

# Pengelompokan Optimal Kabupaten dan Kota Rawan Kriminalitas di Jawa Timur dengan Metode Analisis Kluster Terbaik

M. Fariz Fadillah Mardianto, Kuzairi, Toni Yulianto, Rica Amalia, Faisal  
Jurusan Matematika, Fakultas MIPA, Universitas Islam Madura (UIM)  
Kampus Hijau Komplek PP. Miftahul Ulum Bettet, Pamekasan 69351  
E-mail: fm.fariz@yahoo.com

**Abstrak**—Kriminalitas adalah kejadian melanggar hukum yang mengganggu keamanan di suatu wilayah. Tingginya tingkat kriminalitas dapat menghambat mobilitas masyarakat dan investor. Kegiatan perekonomian, perdagangan, dan pendidikan dapat terganggu di suatu daerah yang dikategorikan sebagai daerah rawan jika keamanan tidak dijamin. Akhir-akhir ini isu kriminalitas menjadi topik yang menarik di Jawa Timur. Jawa Timur dengan luas wilayah yang besar, jumlah penduduk yang banyak dan beragam, serta pembangunan yang pesat memiliki probabilitas tinggi dalam hal kriminalitas. Pengelompokan daerah rawan kriminalitas perlu dilakukan.

Dalam penelitian ini pengelompokan daerah rawan kriminalitas dilakukan dengan metode analisis kluster hierarki diantaranya *single linkage*, *complete linkage*, *average linkage*, dan metode ward, serta analisis kluster non hierarki dengan menggunakan metode *k-means*. Jumlah kluster optimal untuk tiap metode ditentukan dengan statistik Pseudo-F terbesar. Setelah mendapatkan jumlah kluster optimal dengan Pseudo-F, langkah selanjutnya adalah menentukan metode analisis kluster yang terbaik dengan melihat nilai *internal cluster dispersion rate (icdrate)*. Semakin kecil nilai *icdrate*, semakin baik metode tersebut dalam melakukan pengelompokan. Setelah dilakukan pengelompokan didapatkan hasil lima kelompok yang optimal berdasarkan metode *ward* dimana terdapat lima kategori daerah dengan tingkat kriminalitas sangat rawan, rawan, sedang, aman, sangat aman. Berdasarkan hasil analisis kluster dua wilayah disarankan untuk mendapatkan keamanan khusus yaitu Kota Surabaya yang merupakan daerah sangat rawan kriminalitas, dan Kota Malang dengan kategori rawan kriminalitas. Kedua daerah ini merupakan dua kota terbesar di Jawa Timur.

**Kata Kunci**— tingkat kriminalitas, Jawa Timur, analisis kluster, *internal cluster dispersion rate*, *Pseudo-F*

## I. PENDAHULUAN

**K**EAMANAN merupakan salah satu harapan masyarakat yang tinggal di suatu wilayah. Salah satu indikator keamanan adalah rendahnya kasus kriminalitas. Kriminalitas merupakan segala macam bentuk tindakan dan perbuatan yang

merugikan secara ekonomis dan psikologis yang melanggar hukum yang berlaku dalam negara Indonesia serta norma-norma sosial dan agama. Dapat diartikan bahwa, tindak kriminalitas adalah segala sesuatu perbuatan yang melanggar hukum dan melanggar norma-norma sosial, sehingga masyarakat menentangnya [1].

Kriminalitas kembali menjadi persoalan yang dibicarakan oleh masyarakat. Tindakan kriminalitas seperti pencurian, penipuan, penculikan, dan lain sebagainya yang sering diberitakan dapat menciptakan keresahan masyarakat. Keresahan masyarakat dapat menghambat mobilitas terhambat. Kegiatan perekonomian, perdagangan, dan pendidikan dapat terganggu di suatu daerah yang dikategorikan sebagai daerah rawan kriminalitas. Keresahan masyarakat dapat bertambah jika keamanan tidak dijamin oleh aparat keamanan. Oleh sebab itu pengelompokan daerah rawan kriminalitas perlu dilakukan dengan tujuan masyarakat beserta pihak aparat keamanan dapat bersinergi untuk menurunkan angka kriminalitas di daerah tersebut [1].

Provinsi Jawa Timur adalah satu provinsi di Indonesia dengan jumlah tindak pidana kriminalitas yang tinggi. Tahun 2001-2013 jumlah tindak pidana kriminalitas di Jawa Timur selalu berada di urutan lima besar tertinggi. Tahun 2013 terdapat 16.913 tindak pidana kriminalitas [2]. Provinsi ini juga memiliki karakteristik masyarakat yang heterogen, wilayah yang luas, dan budaya yang beragam sehingga menarik untuk diteliti.

Penelitian tentang pengelompokan suatu daerah dilakukan dengan menggunakan analisis berdasarkan data yang dibutuhkan. Dalam analisis ini, data yang menggambarkan banyaknya kejadian kriminalitas tertentu dinyatakan sebagai variabel. Untuk meringkas data dengan peubah banyak, digunakan analisis multivariat, yaitu analisis kluster untuk memahami dan mempermudah interpretasi data multivariat. Analisis ini bertujuan untuk mengelompokkan objek-objek berdasarkan karakteristik diantara objek-objek tersebut [3]. Dari penelitian ini nantinya didapatkan hasil pengelompokan suatu kabupaten dan kota di Jawa Timur berdasarkan tingkat kejadian kriminalitas. Pengelompokan ini diharapkan mampu menunjukkan kelompok-kelompok kabupaten dan kota berdasarkan banyaknya kejadian-kejadian kriminalitas khusus. Berdasarkan hasil pengelompokan, diharapkan ada perhatian intensif dari pihak aparat dan masyarakat sebagai upaya pengendalian angka kriminalitas untuk menciptakan keamanan. Selain itu dari hasil pengelompokan ini dapat

menjadi bahan pertimbangan serta masukan untuk lembaga-lembaga terkait.

Pada penelitian ini dilakukan pengelompokan optimal berdasarkan berbagai macam metode analisis kluster terhadap data-data kriminalitas di Jawa Timur. Jumlah kluster optimal ditentukan dengan melihat nilai Pseudo-F terbesar. Metode analisis kluster terbaik diperoleh dengan melihat nilai *internal cluster dispersion rate (icdrate)* yang terkecil.

## II. TINJAUAN PUSTAKA

### a. Analisis Multivariat

Analisis multivariat adalah analisis beberapa variabel dalam hubungan tunggal atau banyak hubungan [4]. Analisis multivariat juga didefinisikan sebagai analisis dimana masalah yang diteliti bersifat multidimensional dan menggunakan tiga atau lebih variabel [5]. Jika terdapat sebanyak  $n$  objek dan  $p$  variabel, maka observasi objek ke- $i$  dan variabel ke- $j$  yang dinotasikan  $x_{ij}$  dengan  $i = 1, 2, \dots, n$  dan  $j = 1, 2, \dots, p$  dapat ditampilkan pada Tabel 1 sebagai berikut :

**Tabel 1.** Hubungan beberapa variabel dan beberapa objek

	Var 1	Var 2	...	Var j	...	Var p
Objek 1	$x_{11}$	$x_{12}$	...	$x_{1j}$	...	$x_{1p}$
Objek 2	$x_{21}$	$x_{22}$	...	$x_{2j}$	...	$x_{2p}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Objek i	$x_{i1}$	$x_{i2}$	⋮	$x_{ij}$	⋮	$x_{ip}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Objek n	$x_{n1}$	$x_{n2}$	⋮	$x_{nj}$	⋮	$x_{np}$

### b. Pengelompokan Hierarki

Metode hierarki merupakan metode kluster yang hasil pengelompokannya disajikan secara bertingkat atau berjenjang dari  $n, (n - 1), \dots, 1$  kelompok. Fungsi jarak yang seringkali digunakan adalah *Euclidean*, dimana didefinisikan sebagai jarak antara observasi ke- $i$  dan ke- $k$  [3]. Rumus jarak euclid dari objek ke- $i$  menuju objek ke- $h$  dirumuskan pada persamaan (2.1).

$$d(i, h) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{hk})^2}; \quad i = 1, 2, \dots, n; \quad k = 1, 2, \dots, p; \quad i \neq h \quad (2.1)$$

#### a. Single Linkage (Pautan Tunggal/Nearest Neighbor)

*Single linkage* membentuk kelompok-kelompok dari individu dengan menggabungkan jarak terdekat atau terkecil. *Single linkage* dirumuskan pada persamaan (2.2).

$$d_{(ih)g} = \min\{d_{ig}, d_{hg}\} \quad (2.2)$$

dengan  $d_{(ih)g}$  jarak objek  $ih$  dengan objek  $g$ ,  $d_{ig}$  jarak terpendek objek  $i$  dengan objek  $g$ ,  $d_{hg}$  adalah jarak terpendek objek  $h$  dengan objek  $g$ .

#### b. Complete Linkage (Pautan Lengkap/Furthest Neighbor)

*Complete linkage* mengelompokkan semua objek dalam cluster yang berada paling jauh satu sama lainnya. *Complete linkage* dirumuskan pada persamaan (2.3)

$$d_{(ih)g} = \max\{d_{ig}, d_{hg}\} \quad (2.3)$$

dengan  $d_{(ih)g}$  jarak objek pada kelompok ( $ih$ ) dengan objek  $g$ ,  $d_{ig}$  jarak terjauh objek  $i$  dengan objek  $g$ ,  $d_{hg}$  adalah jarak terjauh objek  $h$  dengan objek  $g$ .

#### c. Average Linkage (Pautan Rataan/Between Method)

Metode ini memperlakukan jarak antara dua cluster sebagai jarak rata-rata antara semua pasangan individu. Untuk memperoleh penggabungan (aglomerasi) objek satu dengan lainnya dirumuskan sebagai berikut :

$$d_{(ih)g} = \frac{\sum_i \sum_g d_{ig}}{N_{(ih)} N_g} \quad (2.4)$$

dengan  $d_{(ih)g}$  jarak antara objek  $g$  pada kelompok ( $ih$ ) dan objek  $g$ .  $N_{(ih)}$  adalah banyaknya anggota dalam kelompok ( $ih$ ), sedangkan  $N_g$  banyaknya anggota pada kelompok  $g$ .

#### d. Metode Ward

Pengelompokan metode Ward adalah memperkecil total jumlah kuadrat dalam kelompok. Masing-masing kelompok dibentuk sedemikian hingga menghasilkan kelompok dengan jumlah kuadrat yang terkecil dan dikenal dengan *Error Sums of Squares (ESS)*. Dua kluster yang memiliki peningkatan *ESS* paling minimum akan berkelompok. Jika cluster sebanyak  $c$  maka *SSE* merupakan jumlahan dari  $ESS_c$ .

$$ESS = ESS_1 + ESS_2 + \dots + ESS_c$$

Saat semua cluster bergabung menjadi satu kelompok dari  $N$  item, maka nilai *SSE* dirumuskan :

$$ESS = \sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x})$$

dengan  $x_j$  matrik multivariat objek ke- $j$  dan  $\bar{x}$  adalah mean dari semua objek [3].

#### c. Pengelompokan Non Hierarki

Pengelompokan nonhierarki berawal dari menentukan inisial banyak kelompok. Algoritma k-means bertujuan untuk meminimasi fungsi objektif yang merupakan fungsi error kuadrat. Misal ada  $n$  objek dan  $p$  variabel. Jarak antara objek ke- $i$  dan kelompok ke- $l$  dihitung menggunakan jarak euclid kuadrat dirumuskan persamaan (2.5).

$$D[i, l(i)] = \sum_{j=1}^p [X(i, j) - \bar{X}(l, j)]^2; \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, p \quad (2.5)$$

dengan  $X(i, j)$  nilai antara objek ke- $i$  terhadap variabel ke- $j$ ,  $\bar{X}(l, j)$  rata-rata variabel ke- $j$  terhadap kelompok ke- $l$ ,  $D[i, l(i)]$  jarak euclid antara objek ke- $i$  dan rata-rata kluster dari kluster objek (centroid) [3].

#### d. Pseudo F dan Internal Cluster Dispersion Rate (icdrate)

Jumlah kluster atau kelompok optimal ditentukan dengan melihat nilai statistik Pseudo-F. Berdasarkan penelitian Milligan dan Cooper statistik ini secara komputasi memberikan hasil terbaik untuk menentukan berapa kluster yang optimal. Berikut adalah statistik Pseudo-F :

$$Pseudo - F = \frac{(n - c)R^2}{(c - 1)(1 - R^2)} \quad (2.6)$$

dengan

$$R^2 = (SST - SSW)/SST$$

$$SST = \sum_{i=1}^{n_c} \sum_{j=1}^c \sum_{k=1}^p (x_{ij}^k - \bar{x}^k)^2$$

$$SSW = \sum_{i=1}^{n_c} \sum_{j=1}^c \sum_{k=1}^p (x_{ij}^k - \bar{x}_j^k)^2$$

$R^2$  adalah koefisien determinasi yang menunjukkan kontribusi dari variabel-variabel, *SST* total jumlah dari kuadrat jarak

terhadap rata-rata keseluruhan,  $SSW$  total jumlah dari kuadrat jarak sampel terhadap rata-rata kelompoknya,  $n$  banyaknya sampel,  $c$  banyak kelompok,  $n_c$  banyak data pada kelompok ke- $i$ ,  $p$  banyak variabel,  $x_{ij}^k$  sampel ke- $i$  pada kelompok ke- $j$  dan variabel ke- $k$ ,  $\bar{x}^k$  rata-rata seluruh sampel pada variabel- $k$ ,  $\bar{x}_j^k$  rata-rata sampel pada kelompok ke- $j$  dan variabel ke- $k$  [6].

Kebaikan dalam kluster adalah homogenitas yang tinggi antar anggota dalam satu kluster (*within-cluster*) dan heterogenitas yang tinggi antar kluster (*between-cluster*). *Icdrate* adalah untuk menilai homogenitas dalam kelompok dan heterogenitas antar kelompok. Semakin kecil nilai *icdrate*, semakin baik hasil pengelompokannya [6]. Perhitungan *internal cluster dispersion rate (icdrate)* dirumuskan persamaan (2.7).

$$icdrate = 1 - \frac{SSB}{SST} = 1 - \frac{SST - SSW}{SST} = 1 - R^2 \quad (2.7)$$

e. *Multivariate Analysis of Varians (MANOVA)*

*Multivariate Analysis of Varians (MANOVA)* adalah teknik yang digunakan untuk membandingkan rata-rata dua populasi atau lebih.

Uji pengaruh perlakuan (MANOVA) :

Hipotesis :

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_c = 0$$

$$H_1 : \text{minimal 1 pasang } \mu_j \neq 0 \quad (j = 1, 2, \dots, c)$$

Statistik uji :

$$\Lambda^* = \frac{|W|}{|B + W|}$$

Kriteria pengujian :

$$H_0 \text{ ditolak jika } \Lambda^* > F_{n_c-1, n-n_c}(\alpha).$$

**Tabel 2.** MANOVA

Sumber variasi	df	Sum of squares	Wilks Criterion
Perlakuan	$K-1$	<b>B</b>	$\frac{ W }{ W+B } \sim \Lambda_{\alpha; (p, n-K, K-1)}$
Residual	$\sum_{j=1}^K n_j - K$	<b>W</b>	
Total	$\sum_{j=1}^K n_j - 1$	<b>B + W</b>	

Keterangan :

$$B = \sum_{j=1}^K n_j (\bar{x}_j - \bar{x})(\bar{x}_j - \bar{x})'$$

$$W = \sum_{j=1}^K \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)(x_{ij} - \bar{x}_j)'$$

$$T = \sum_{j=1}^K \sum_{i=1}^{n_j} (x_{ij} - \bar{x})(x_{ij} - \bar{x})'$$

dengan :

**W** : Matriks *Sum of Square Residuals (SSW)*

**B** : Matriks *Sum of Square Treatment (SST)*

$n$  : jumlah sampel

$c$  : banyaknya kelompok

$n_c$  : banyak anggota pada kelompok  $c$

Uji MANOVA dilakukan setelah data memenuhi asumsi-asumsi berikut :

a. Matriks varians kovarians antar perlakuan homogen.

Variabel  $X$  dengan  $p$  variabel dikatakan berdistribusi multivariat normal dengan parameter  $\mu$  dan  $\Sigma$  bila memiliki fungsi densitas pada persamaan

$$f(X_1, X_2, \dots, X_p) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{p/2}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)$$

Pemeriksaan multivariat normal dilakukan dengan cara membuat q-q plot dari  $d_i^2$  dan  $q_i$ .

b. Setiap populasi memiliki distribusi multivariat normal.

Matriks varians kovarians dapat diuji dengan menggunakan uji Box's M, hipotesis dan statistik uji Box's M dirumuskan :

Hipotesis :

$$H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_p$$

$$H_1 : \text{minimal ada satu } \Sigma_i \text{ dan } \Sigma_j \text{ yang berbeda, dengan } i \neq j$$

Statistik Uji :

$$X_{hitung}^2 = -2(1 - c_1) \left( \frac{1}{2} \sum_{j=1}^c v_j \ln|S_i| - \frac{1}{2} \ln|S_{pool}| \left\| \sum_{j=1}^c v_j \right\| \right)$$

dengan

$$S_{pool} = \frac{\sum_{j=1}^c v_j S_i}{\sum_{j=1}^c v_j}$$

$$c_1 = \left( \sum_{j=1}^c \frac{1}{v_j} - \frac{1}{\sum_{j=1}^c v_j} \right) \left( \frac{2p^2 + 3p - 1}{6(p+1)(k-1)} \right) v_i = n_i - 1$$

Kriteria Pengujian :

Keputusan gagal tolak  $H_0$  jika  $X_{hitung}^2 \leq X_{1/2(k-1)p(p+1)}^2$  [6].

f. *Tindak Pidana Kriminalitas*

Tindak pidana adalah suatu perbuatan yang dilakukan oleh seseorang dimana perbuatan tersebut melanggar ketentuan perundang-undangan yang diancam dengan sanksi terhadap pelanggaran tersebut, dimana perbuatan yang melanggar ketentuan perundang-undangan tersebut melahirkan sanksi yang bersifat pidana, sanksi bersifat perdata, ataupun sanksi yang bersifat administrasi [7]. Jenis-jenis kejahatan yang datanya diterbitkan oleh Badan Pusat Statistik (BPS) Jawa Timur dan digunakan sebagai variabel penelitian ini adalah sebagai berikut :

- Pembunuhan adalah perbuatan dengan sengaja menghilangkan jiwa orang lain (X1).
- Perkosaan adalah pemaksaan perempuan yang bukan istrinya untuk bersetubuh dengan kekerasan atau ancaman kekerasan (X2).
- Penganiayaan ringan adalah perbuatan dengan sengaja mengakibatkan rusaknya kesehatan orang lain yang tidak menimbulkan halangan bagi korban untuk melakukan pekerjaan sehari-hari (X3).
- Penganiayaan berat adalah perbuatan dengan sengaja mengakibatkan rusaknya kesehatan orang lain sampai dengan mengakibatkan korban luka/cacat atau menjadi sakit sehingga tidak mampu melakukan pekerjaan sehari-hari dengan sempurna (X4).
- Penculikan atau perampasan kemerdekaan adalah perbuatan melarikan seseorang dengan melawan hak, dengan maksud menjadikan orang tersebut di bawah kekuasaan diri pelaku atau orang lain (X5).
- Pembakaran dengan sengaja adalah perbuatan yang dengan sengaja membakar (misalnya rumah, hutan) yang

dapat mendatangkan bahaya bagi barang, jiwa atau badan (X6).

- g. Pengrusakan atau penghancuran adalah perbuatan yang dengan sengaja dilakukan dan berakibat kerusakan pada barang milik pribadi atau fasilitas umum (X7).
- h. Pencurian dengan pemberatan adalah pencurian barang atau ternak yang didahului, disertai dan diikuti dengan pengrusakan (X8).
- i. Pencurian biasa (ringan) adalah pencurian barang atau ternak bukan miliknya dengan maksud untuk memilikinya yang didahului, tanpa disertai dengan kekerasan terhadap orang/ korban (X9).
- j. Pencurian dengan kekerasan adalah pencurian barang atau ternak yang didahului, disertai atau diikuti dengan kekerasan terhadap orang dengan maksud akan menyiapkan atau memudahkan pencurian itu serta memudahkan atau memberi kesempatan pelaku melarikan diri, atau jika tertangkap basah barang yang dicuri tetap ada ditangan pelaku (X10).
- k. Pencurian dalam keluarga adalah pencurian yang dilakukan oleh anggota keluarga terhadap anggota keluarga yang lain (X11).
- l. Penipuan adalah perbuatan dengan maksud hendak menguntungkan diri sendiri atau orang lain dengan melawan hak, baik dengan memakai nama palsu atau keadaan palsu, baik dengan akal dan tipu muslihat atau dengan perkataan-perkataan bohong, membujuk orang supaya memberikan sesuatu barang, membuat hutang atau menghapuskan piutang (X12).
- m. Penadahan adalah membeli atau mengumpulkan barang hasil curian (X13).
- n. Pencurian kendaraan bermotor adalah pencurian jenis kendaraan bermotor baik kendaraan beroda dua maupun kendaraan beroda empat (X14).
- o. Tindak pidana kriminalitas lainnya (X15).

### III. PEMBAHASAN

#### g. Deskripsi Statistik

Secara statistik pada tahun 2013, deskripsi mengenai banyaknya tindakan kriminalitas di Provinsi Jawa Timur dapat diketahui berdasarkan ukuran pemusatan dan ukuran seperti ditampikan Tabel 3.

**Tabel 3.** Deskripsi Statistik Banyaknya Kejadian Kriminalitas Pada Tahun 2013 di Jawa Timur (dalam kasus)

Variabel	Rata-Rata	Min	Maks	Daerah Maks
X1	3	0	11	Kab Bangkalan
X2	3	0	16	Kab Jember
X3	26	0	201	Kab Jember
X4	43	0	267	Kota Surabaya
X5	0	0	5	Kab Jombang
X6	3	0	89	Kota Surabaya
X7	10	0	57	Kab Jember
X8	129	15	678	Kota Malang
X9	39	0	376	Kota Malang
X10	25	1	184	Kota Surabaya
X11	1	0	23	Kab Jember
X12	93	4	577	Kota Surabaya
X13	6	0	89	Kab Bojonegoro

X14	112	0	1399	Kota Surabaya
X15	6	0	38	Kota Kediri

Selanjutnya sebelum dilakukan analisis kluster ada asumsi yang dipenuhi yaitu pengujian multikolinieritas. Berdasarkan nilai VIF (*variance index factor*) untuk semua variabel kurang dari 10. Ini berarti tidak ada gejala multikolinieritas atau hubungan linieritas yang erat terhadap data.

#### h. Hierarchical Clustering

Metode pengelompokan hierarki dalam penelitian ini adalah penggabungan (*agglomerative*). Jarak tiap objek (kabupaten/kota) yang dihitung dengan jarak euclid dengan rumusan pada persamaan (2.1), dimana  $i = 1, 2, \dots, 38$ ,  $h = 1, 2, \dots, 38$ , dengan kondisi  $i \neq h$  dan  $k = 1, 2, \dots, 15$ . Setelah mendapatkan jarak Euclid dari satu objek ke objek yang lain, maka dilanjutkan dengan membuat kluster.

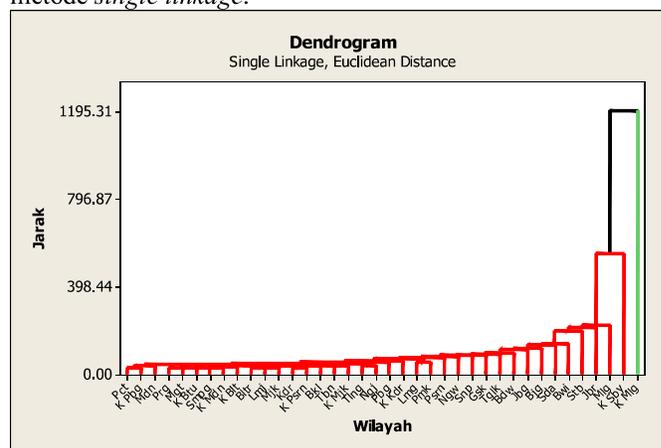
##### 1. Single Linkage

Metode ini membentuk kelompok dari objek-objek dengan menggabungkan jarak paling pendek atau kemiripan yang paling besar. Pada tahap ini untuk menentukan *agglomeration schedule* dari tiap kluster dan perhitungan nilai Pseudo-F dilakukan secara komputasi. Nilai Pseudo-F untuk banyak kluster 2-7 disajikan pada Tabel 3 sebagai berikut :

**Tabel 4.** Nilai Pseudo-F untuk Metode *Single Linkage*

Banyak Kluster	Nilai Pseudo-F
2	2,3542
3	1,4875
4	0,8361
5	0,6225
6	0,4447
7	0,3116

Banyak kluster optimal berdasarkan Tabel 4 ada dua kluster. Adapun anggota tiap kluster adalah Kota Malang ada di kluster 2, sedangkan 37 kabupaten dan kota lain berada di kluster 1. Gambar 1 menyajikan hasil dendrogram untuk metode *single linkage*.



**Gambar 1.** Dendrogram pada metode *single linkage*

##### 2. Complete Linkage

Metode ini membentuk kelompok dari objek-objek yang berada paling jauh satu sama lainnya. Pada tahap ini untuk menentukan *agglomeration schedule* dari tiap kluster dan perhitungan nilai Pseudo-F dilakukan secara komputasi. Nilai Pseudo-F untuk banyak kluster 2-7 disajikan pada Tabel 5 sebagai berikut :

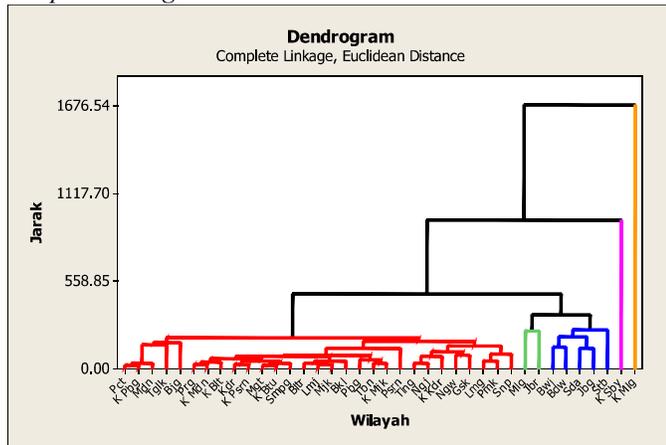
**Tabel 5.** Nilai Pseudo-F untuk Metode *Complete Linkage*

Banyak Kluster	Nilai Pseudo-F
2	5,7866
3	6,9677
4	6,1907
<b>5</b>	<b>6,9760</b>
6	3,7498
7	5,0172

Banyak kluster optimal berdasarkan Tabel 5 ada lima kluster dengan anggota sebagai berikut :

- Anggota Kluster 1 adalah semua kabupaten dan kota di Jawa Timur selain daerah yang masuk anggota kluster 2,3,4, dan 5.
- Anggota Kluster 2 adalah Kab Jember, dan Kab Malang.
- Anggota Cluster 3 adalah Kab Sidoarjo, Kab. Banyuwangi, Kab. Situbondo, Kab. Bondowoso dan Kab Jombang.
- Anggota Kluster 4 adalah Kota Malang.
- Anggota Kluster 5 adalah Kota Surabaya.

Gambar 2 menyajikan hasil dendrogram untuk metode *complete linkage*.



**Gambar 2.** Dendrogram pada metode *complete linkage*

### 3. Average Linkage

Metode *average linkage* juga disebut metode *between group*. Metode ini akan mengelompokkan objek berdasarkan jarak rata-rata yang didapat dengan melakukan rata-rata semua objek terlebih dahulu. Pada tahap ini untuk menentukan *agglomeration schedule* dari tiap kluster dan perhitungan nilai Pseudo-F dilakukan secara komputasi. Nilai Pseudo-F untuk banyak kluster 2-7 disajikan pada Tabel 6 sebagai berikut :

**Tabel 6.** Nilai Pseudo-F untuk Metode *Average Linkage*

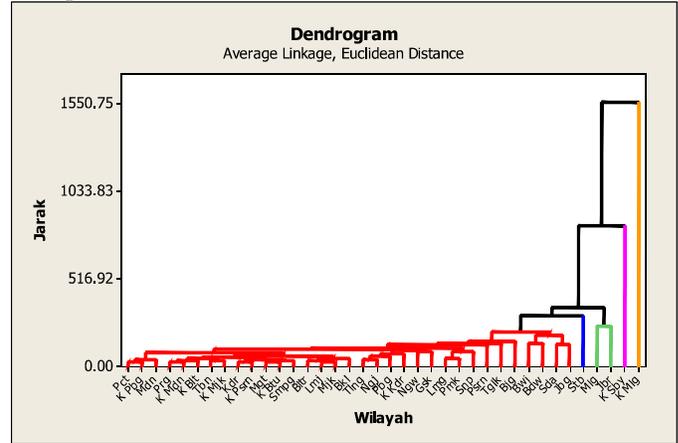
Banyak Kluster	Nilai Pseudo-F
2	2,9263
3	3,0404
4	2,9509
<b>5</b>	<b>4,2610</b>
6	2,3061
7	4,1169

Banyak kluster optimal berdasarkan Tabel 6 ada lima kluster dengan anggota sebagai berikut :

- Anggota Kluster 1 adalah semua kabupaten dan kota di Jawa Timur selain daerah yang masuk anggota kluster 2,3,4, dan 5.
- Anggota Kluster 2 adalah Kab. Jember dan Kab. Malang.
- Anggota Kluster 3 adalah Kab. Situbondo.

- Anggota Kluster 4 adalah Kota Malang.
- Anggota Kluster 5 adalah Kota Surabaya.

Gambar 3 menyajikan hasil dendrogram untuk metode *average linkage*.



**Gambar 3.** Dendrogram pada metode *average linkage*

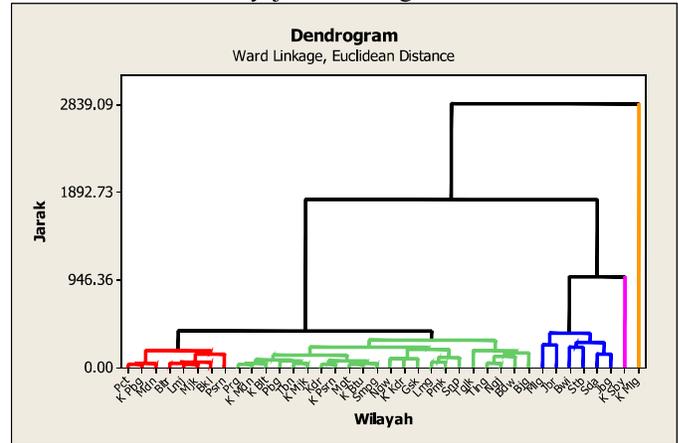
### 4. Metode Ward

Metode Ward memperkecil total jumlah kuadrat dalam kelompok. Masing-masing kelompok dibentuk sedemikian hingga menghasilkan kelompok dengan jumlah kuadrat yang terkecil dan dikenal dengan *Error Sums of Squares (ESS)*. Pada tahap ini untuk menentukan nilai Pseudo-F dilakukan secara komputasi. Nilai Pseudo-F untuk banyak kluster 2-7 disajikan pada Tabel 7 sebagai berikut :

**Tabel 7.** Nilai Pseudo-F untuk Metode *Complete Linkage*

Banyak Kluster	Nilai Pseudo-F
2	46.054
3	47.044
4	72.641
<b>5</b>	<b>149.430</b>
6	26.566
7	21.446

Banyak kluster optimal berdasarkan Tabel 7 ada lima kluster . Gambar 4 berikut menyajikan dendrogram untuk metode ward.



**Gambar 4.** Dendrogram pada metode ward

Hasil dari pengklusteran sebagai berikut :

- Anggota Kluster 1 adalah Kota Probolinggo, Kab. Bangkalan, Kab. Madiun, Kab. Pacitan, Kab. Blitar, Kab. Lumajang, Kab. Mojokerto, Kab. Pasuruan.
- Anggota Kluster 2 adalah semua kabupaten kota yang tidak masuk anggota kluster 1,3,4,dan 5.

3. Anggota Kluster 3 adalah Kab Malang, Kab. Jember, Kab. Situbondo, Kab. Banyuwangi, Kab. Sidoarjo, dan Kab. Jombang.
4. Anggota Kluster 4 adalah Kota Malang.
5. Anggota Kluster 5 adalah Kota Surabaya.

*i. Nonhierarchical Clustering*

Prosedur pengelompokan nonhierarki tidak dilakukan secara bertahap dan jumlah kelompoknya juga ditentukan terlebih dahulu. Metode yang digunakan dalam pengelompokan nonhierarki adalah *k-means*. Tahapan analisis *k-means* adalah sebagai berikut :

1. Membagi item-item (objek) ke dalam *k* kluster. *Centroid* ditentukan sebarang objek secara random.
2. Menghitung jarak setiap objek dengan setiap centroid.
3. Selanjutnya menghitung kembali *centroid* baru yang merupakan rata-rata seluruh variabel pada tiap kluster.
4. Menghitung jarak setiap objek dengan tiap centroid yang baru, sama seperti langkah 2.
5. Mengulangi kembali langkah 2 dan 3 sampai tidak ada perubahan anggota kluster.

Proses tersebut dapat digambarkan sebagai algoritma yang dapat memudahkan proses komputasi pengklusteran dengan metode *k-means*. Sama seperti metode sebelumnya, dalam metode *k-means* dapat juga ditentukan nilai Pseudo-F untuk tiap kluster. Nilai Pseudo-F dihitung secara komputasi. Nilai Pseudo-F untuk banyak kluster 2-7 disajikan pada Tabel 8 sebagai berikut :

**Tabel 8.** Nilai Pseudo-F untuk Metode *K-Means*

Banyak Kluster	Nilai Pseudo-F
2	66,8117
3	68,3958
<b>4</b>	<b>68,4800</b>
5	67,4859
6	67,4079
7	67,3708

Banyak kluster optimal berdasarkan Tabel 8 ada empat kluster. Dari proses ini diperoleh anggota tiap kluster sebagai berikut :

- a. Kluster 1 dengan anggota semua kabupaten dan kota di Jawa Timur selain daerah yang masuk anggota kluster 2,3, dan 4.
- b. Kluster 2 dengan anggota Kab. Malang, Kab. Jember, Kab. Situbondo, Kab. Bondowoso, Kab. Banyuwangi, Kab. Sidoarjo, Kab. Jombang.
- c. Kluster 3 dengan anggota Kab. Trenggalek.
- d. Kluster 4 dengan anggota Kota Surabaya, dan Kota Malang.

*j. Pemilihan Metode Terbaik*

Untuk mendapatkan hasil pengelompokan optimal untuk kabupaten dan kota terbaik apakah menggunakan metode *single linkage*, *complete linkage*, *average linkage*, metode *ward*, atau metode *k-means*, maka keenam metode akan dibandingkan berdasarkan tingkat kebaikan hasil pengelompokannya. Kebaikan hasil pengelompokan dapat diketahui berdasarkan *internal cluster dispersion rate* (*icdrate*). Semakin kecil nilai *icdrate*, semakin baik metode tersebut dalam melakukan pengelompokan. Tabel 9

menunjukkan nilai  $R^2$  dan *icdrate* yang digunakan untuk kebaikan metode kluster.

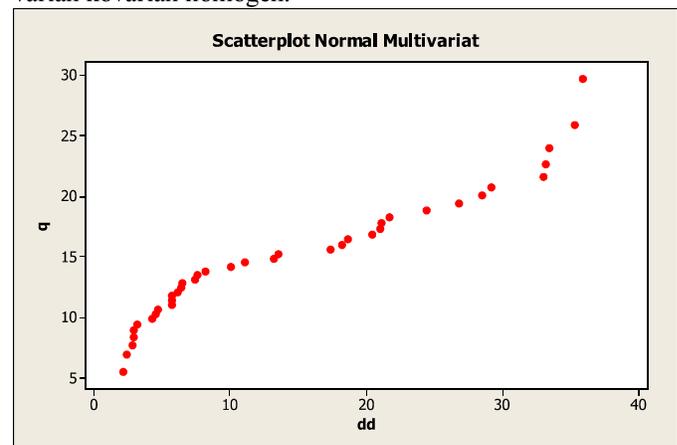
**Tabel 9.** Nilai *icdrate* pada setiap metode pengelompokan

Metode	Kluster	$R^2$	<i>icdrate</i>
<i>Single Linkage</i>	2	0,9386	0,0614
<i>Complete Linkage</i>	5	0,9299	0,0701
<i>Average Linkage</i>	5	0,9744	0,0256
<b><i>Ward</i></b>	<b>5</b>	<b>0,9998</b>	<b>0,0002</b>
<i>K-Means</i>	4	0,9451	0,0549

Berdasarkan Tabel 9 nilai *icdrate* terkecil dimiliki oleh metode *ward* yaitu sebesar 0.0002 dengan nilai  $R^2$  terbesar yaitu 0,9998. Dalam penelitian ini metode *ward* dengan lima kluster optimal dipilih untuk mengelompokan kabupaten dan kota di Jawa Timur berdasarkan tingkat kerawanan kriminalitas.

*k. Evaluasi Hasil Pengelompokan*

Hasil pengelompokan dengan metode *ward* untuk 5 kluster dievaluasi dengan pengujian atau evaluasi apakah antar kelompok terdapat perbedaan yang signifikan. Evaluasi dengan menggunakan MANOVA. Asumsi yang harus dipenuhi adalah berdistribusi multivariat normal dan matriks varian kovarian homogen.



**Gambar 5.** Plot distribusi dari uji normal multivariat

Gambar 5 adalah plot dari uji normal multivariat. Berdasarkan hasil pengujian maka data kluster berdistribusi normal dengan indikasi plot pada Gambar 5 dan banyaknya  $d_i^2 \leq \chi_{tabel}^2$  adalah 50,9% yang masih di sekitar 50% sehingga hipotesis nol ditolak. Uji homogenitas matriks varian-kovarian dengan uji box's M pada Tabel 10 menunjukkan bahwa terdapat perbedaan rata-rata antar kluster karena p-value >  $\alpha = 0,05$ . Dengan ini hasil kluster layak untuk digunakan.

**Tabel 10.** Uji Box's M

Box's M	158,451
F	1,953
p-value	0,433

*l. Karakteristik Tiap Kelompok*

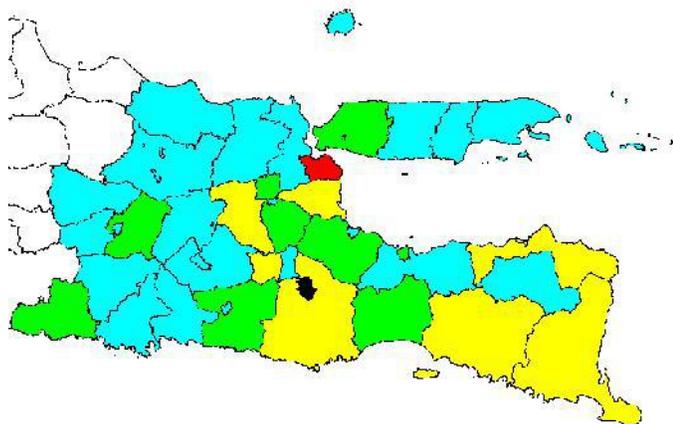
Deskripsi statistik untuk setiap kluster atau setiap kelompok dengan variabel-variabel yaitu jumlah kasus kriminalitas tertentu ditampilkan dalam Tabel 11. Secara umum diketahui bahwa rata-rata jumlah kasus kriminalitas untuk tiap kelompok tertentu memiliki karakteristik tertentu

yang dapat diteliti lebih lanjut berdasarkan literatur yang sesuai dengan bidang permasalahan.

**Tabel 11.** Deskripsi statistik untuk setiap kelompok

Variabel	Kelompok (Ukuran Rata-Rata Kasus)				
	1	2	3	4	5
X1	3	2	3	1	5
X2	2	2	7	0	14
X3	9	14	89	50	39
X4	15	21	107	131	267
X5	0	0	1	0	0
X6	0	0	3	0	89
X7	3	7	26	26	20
X8	60	81	214	678	630
X9	9	31	50	376	0
X10	28	12	40	57	184
X11	0	1	4	0	4
X12	15	54	174	512	577
X13	2	7	6	0	30
X14	26	71	122	1399	287
X15	4	5	10	0	22

Gambar 6 adalah pemetaan daerah rawan kriminalitas berdasarkan hasil analisis kluster pada penelitian ini.



**Gambar 6.** Pemetaan Wilayah Rawan Kriminalitas di Jawa Timur

Keterangan

Merah : Sangat Rawan

Hitam : Rawan

Kuning : Sedang

Biru : Aman

Hijau : Sangat Aman

Deskripsi singkat tentang karakteristik kelompok dibahas singkat dalam Tabel 12-Tabel 14.

**Tabel 12.** Karakteristik Kelompok 1.

Kelompok	Karakteristik
1: Daerah dengan tingkat kriminalitas sangat rendah (daerah sangat aman). Anggotanya Kab Pacitan, Kab Blitar, Kab Lumajang, Kab Pasuruan, Kab Mojokerto, Kab Madiun, Kab Bangkalan, dan Kota Probolinggo.	<ul style="list-style-type: none"> <li>- Sebagian besar jumlah kasus kriminalitas paling rendah dibandingkan kelompok 2, 3, 4, dan 5. Jumlah kejadian antara 54 (Kota Probolinggo) sampai 341 (Kab Pasuruan) kasus.</li> <li>- Rata-rata kasus penganiayaan, curat, penipuan, pencurian ringan dan curanmor terkecil daripada kelompok lainnya.</li> <li>- Jumlah kasus penculikan, pembakaran, dan pencurian dalam keluarga nol (<i>zero accident</i>).</li> </ul>

**Tabel 13.** Karakteristik Kelompok 2, 3, dan 4.

Kelompok	Karakteristik
2: Daerah dengan tingkat kriminalitas rendah (daerah aman). Anggotanya Kab Ponorogo, Kab Trenggalek, Kab Tulungagung, Kab Blitar, Kab Kediri, Kab Bondowoso, Kab Probolinggo, Kab Mojokerto, Kab Nganjuk, Kab Magetan, Kab Ngawi, Kab Bojonegoro, Kab Tuban, Kab Lamongan, Kab Gresik, Kab Sampang, Kab Pemekasan, Kab Sumenep, Kota Kediri, Kota Blitar, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Batu.	<ul style="list-style-type: none"> <li>- Sebagian besar jumlah kasus kriminalitas lebih rendah dibandingkan kelompok 3,4, dan 5, 179 (Kab Magetan) sampai 550 (Kab Bondowoso) kasus.</li> <li>- Rata-rata kasus curas terkecil daripada kelompok lainnya.</li> <li>- Jumlah kasus penculikan, dan pembakaran nol (<i>zero accident</i>).</li> <li>- Tersebar di 64% wilayah di Jawa Timur. Sebagian besar wilayah Jawa Timur termasuk kategori aman.</li> <li>- Semua wilayah Pantura Barat Jawa Timur dan Madura (kecuali Bangkalan) masuk dalam kategori aman.</li> </ul>
3: Daerah dengan tingkat kriminalitas sedang (daerah transisi). Anggotanya Kab Malang, Kab Jember, Kab Banyuwangi, Kab Situbondo, Kab Sidoarjo, dan Kab Jombang.	<ul style="list-style-type: none"> <li>- Sebagian besar jumlah kasus kriminalitas lebih rendah dibandingkan kelompok 4, dan 5, 596 (Kab Jombang) sampai 1.143 (Kab Jember) kasus.</li> <li>- Rata-rata kasus penculikan terbesar, namun hanya 1 kasus saja.</li> <li>- Tidak ada kejadian <i>zero accident</i>.</li> <li>- Wilayah yang masuk kategori ini sebagian besar adalah wilayah dengan luas terbesar dan kecamatan terbanyak.</li> <li>- Wilayah yang masuk kategori ini sebagian besar wilayah yang perkembangan daerah perkotaannya tumbuh, dan banyak dihuni pendatang.</li> </ul>
4: Daerah dengan tingkat kriminalitas tinggi (daerah rawan). Anggotanya Kota Malang.	<ul style="list-style-type: none"> <li>- Jumlah kasus kriminalitas tertinggi di Jawa Timur, yaitu 3.230 kasus, dan paling tinggi dibandingkan kelompok lainnya.</li> <li>- Jumlah kasus penganiayaan ringan, curat, pencurian ringan, dan curanmor tertinggi dibandingkan kelompok lain.</li> <li>- Jumlah kasus curanmor 1.399. Jumlah ini sangat jauh selisihnya dibandingkan dengan daerah terbesar kedua yaitu Surabaya dimana terdapat 287 kasus.</li> <li>- Untuk kasus perkosaan, penculikan, pembakaran dengan sengaja, pencurian dalam keluarga, dan pencurian lain 0 (<i>zero accident</i>).</li> <li>- Wilayah ini kota terbesar dengan fasilitas terlengkap setelah Surabaya. Banyak pendatang dan pusat pendidikan.</li> </ul>

**Tabel 14.** Karakteristik Kelompok 5.

Kelompok	Karakteristik
5: Daerah dengan tingkat kriminalitas sangat tinggi (daerah sangat rawan). Anggotanya Kota Surabaya.	<ul style="list-style-type: none"> <li>- Jumlah kasus kriminalitas sebesar 2.168 kasus.</li> <li>- Jumlah kasus pembunuhan, pemerkosaan, penganiayaan berat, pembakaran, curas, penipuan, dan penadahan barang curian tertinggi dibandingkan kelompok lain.</li> <li>- Banyak kasus kriminalitas yang memiliki predikat tertinggi terbanyak dibandingkan kelompok lain.</li> <li>- Untuk kasus penculikan dan pencurian ringan 0 (<i>zero accident</i>).</li> <li>- Wilayah terpadat di Jawa Timur, dengan penduduk terbanyak, dan jumlah pendatang terbanyak.</li> <li>- Ibu kota Provinsi Jawa Timur dengan fasilitas yang paling lengkap, biaya hidup tinggi.</li> <li>- Jumlah kasus kriminalitas diprediksi terus meningkat.</li> </ul>

Pengelompokan tersebut berdasarkan data tahun 2013. Penelitian daerah rawan kriminalitas perlu dilakukan berkala tiap tahun. Untuk penelitian selanjutnya jumlah kluster optimal bisa juga berubah sesuai nilai Pseudo-F data kluster. Metode terbaik yang digunakan juga dapat berubah sesuai dengan nilai *icdrate* terkecil. Selain itu keanggotaan kluster dapat berubah, daerah yang masuk kategori aman bisa jadi masuk kateogri rawan atau sebaliknya.

Penelitian ini hanya menerapkan pengelompokan optimal dari kumpulan data multivariat dengan menggunakan metode kluster yang terbaik. Hasil penelitian ini sesuai jika dibandingkan dengan kondisi nyata kriminalitas kabupaten kota di Jawa Timur.

#### IV. KESIMPULAN

Berdasarkan pembahasan maka dapat dibuat kesimpulan sebagai berikut:

1. Hasil analisis kluster dari 38 kabupaten dan kota berdasarkan banyaknya kasus-kasus kriminalitaas tertentu dengan menggunakan data tahun 2013, ada 5 kelompok optimal sesuai metode terbaik yaitu metode *ward*. Pembagian kelompoknya adalah sebagai berikut:
  - a. Kelompok 1 daerah sangat aman: Pacitan, Blitar, Lumajang, Pasuruan, Mojokerto, Madiun, Bangkalan, dan Kota Probolinggo.
  - b. Kelompok 2 daerah aman: Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Bondowoso, Probolinggo, Mojokerto, Nganjuk, Ngawi, Bojonegoro, Tuban, Lamongan, Gresik, Sampang, Pamekasan, Sumenep, Kota Kediri, Kota Blitar, Kota Pasuruan, Kota Mojokerto, Kota Madiun, dan Kota Batu.
  - c. Kelompok 3 daerah transisi : Malang, Jember, Banyuwangi, Sidoarjo, Jombang, dan Situbondo.

- d. Kelompok 4 daerah rawan yaitu Kota Malang.
  - e. Kelompok 5 daerah sangat rawan yaitu Kota Surabaya. Secara umum sebagian besar kabupaten kota di Jawa Timur termasuk daerah aman.
2. Ciri atau karakteristik dari tiap kluster adalah :
- a. Untuk daerah di kelompok 1 rata-rata kasus penganiayaan, curat, penipuan, pencurian ringan dan curanmor terkecil daripada kelompok lainnya.
  - b. Untuk daerah di kelompok 2 rata-rata kasus curas terkecil daripada kelompok lainnya..
  - c. Untuk daerah di kelompok 3 rata-rata kasus penculikan terbesar, namun hanya 1 kasus saja.
  - d. Untuk daerah di kelompok 4 jumlah kasus penganiayaan ringan, curat, pencurian ringan, dan curanmor tertinggi dibandingkan kelompok lain.
  - e. Untuk daerah di kelompok 5 jumlah kasus pembunuhan, pemerkosaan, penganiayaan berat, pembakaran, curas, penipuan, dan penadahan barang curian tertinggi dibandingkan kelompok lain.

#### DAFTAR PUSTAKA

- [1] Susanto, I.S. (2010). "*Statistik Kriminal, sebagai Konstruksi Sosial*", Yogyakarta: Genta Publishing.
- [2] Badan Pusat Statistik. (2014). "*Jumlah Tindak Pidana menurut Polda 2000-2013*", Jakarta: BPS.
- [3] Johnson, R.A. & Wichern, D.W. (2002). "*Applied Multivariate Statistical Analysis*", 5th ed. New Jersey: Prentice Hall International Inc.
- [4] Hair, Joseph F., dkk. (1998). "*Multivariate Data Analysis Fifth Edition*". New Jersey : Prentice-Hall.
- [5] Kuncoro, M . (2003). *Metode Riset Untuk Bisnis dan Ekonomi*. Jakarta: Erlangga.
- [6] Lzulf, I . (2013). *Analisis Cluster Kabupaten/Kota berdasarkan Tingkat Pencemaran Udara*. Surabaya: Tugas Akhir ITS.
- [7] Badan Pusat Statistik. (2014). "*Laporan Eksekutif Statistik Politik dan Keamanan Provinsi Jawa Timur 2013*", Surabaya: BPS Jatim.