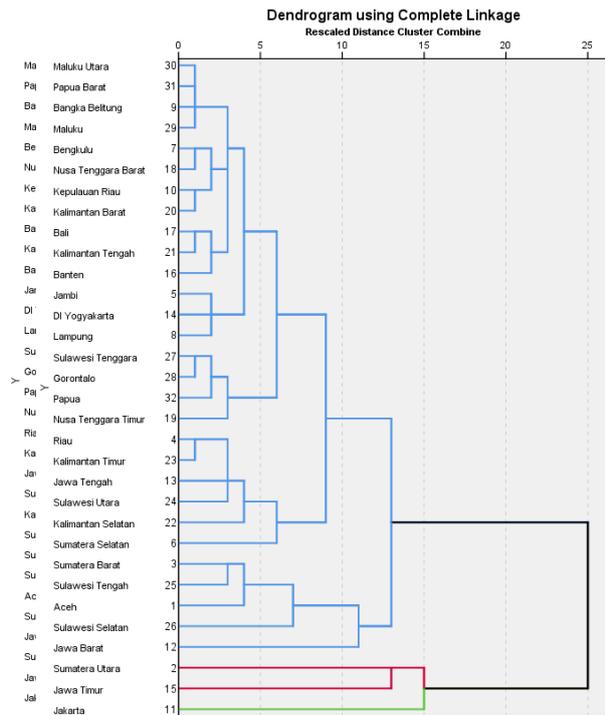
Pada metode *Complete Linkage*, proses penggabungan dilakukan dengan objek yang mempunyai jarak maksimum atau jarak terjauh antar objek. Untuk membentuk gerombol pada metode ini dimulai dengan jarak *Euclidean* yang paling dekat antar provinsi berdasarkan lampiran 2 adalah Provinsi Maluku Utara (nomor urut 30) dan Provinsi Papua Barat (nomor urut 31) maka diperoleh jarak antar gerombol (30,31). Berikut perhitungan jarak maksimum gerombol (30,31) dengan gerombol 1,2 dan 3 sebagai contoh.

$$d_{(3031)1} = \max\{d_{301}, d_{3111}\} = \min\{2.624,32, 2.477,76\} = 2.624,32$$

$$d_{(3031)2} = \max\{d_{302}, d_{313}\} = \min\{7.983,40, 7.904,59\} = 7.983,40$$

$$d_{(3031)3} = \max\{d_{303}, d_{314}\} = \min\{2.971,25, 2.797,79\} = 2.971,25$$

Demikian seterusnya, perhitungan jarak maksimum dilakukan sehingga seluruh provinsi tergabung menjadi tiga gerombol seperti diilustrasikan di dalam gambar 2.



Gambar 2. Dendrogram Complete Linkage

Gambar 2 tampak bahwa provinsi yang tergabung dalam tiga gerombol dituliskan dalam tabel 2.

Tabel 2. Pengelompokan Provinsi dengan Metode Complete Linkage

Kelompok	Anggota Kelompok	
1	Aceh	Nusa Tenggara Timur
	Sumatera Barat	Kalimantan Barat
	Riau	Kalimantan Tengah
	Jambi	Kalimantan Selatan
	Sumatera Selatan	Kalimantan Timur
	Bengkulu	Sulawesi Utara
	Lampung	Sulawesi Tengah
	Bangka Belitung	Sulawesi Selatan
	Kepulauan Riau	Sulawesi Tenggara
	Jawa Barat	Gorontalo
	Jawa Tengah	Maluku
	DI Yogyakarta	Maluku Utara
	Banten	Papua Barat
	Bali	Papua
Nusa Tenggara Barat		
2	Sumatera Utara	
	Jawa Timur	
3	Jakarta	

Sumber : Data Sekunder (2018), diolah

3.5 Pengelompokan Dengan Metode Average Linkage

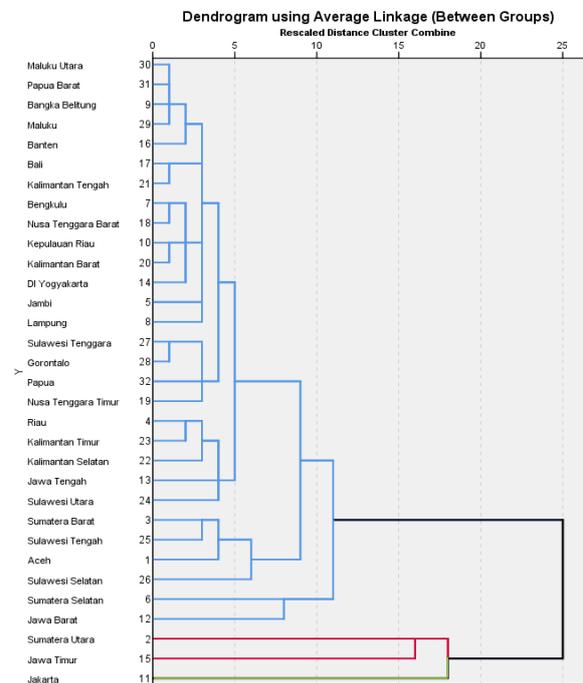
Pada metode Average Linkage menghitung jarak antar objek yang disebut sebagai jarak rata-rata dimana jarak tersebut dihitung pada masing-masing gerombol. Berdasarkan persamaan (2.4) dilakukan perhitungan jarak antara gerombol (30,31) dengan gerombol lain. Diberikan contoh perhitungan antara gerombol (30,31) dengan gerombol 1,2, dan 3, yaitu.

$$d_{(3031),1} = \frac{d_{(301)} + d_{(311)}}{2} = \frac{2.624,32 + 2.477,76}{2} = 2.551,04$$

$$d_{(3031),2} = \frac{d_{(302)} + d_{(312)}}{2} = \frac{7.983,40 + 7.904,59}{2} = 7.944,00$$

$$d_{(3031),3} = \frac{d_{(303)} + d_{(313)}}{2} = \frac{2.971,25 + 2.797,79}{2} = 2.884,52$$

Demikian seterusnya, proses perhitungan dilakukan dengan cara yang sama sehingga semua gerombol tergabung dalam tiga gerombol seperti tampak pada gambar 3 sebagai berikut.



Gambar 3. Dendrogram Average Linkage

Pada gambar 3 dapat dilihat bahwa hasil pengelompokan pada metode Average Linkage disajikan dalam tabel 3.

Tabel 3. Pengelompokan Provinsi dengan Metode *Average Linkage*

Kelompok	Anggota Kelompok	
1	Aceh	NTT
	Sumatera Barat	Kalimantan Barat
	Riau	Kalimantan Tengah
	Jambi	Kalimantan Selatan
	Sumatera Selatan	Kalimantan Timur
	Bengkulu	Sulawesi Utara
	Lampung	Sulawesi Tengah
	Bangka Belitung	Sulawesi Selatan
	Kepulauan Riau	Sulawesi Tenggara
	Jawa Barat	Gorontalo
	Jawa Tengah	Maluku
	DI Yogyakarta	Maluku Utara
	Banten	Papua Barat
	Bali	Papua
NTB		
2	Sumatera Utara	
	Jawa Timur	
3	Jakarta	

Sumber : Data Sekunder (2018), diolah

3.6 Pengelompokan Dengan Metode *K-Means*

Banyaknya gerombol yang ingin dibentuk dengan metode *K-Means* pada penelitian ini adalah tiga. Sehingga terdapat tiga buah *centroid* (pusat gerombol) dimana c_1 (*centroid* pertama), c_2 (*centroid* kedua), c_3 (*centroid* ketiga) dengan bantuan SPSS, nilai *centroid* dapat dilihat pada tampilan *initial cluster center* pada tabel 4.

Tabel 4. *Initial Cluster Center*

Jenis Kejahatan	Initial Cluster Centers		
	Cluster		
	1	2	3
X1	69	49	23
X2	1.601	688	167
X3	404	496	0
X4	1.324	491	52
X5	70	56	18
X6	210	195	64
X7	67	17	2
X8	1.642	854	69
X9	5.569	7.523	27
X10	5.969	1.693	36
X11	2.432	666	16
X12	52	112	35

Sumber : Data Sekunder (2018), diolah

Sehingga diperoleh c_1 merupakan nilai dari tiap variabel untuk Provinsi Jakarta, c_2 merupakan nilai dari tiap variabel untuk Provinsi Jawa Timur, c_3 merupakan nilai dari tiap

variabel untuk Provinsi Maluku Utara dengan nilai masing-masing seperti pada tabel 5.

Perhitungan jarak setiap objek dengan c_1 , c_2 , dan c_3 menggunakan persamaan Jarak *Euclidean*. Demikian seterusnya sehingga proses ini memperoleh nilai *centroid* terbaru yang sama dengan final cluster centers output dari SPSS diuraikan dalam tabel 5 sebagai berikut.

Tabel 5. *Final Cluster Center*

Jenis Kejahatan	Final Cluster Centers		
	Cluster		
	1	2	3
X1	69	105	34
X2	1601	1824	313
X3	404	1620	595
X4	1324	246	316
X5	70	121	44
X6	210	98	111
X7	67	48	7
X8	1642	1500	751
X9	5569	6557	706
X10	5969	1881	604
X11	2432	1838	494
X12	52	84	37

Sumber : Data Sekunder (2018), data diolah

Dapat dilihat nilai ketiga *centroid* baru tersebut sudah tetap atau tidak mengalami perubahan, maka proses pengelompokan berhenti. Sehingga untuk kelompok yang terbentuk pada tabel 6.

Tabel 6. Pengelompokan Provinsi dengan Metode *K-Means*

Kelompok	Anggota Kelompok	
1	Jakarta	
2	Sumatera Utara	Jawa Timur
3	Aceh	NTT
	Sumatera Barat	Kalimantan Barat
	Riau	Kalimantan Tengah
	Jambi	Kalimantan Selatan
	Sumatera Selatan	Kalimantan Timur
	Bengkulu	Sulawesi Utara
	Lampung	Sulawesi Tengah
	Bangka Belitung	Sulawesi Selatan
	Kepulauan Riau	Sulawesi Tenggara
	Jawa Barat	Gorontalo
	Jawa Tengah	Maluku
	DI Yogyakarta	Maluku Utara
	Banten	Papua Barat
	Bali	Papua
NTB		

Sumber : Data Sekunder (2018), diolah

3.7 Pemilihan Metode Terbaik

Pada tahap ini akan dilakukan pemilihan metode terbaik pada metode *Single Linkage*,

Complete Linkage, Average Linkage, dan K-Means yang telah dijelaskan sebelumnya telah diperoleh tiga kelompok atau gerombol. Hasil perhitungan nilai rasio S_w terhadap S_b pada masing-masing metode dapat dilihat pada tabel 7.

Tabel 7. Perbandingan Nilai Rasio Simpangan Baku (s)

No	Metode	Nilai Simpangan Baku (s)
1	Single Linkage	0,18
2	Complete Linkage	0,34
3	Average Linkage	0,34
4	K-Means	0,34

Sumber : Data Sekunder (2018), data diolah

3.8 Analisis Biplot

Setelah terbentuknya kelompok-kelompok provinsi, maka akan dilihat vektor jenis akan dilihat vektor jenis kejahatan mana yang memberikan sumbangan keragaman relative besar terhadap masing-masing kelompok provinsi yang selanjutnya disebut sebagai peubah penciri dari masing-masing kelompok provinsi tersebut. Untuk itu akan dilakukan pembentukan plot dua dimensi, dimana data yang digunakan merupakan data nilai rata-rata dari setiap jenis kejahatan yang dimiliki masing-masing kelompok.

Untuk memperoleh plot dua dimensi dari masing-masing kelompok provinsi dan jenis kejahatan, maka langkah awal yang perlu dilakukan adalah mencari titik koordinat dari masing-masing kelompok provinsi dan jenis kejahatan. Koordinat dari masing-masing kelompok provinsi dapat diperoleh dari matriks G , sedangkan koordinat dari masing-masing jenis kejahatan dapat diperoleh dari matriks H .

Dengan menggunakan bantuan program R , maka diperoleh matriks G yang memuat skor komponen dari dua komponen utama dari data sebagai berikut.

Tabel 9. Skor Komponen 1 dan Komponen 2

Kelompok	Skor Komponen 1	Skor Komponen 2
Kelompok 1	-0.8159858	0.02887711
Kelompok 2	0.3829846	-0.72110296
Kelompok 3	0.4330012	0.69222585

Sumber : Data Sekunder (2018), data diolah

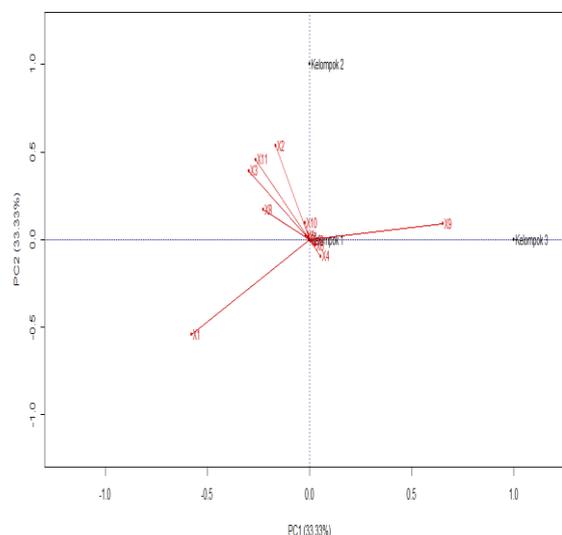
Untuk memperoleh matriks H dilakukan dengan hal yang sama dengan menggunakan bantuan R , maka diperoleh matriks H yang memuat dua vektor eigen pertama yang merupakan titik-titik koordinat dari masing-masing jenis kejahatan disajikan pada tabel 10 sebagai berikut.

Tabel 10. Tabel Vektor Eigen 1 dan Vektor Eigen 2

Jenis Kejahatan	Vektor eigen 1	Vektor eigen 2
X1	53.804002	-81.14962
X2	1140.852913	-1647.21868
X3	785.193724	-1617.65098
X4	-72.909832	349.98699
X5	59.065752	-94.07172
X6	-9.182813	138.29711
X7	30.215811	-44.93738
X8	554.163239	-933.05773
X9	4690.051125	1201.00817
X10	886.435541	-296.70132
X11	984.700445	-1692.63612
X12	39.074529	38.23995

Sumber : Data Sekunder (2018), data diolah

Setelah diperoleh koordinat dari masing-masing kelompok provinsi dan jenis kejahatan, maka selanjutnya dibentuk plot untuk masing-masing kelompok dan jenis kejahatannya. Berikut ini adalah penggambaran biplot yang menampilkan titik untuk setiap kelompok provinsi dan jenis kejahatan.



Gambar 4. Hasil Pemetaan Biplot

Pada gambar 4 merupakan hasil pengelohan dari analisis yang memiliki keragaman data atau ukuran kesesuaian biplot yang diterangkan oleh dua sumbu utama biplot yaitu sebesar 66,66%, dimana sumbu pertama (PC1) mampu menerangkan keragaman data sebesar 33,33% dan sumbu kedua (PC2) mampu menerangkan keragaman 33,33%. Hal ini menunjukkan bahwa analisis biplot hanya mampu menjelaskan informasi sebesar 66,66% dari keseluruhan data dan terdapat kehilangan informasi sebesar 33,34%.

Untuk mengetahui jenis kejahatan yang menjadi penciri dari masing-masing kelompok provinsi maka akan dilakukan proyeksi ortogonal dari masing-masing kelompok provinsi terhadap setiap vektor jenis kejahatan untuk mengetahui panjang proyeksi vektor dimana semakin besar panjang proyeksinya hal ini berarti semakin besar juga ciri yang diwakili kelompok provinsi berdasarkan peubah atau jenis kejahatan yang bersesuaian.

Tabel 11. Panjang Proyeksi Vektor

Jenis Kejahatan	Kelompok 1	Kelompok 2	Kelompok 3
X1	-0.4749777	0.8126385	-0.3376608
X2	-0.4883362	0.8108654	-0.3225292
X3	-0.3822941	0.8159578	-0.4336637
X4	0.1946849	-0.7840544	0.5893695
X5	-0.4583579	0.814355	-0.3559971
X6	0.0828754	-0.7448926	0.6620172
X7	-0.4792744	0.8121073	-0.3328329
X8	-0.4415097	0.8155673	-0.3740577
X9	-0.7833159	0.1921283	0.5911877
X10	-0.782957	0.5920621	0.1908949
X11	-0.4352816	0.8158861	-0.3806045
X12	-0.5629858	-0.2306449	0.7936308

Sumber : Data Sekunder (2018), data diolah

Berdasarkan tabel 11 dapat dilihat bahwa kelompok 1 memiliki pangjan proyeksi vektor pada jenis kejahatan X4 atau jenis kejahatan kekerasan dalam rumah tangga (KDRT), sehingga untuk peubah penciri pada kelompok 1 yaitu jenis kejahatan KDRT. Kemudian untuk peubah penciri setiap kelompok akan dirangkum dalam tabel 12.

Tabel 12. Peubah Penciri Setiap Kelompok

Kelompok	Peubah Penciri
1	Jenis Kejahatan Kekerasa Dalam Rumah Tangga (X4)
2	Jenis Kejahatan Penganiayaan Ringan (X3)
3	Jenis Kejahatan Korupsi (X12)

Sumber : Data Sekunder (2018), data diolah

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Berdasarkan hasil penelitian yang dilakukan pada bab sebelumnya bahwa untuk pengelompokan tingkat kriminalitas dengan metode *Agglomerative* dan *K-Means* pada 32 provinsi dapat disimpulkan untuk metode terbaik dalam pengelompokan tingkat kriminalitas yaitu metode *Single Linkage*.

Terdapat tiga kelompok dalam metode *Single Linkage*. Kelompok pertama terdiri dari Provinsi Aceh, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Riau, Jakarta, Jawa Barat, Jawa, DI Yogyakarta, Banten, Bali, NTB, NTT, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Maluku, Maluku Utara, Papua Barat, dan Papua. Sedangkan untuk kelompok kedua dan ketiga yaitu Provinsi Sumatera Utara dan Provinsi Jawa Timur. Kemudian untuk peubah penciri pada ketiga kelompok tersebut yaitu kelompok pertama perubah pencirinya adalah jenis kejahatan KDRT sedangkan untuk kelompok kedua peubah pencirinya yaitu jenis kejahatan penganiayaan ringan dan untuk kelompok ketiga peubah pencirinya adalah jenis kejahatan korupsi.

4.2 Saran

Pada penelitian ini hanya mengkaji tentang empat metode pengelompokan dalam analisis gerombol yaitu *Single Linkage*, *Complete Linkage*, *Average Linkage*, dan metode *K-Means* dengan penentuan jumlah kelompok adalah tiga kelompok. Bagi peneliti yang ingin melanjutkan tentang analisis gerombol dapat menggunakan

metode-metode analisis gerombol yang lain seperti metode *Ward* dan *Two Step Cluster* sebagai perbandingan.

DAFTAR PUSTAKA

- Barakbah, Ali & Arai, Kohei. 2004. *Determining Constraints Of Moving Variance to Find Global Optimum and Make Automatic Clustering*. Industrial Electronics Seminar, 409-413.
- Bunkers WJ, et al. 1996. *Definition of Climate Regions in the Northern Plains Using an Objective Cluster Modification Technique*. J.Climate 9:130-146.
- Everitt, Brian.S.et.al. 2011. *Cluster Analysis*. 5th Edition. United Kingdom. John Wiley and Son,Ltd.
- Gudono. 2011. *Analisis Data Multivariat*. Edisi Pertama. Yogyakarta: BPFE.
- Jolliffe, I.T 2002. *Principal Component Analysis*. Springer-verlag. New York.
- Johnson, R.A.&Dean.W.Wichern. 2007. *Applied Multivariate Statitical Analysis*. Sixth Edition. United State of America: Pearson Education Inc.
- Rencher, A.C. 2002. *Methods of Multivariate Analysis Second Edition*. New Yowk: John Wiley And Son,Inc.