

**ANALISIS KLASTER METODE WARD DAN AVERAGE LINKAGE  
DENGAN VALIDASI DUNN INDEX DAN KOEFISIEN KORELASI COPHENETIC  
(Studi Kasus: Kecelakaan Lalu Lintas Berdasarkan Jenis Kendaraan  
Tiap Kabupaten/Kota di Jawa Tengah Tahun 2018)**

**Sisca Indah Pratiwi<sup>1</sup>, Tatik Widiharih<sup>2</sup>, Arief Rachman Hakim<sup>3</sup>**

<sup>1,2,3</sup> Departemen Statistika FSM Universitas Diponegoro

[widiharih@gmail.com](mailto:widiharih@gmail.com)

**ABSTRACT**

Based on Central Java Regional Police data, traffic accidents from 2017 to 2018 increased from 17.522 to 19.016 or 8,54 percent. To reduce the number of traffic accidents in Central Java, the initial step was carried out by grouping districts/cities that had the same accident level characteristics based on vehicle type with cluster analysis. The *ward* and *average linkage* method is a hierarchical cluster analysis method. *ward* method can maximize cluster homogeneity. While the *average linkage* method can generate clusters with small cluster variants. In this study using a measure of squared euclidean distance to measure the similarity between pairs of objects. To determine the quality of clustering results, the validation dunn index and cophenetic coefficients correlation are used. Based on the results of the clustering, the optimal number of clusters is obtained at  $q = 5$  for the *average linkage* method with the results of validation dunn index = 0,08571196 and the  $rcoph = 0,687458$ .

**Keywords:** Accidents, Cluster Analysis, Ward Method, Average linkage, Squared Euclidean Distance, Dunn Index, Cophenetic Correlation Coefficient

**1. PENDAHULUAN**

Analisis kluster merupakan salah satu metode statistika yang dapat digunakan untuk melakukan proses pengelompokan objek, sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam kluster yang sama. Metode dalam analisis kluster yaitu hierarki dan non hierarki. Analisis kluster hierarki terdiri dari metode aglomeratif dan devisif. Penggunaan metode aglomeratif dimulai dengan menganggap semua objek sebagai kluster, kemudian menghitung jarak antar kluster dan menggabungkan objek dengan jarak antar kluster yang paling dekat, hingga semua objek bergabung menjadi  $q$  kluster (Johnson dan Winchern, 2007).

Metode pada analisis kluster hierarki aglomeratif diantaranya yaitu *single linkage*, *complete linkage*, *average linkage*, *ward* dan *centroid*. Pada penelitian ini menggunakan metode *ward* dan *average linkage*. Menurut Hair, *et al* (2010), metode *ward* dapat meminimumkan jumlah kuadrat sesatan dan dapat memaksimumkan homogenitas dalam kluster. Sedangkan metode *average linkage* dapat menghasilkan kluster dengan varian dalam kluster yang kecil. Selain itu, kriteria penggabungan kluster pada metode *average linkage* tidak berdasarkan jarak ekstrim (terdekat dan terjauh) seperti pada *single linkage* dan *complete linkage*, melainkan berdasarkan nilai rata-rata jarak dari semua objek. Untuk menentukan metode terbaik diantara metode *ward* dan *average linkage* maka perlu dilakukan validasi. Dalam penelitian ini menggunakan validasi *dunn index* dan koefisien korelasi *cophenetic*. Metode pengklusteran dengan nilai *dunn index* dan koefisien korelasi *cophenetic* terbesar merupakan metode terbaik.

Penerapan metode analisis kluster telah banyak digunakan di berbagai bidang ilmu seperti biologi, kimia, ekonomi, kesehatan, dan sosial. Pada penelitian ini akan dilakukan pengelompokan jumlah kecelakaan lalu lintas berdasarkan jenis kendaraan. Kecelakaan lalu lintas kerap menjadi masalah utama dalam pemerintahan dan sosial. Penyebab utama kematian di Indonesia salah satunya adalah kecelakaan lalu lintas. Data kecelakaan lalu lintas Polda Jawa Tengah menyatakan bahwa, sepeda motor merupakan jenis kendaraan yang paling banyak terlibat kecelakaan apabila dibandingkan dengan jenis kendaraan lain seperti mobil penumpang, mobil

barang, mobil bus, dan kendaraan tidak bermotor. Metode analisis kluster digunakan untuk dapat mengetahui wilayah kabupaten/kota yang memiliki karakteristik tingkat kecelakaan yang sama berdasarkan jenis kendaraan.

## 2. TINJAUAN PUSTAKA

### 2.1. Analisis Kluster

Analisis kluster merupakan teknik multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya, sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam kluster yang sama. Kluster yang terbentuk memiliki homogenitas internal yang tinggi dan heterogenitas eksternal yang tinggi. Analisis kluster dapat mengelompokkan  $n$  objek berdasarkan  $p$  variabel yang secara relatif mempunyai kesamaan karakteristik diantara objek-objek tersebut, sehingga keragaman di dalam suatu kelompok lebih kecil dibandingkan keragaman antar kelompok. Apabila terdapat  $n$  objek dan  $p$  variabel, maka observasi  $x_{ik}$  dengan  $i = 1, 2, \dots, n$  dan  $k = 1, 2, \dots, p$  (Hair, *et al.*, 2010). Susunan observasi analisis kluster dapat digambarkan sebagai berikut:

**Tabel 1.** Layout Data Multivariat

	Variabel 1	Variabel 2	Variabel 3	...	Variabel p
Objek 1	$X_{11}$	$X_{12}$	$X_{13}$	...	$X_{1p}$
Objek 2	$X_{21}$	$X_{22}$	$X_{23}$	...	$X_{2p}$
Objek 3	$X_{31}$	$X_{32}$	$X_{33}$	...	$X_{3p}$
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
Objek n	$X_{n1}$	$X_{n2}$	$X_{n3}$	...	$X_{np}$

### 2.2. Tahapan Analisis Kluster

Berikut merupakan tahapan-tahapan dalam melakukan analisis kluster, tahapan pertama yang dilakukan yaitu dengan melakukan uji asumsi. Terdapat dua asumsi dalam analisis kluster yang harus dipenuhi yaitu:

#### a. Sampel Mewakili Populasi (*Sample Representative*)

Uji KMO ini mengukur kecukupan sampling secara menyeluruh dan mengukur kecukupan sampling untuk setiap indikator/variabel. Berikut uji hipotesis untuk melihat apakah sampel dapat mewakili populasi yang ada :

#### Hipotesis

$H_0$  : Sampel mewakili populasi atau sampel representatif

$H_1$  : Sampel tidak mewakili populasi atau sampel tidak representatif

#### Statistik Uji

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

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

dengan :

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

$$\rho_{X_k X_l, X_m} = \frac{r_{X_k X_l} - r_{X_k X_m} r_{X_l X_m}}{\sqrt{(1 - r_{X_k X_m}^2)(1 - r_{X_l X_m}^2)}} \quad (3)$$

#### Kriteria Uji

Sampel dikatakan dapat mewakili populasi atau sampel representatif apabila diperoleh nilai KMO berkisar antara 0,5 sampai dengan 1.

#### b. Tidak terjadi multikolinearitas

Multikolinieritas adalah suatu peristiwa dimana terjadi korelasi yang kuat antara dua atau lebih variabel kelompok. Multikolinieritas merupakan masalah yang perlu diperhatikan dalam analisis multivariat pada umumnya, karena pengaruhnya yang sangat besar dalam menghasilkan solusi, sehingga mengganggu proses analisis. Menurut Gujarati (2009), untuk mengetahui ada tidaknya multikolinieritas dapat digunakan nilai *Variance Inflation Factor* (VIF). Berikut uji hipotesis untuk melihat multikolinieritas data :

**Hipotesis**

H<sub>0</sub> : Tidak ada hubungan linier antar variabel

H<sub>1</sub> : Ada hubungan linier antar variabel

**Statistik Uji**

$$VIF_i = \frac{1}{1-R_k^2}, k = 1,2,\dots,p \tag{4}$$

dimana  $VIF_k$  adalah *Variance Inflation Factor* variabel  $k$ , dan  $R_k^2$  adalah koefisien determinasi yang diperoleh bila nilai variabel ke- $k$  diregresikan dengan variabel lainnya.

**Kriteria Uji**

Jika nilai VIF > 10 maka H<sub>0</sub> ditolak sehingga terjadi multikolinieritas antar variabel.

Tahapan selanjutnya yaitu menghitung jarak *squared euclidean*. Ukuran jarak *squared euclidean* adalah jumlah kuadrat perbedaan nilai untuk setiap variabel tanpa mengambil akar kuadratnya. Dengan tidak mengambil akar kuadratnya kembali, proses perhitungan menjadi lebih cepat. Selain itu ukuran jarak *squared euclidean* juga disarankan untuk metode *ward* dan dapat digunakan pada metode *average linkage* (Hair, et al. 2010). Jarak *Squared Euclidean* antara kelompok objek ke- $i$  dan kelompok objek ke- $j$  dari  $k$  variabel didefinisikan sebagai berikut:

$$d_{ij} = \sum_{k=1}^p (x_{ik} - x_{jk})^2 \tag{5}$$

dimana  $d_{ij}$  = jarak antara objek ke- $i$  dan objek ke- $j$ ,  $p$  = jumlah variabel klaster,  $x_{ik}$  = data objek ke- $i$  pada variabel ke- $k$ ,  $x_{jk}$  = data objek ke- $j$  pada variabel ke- $k$ .

Proses pembentukan klaster dapat dilakukan dengan dua cara, yaitu dengan metode hierarki dan non hierarki. Pada metode hierarki terdiri dari metode *agglomerative* dan metode *devisive*. Metode *agglomerative* terdiri dari tiga metode, yaitu metode *linkage*, metode *variance*, dan metode *centroid*. Dimana *linkage* terdiri dari metode *single linkage*, *complete linkage*, dan *average linkage*. Sedangkan pada metode *variance* terdiri dari metode *ward*. Metode non hierarki terdiri dari tiga metode, yaitu metode *sequential threshold*, *metode parallel threshold*, dan metode *optimizing partitioning*. Pada penelitian ini menggunakan metode *ward* dan *average linkage*.

a. Metode *Ward*

Metode *Ward* adalah metode varians dalam klastering hierarki yang bersifat agglomeratif, untuk memperoleh klaster yang memiliki varian internal sekecil mungkin. Jarak antara dua klaster yang terbentuk pada metode *Ward* adalah sum of squares di antara dua klaster tersebut. Metode *Ward* didasarkan pada kriteria sum squares error (SSE) dengan ukuran kehomogenan antara dua objek berdasarkan SSE yang paling minimal. SSE hanya dapat dihitung jika klaster memiliki elemen lebih dari satu objek. SSE klaster yang hanya memiliki satu objek adalah nol (Gudono, 2014). Berikut merupakan langkah pengklasteran dengan metode *ward*:

- 1) Menghitung SSE antara dua klaster dengan rumus

$$SSE_{ij} = \frac{1}{2} \sum_{k=1}^p (x_{ik} - x_{jk})^2 \tag{6}$$

dimana  $SSE_{ij}$ : *Sum Squares Error* antar pasangan objek  $i$  dan  $j$ ,  $x_{ik}$ : nilai objek  $i$  pada variabel ke- $k$ , dan  $x_{jk}$ : nilai objek  $j$  pada variabel ke- $k$

- 2) Mencari SSE terkecil antar dua klaster dan menggabungkan dua klaster tersebut menjadi satu klaster. Dengan demikian N klaster secara sistematis menjadi N-1.
- 3) Mengulang langkah (2) sampai diperoleh q klaster.

b. Metode *Average Linkage*

*Average linkage* adalah kesamaan rata-rata dari semua individu dalam satu kelompok dengan semua individu dikelompok lainnya. Metode ini tidak bergantung pada nilai ekstrim (pasangan terdekat atau terjauh) seperti halnya pada *single linkage* dan *complete linkage*. Pendekatan *average linkage*, sebagai jenis perpaduan antara metode *single linkage* dan *complete linkage*, sehingga cenderung menghasilkan kluster dengan variasi dalam kluster yang kecil (Hair et al, 2010). Berikut merupakan langkah pengklasteran dengan metode *average linkage*:

- 1) Tentukan kluster yang mempunyai jarak *squared euclidean* terdekat, misal objek U dan objek V mempunyai jarak terdekat kemudian gabungkan, hasil gabungannya adalah kluster UV.
- 2) Dari kluster UV yang terbentuk kemudian dicari jarak rata-rata antar kluster (UV) dengan objek lainnya (W)

$$d_{(uv)w} = \frac{\sum_i^{N(uv)} \sum_k^{N_w} d_{ij}}{N(uv)N_w} \quad (7)$$

dimana  $d_{ij}$  adalah jarak antara objek  $i$  dalam kluster (UV) dan objek  $j$  dalam kluster W, dan  $N_{uv}$  dan  $N_w$  berturut-turut adalah banyaknya objek dalam kluster (UV) dan W.

- 3) Diperoleh matriks jarak yang baru
- 4) Mengulang langkah (1) dan (2) sampai diperoleh q kluster.

Menurut Gudono (2014), tahapan selanjutnya setelah diperoleh hasil pengklasteran yaitu mengecek seberapa tinggi mutu klasterisasi tersebut dengan melihat kemampuan kluster tersebut untuk membedakan data yang ada sesuai dengan variabel atau karakteristik mengenai subjek yang digunakan untuk klasterisasi (Gudono, 2014). Berikut merupakan validasi yang akan digunakan:

a. *Dunn Index*

*Dunn Index* merupakan salah satu metode validasi kluster untuk evaluasi kuantitatif dari hasil klustering. Menurut Bazdek dalam Uyun dan Subanar (2007), *Dunn Index* adalah fungsi validitas yang mampu memberikan hasil penilaian yang efektif untuk aplikasi yang menggunakan beberapa metode clustering yang berbeda.

Menurut Dunn dalam Ansari (2011), rumus *Dunn Index* yaitu sebagai berikut:

$$DU = \left\{ \frac{\min_{\substack{1 \leq i \leq k \\ i+1 \leq j \leq q}} (d(C_i, C_j))}{\max_{1 \leq l \leq q} d(C_l)} \right\} \quad (8)$$

dengan

$$d(C_i, C_j) = \min_{i \in C_i, j \in C_j} d_{ij} \quad (9)$$

$$d(C_l) = \max_{i, j \in C_l} d_{ij} \quad (10)$$

dimana DU: *Dunn Index*, q: jumlah kluster,  $d(C_i, C_j)$ : jarak *squared euclidean* antar pasangan objek pada kluster  $i$  dan kluster  $j$  (*intercluster distance*),  $d(C_l)$ : jarak *squared euclidean* antar anggota dalam kluster  $l$  (*intracluster distance*),  $d_{ij}$ : jarak *squared euclidean* antara objek  $i$  dan objek  $j$ .

Menurut Dunn dalam Iezzi (2012), nilai *dunn index* berada pada interval [0,1]. Semakin tinggi nilai *Dunn Index* maka semakin baik kluster yang terbentuk.

b. Koefisien Korelasi *Cophenetic*

Salah satu ukuran yang dapat digunakan untuk menguji validitas hasil clustering adalah koefisien korelasi *cophenetic*. Koefisien korelasi *cophenetic* merupakan koefisien korelasi antara elemen-elemen asli matriks ketidakmiripan (matriks jarak *squared euclidean*) dan elemen-elemen yang dihasilkan oleh dendrogram (matriks *cophenetic*) (Silva

dan Dias, 2013). Menurut Saracli, dkk (2013), berikut merupakan formula untuk menghitung koefisien korelasi *cophenetic*:

$$r_{coph} = \frac{\sum_{i < j} (d_{ij} - \bar{d})(d_{cij} - \bar{d}_c)}{\sqrt{[\sum_{i < j} (d_{ij} - \bar{d})^2][\sum_{i < j} (d_{cij} - \bar{d}_c)^2]}} \quad (11)$$

dimana  $r_{coph}$ : koefisien korelasi *cophenetic*,  $d_{ij}$ : jarak *Squared Euclidean* antara objek  $i$  dan  $j$ ,  $\bar{d}$ : rata-rata jarak *squared euclidean* antara objek  $i$  dan  $j$ ,  $d_{cij}$ : jarak *cophenetic* antara objek  $i$  dan  $j$ ,  $\bar{d}_c$ : rata-rata jarak *cophenetic* antara objek  $i$  dan  $j$ .

Nilai koefisien korelasi *cophenetic* berkisar antara -1 dan 1, nilai  $r_{coph}$  mendekati 1 berarti solusi yang dihasilkan dari proses klustering baik.

Langkah terakhir yaitu menginterpretasi dan membuat profil kluster meliputi pengujian pada masing-masing kluster yang terbentuk untuk memberikan nama atau keterangan secara tepat sebagai gambaran sifat dari kluster tersebut. Menjelaskan bagaimana mereka bisa berbeda secara relevan pada tiap dimensi. Ketika memulai proses interpretasi digunakan rata-rata *centroid* setiap kluster pada setiap variabel (Supranto, 2004).

### 3. METODOLOGI PENELITIAN

#### 3.1. Sumber dan Jenis Data

Data yang digunakan merupakan data sekunder yang diperoleh dari Kepolisian Daerah Jawa Tengah. Data tersebut merupakan data kecelakaan lalu lintas berdasarkan jenis kendaraan sepeda motor, mobil penumpang, mobil bus, mobil barang, dan kendaraan tidak bermotor tiap kabupaten/kota di Provinsi Jawa Tengah tahun 2018.

#### 3.2. Variabel Penelitian

Dalam penelitian ini data dikelompokkan menurut 35 kabupaten/kota di Provinsi Jawa Tengah. Adapun variabel-variabel yang digunakan yaitu:

- X<sub>1</sub>: Sepeda Motor
- X<sub>2</sub>: Mobil Penumpang
- X<sub>3</sub>: Mobil Bus
- X<sub>4</sub>: Mobil Barang
- X<sub>5</sub>: Kendaraan Tidak Bermotor

#### 3.3. Teknik Analisis Data

Pada penelitian ini digunakan analisis kluster dengan metode pengklasteran *ward* dan *average linkage*. Data penelitian ini diolah dengan menggunakan software R i386 3.3.2.

Tahapan analisis data yang digunakan adalah sebagai berikut:

1. Mengumpulkan data jumlah kecelakaan lalu lintas dari Kepolisian Daerah Jawa Tengah
2. Melakukan pengujian asumsi analisis kluster
  - a. Uji sampel mewakili populasi
  - b. Tidak Terjadi Multikolinearitas
3. Melakukan pengukuran jarak *squared euclidean* antar objek.
4. Setiap data objek dianggap sebagai kluster, dimana N adalah jumlah objek, dalam penelitian ini N=35.
5. Membentuk matriks jarak *Squared euclidean* berdasarkan data tersebut dengan ukuran N × N.
6. Melakukan proses pengklasteran dengan metode *ward*
  - a. Menghitung SSE antara dua kluster.
  - b. Mencari SSE terkecil antar dua kluster dan menggabungkan dua kluster tersebut menjadi satu kluster. Dengan demikian N kluster secara sistematis menjadi N-1.
  - c. Mengulang langkah (b) sampai diperoleh jumlah kluster 3,4, dan 5
7. Melakukan proses pengklasteran dengan metode *average linkage*

- c. Tentukan kluster yang mempunyai jarak *squared euclidean* terdekat, misal objek U dan objek V mempunyai jarak terdekat kemudian gabungkan, hasil gabungannya adalah kluster UV.
  - d. Dari kluster UV yang terbentuk kemudian dicari jarak rata-rata antar kluster (UV) dengan objek lainnya (W)
  - e. Diperoleh matriks jarak yang baru
  - f. Mengulang langkah (a) dan (b) sampai diperoleh jumlah kluster 3, 4, dan 5.
8. Menentukan kebaikan metode pengklasteran dengan:
- a. *Dunn Index*
  - b. Koefisien Korelasi *Cophenetic*
9. Melakukan interpretasi hasil analisis.

#### 4. ANALISIS DAN PEMBAHASAN

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

1) Sampel Mewakili Populasi

Berikut merupakan uji hipotesis untuk melihat apakah sampel dapat mewakili populasi yang ada:

- a. Hipotesis
  - $H_0$  : Sampel mewakili populasi atau sampel representatif
  - $H_1$  : Sampel tidak mewakili populasi atau sampel tidak representatif
- b. Statistik Uji
 

Berdasarkan hasil pengujian, diperoleh nilai KMO sebesar 0,6309859.
- c. Kriteria Uji
 

Sampel dikatakan mewakili populasi atau sampel representatif apabila diperoleh nilai KMO berkisar antara 0,5 sampai dengan 1, sehingga  $H_0$  diterima.
- d. Kesimpulan
 

Berdasarkan uji sampel mewakili populasi, diperoleh nilai KMO sebesar 0,6309858. Dimana nilai tersebut berkisar antara 0,5 sampai dengan 1. Jadi, dapat disimpulkan bahwa  $H_0$  diterima, maka sampel mewakili populasi atau sampel representatif.

2) Multikolinearitas

Berikut uji hipotesis untuk melihat multikolinearitas data:

- a. Hipotesis
  - $H_0$  : Tidak ada hubungan linier antar variabel (tidak terjadi multikolinearitas)
  - $H_1$  : Ada hubungan linier antar variabel (terjadi multikolinearitas)
- b. Statistik Uji
 

Berdasarkan hasil pengujian multikolinearitas, diperoleh hasil sebagai berikut:

**Tabel 1.** Nilai VIF Masing-Masing Variabel

Variabel	VIF	Keterangan
X1	3,688580	$VIF \leq 10$
X2	3,139576	$VIF \leq 10$
X3	1,922520	$VIF \leq 10$
X4	1,668102	$VIF \leq 10$
X5	1,513172	$VIF \leq 10$

- c. Kriteria Uji
 

Jika nilai  $VIF > 10$  maka  $H_0$  ditolak sehingga terjadi multikolinearitas antar variabel.
- d. Kesimpulan
 

Berdasarkan *output* pengujian multikolinearitas, diperoleh nilai  $VIF \leq 10$  untuk kelima variabel tersebut, maka  $H_0$  diterima. Jadi tidak terjadi multikolinearitas pada kelima variabel tersebut.

Tahap selanjutnya yaitu melakukan perhitungan jarak menggunakan ukuran jarak *squared euclidean* dengan rumus seperti pada persamaan 5. Jarak *Squared Euclidean* dihitung antar pasangan objek  $i$  dan objek  $j$  pada variabel ke- $k$ . Karena terdapat 35 objek atau kabupaten/kota maka matriks jarak *squared euclidean* yang terbentuk berukuran  $35 \times 35$ .

Proses pengklasteran pada tugas akhir ini mengelompokkan kabupaten/kota di Provinsi Jawa Tengah berdasarkan jumlah kecelakaan lalu lintas menurut jenis kendaraan dengan menggunakan metode *ward* dan *average linkage* dengan jumlah klaster 3, 4, dan 5. Kemudian akan dipilih jumlah klaster dan metode terbaik berdasarkan nilai validasi *dunn index* dan koefisien korelasi *cophenetic*. Berikut ini merupakan langkah proses pengklasteran dengan menggunakan metode *ward* dan *average linkage*:

1. Metode *Ward*

Proses perhitungan pengklasteran dengan metode *ward* adalah sebagai berikut:

- a. Menghitung *Sum Squares Error* antar pasangan objek ( $SSE_{ij}$ ) dengan menggunakan persamaan 6. Karena terdapat 35 objek atau kabupaten/kota maka matriks  $SSE_{ij}$  yang terbentuk berukuran  $35 \times 35$
- b. Setelah nilai *Sum Squares Error* antar pasangan objek diperoleh, maka langkah selanjutnya yaitu penggabungan antar objek (kabupaten/kota) dengan nilai  $SSE_{ij}$  terkecil.
- c. Mengulang langkah (b) hingga terbentuk 3, 4, dan 5 klaster.

2. Metode *Average Linkage*

Proses perhitungan pengklasteran dengan metode *average linkage* adalah sebagai berikut:

- a. Menentukan pasangan objek yang memiliki jarak *squared euclidean* terdekat, misal objek  $U$  dan objek  $V$  mempunyai jarak terdekat kemudian digabungkan. Hasil penggabungannya adalah klaster  $UV$ .
- b. Dari klaster  $UV$  yang terbentuk kemudian dicari jarak rata-rata antar klaster ( $UV$ ) dengan objek lainnya ( $W$ ) menggunakan persamaan 7.
- c. Diperoleh matriks jarak baru
- d. Kemudian mengulang langkah (a) dan (b) hingga terbentuk 3, 4, dan 5 klaster.

Langkah selanjutnya, setelah diperoleh hasil pengklasteran yaitu menentukan metode pengklasteran terbaik dengan melakukan uji *dunn index* dan koefisien korelasi *cophenetic*.

1. Validasi *Dunn Index*

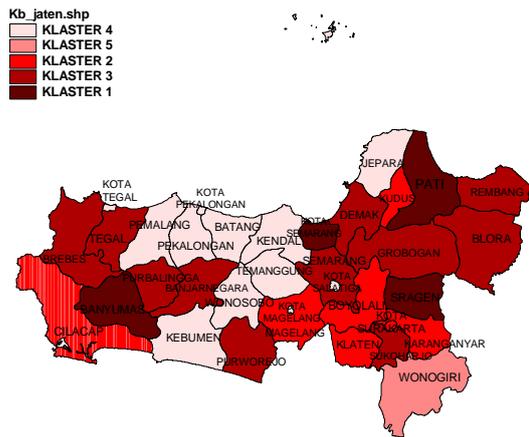
Berikut merupakan hasil perhitungan validasi *dunn index* pada metode *ward* dan *average linkage*:

**Tabel 2.** Nilai *Dunn Index* pada Metode *Ward* dan *Average linkage*

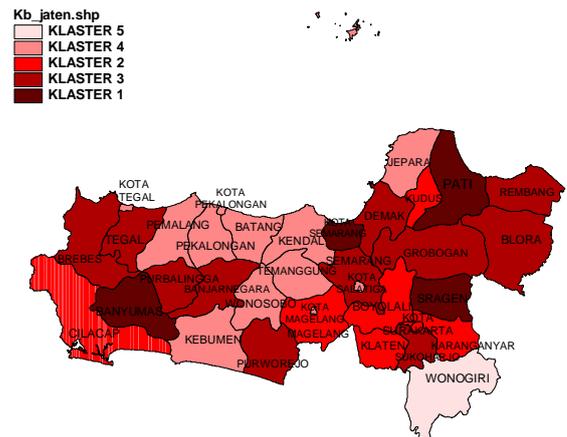
Metode Pengklasteran	Klaster	Nilai <i>Dunn Index</i>
<i>Ward</i>	3	0.0220435
	4	0.0736874
	5	0.01660025
<i>Average linkage</i>	3	0.0220435
	4	0.0736874
	5	0.08571196

Berdasarkan, nilai *dunn index* untuk jumlah klaster lima pada metode *average linkage* merupakan nilai yang paling tinggi yaitu 0.08571196. Nilai *dunn index* yang tinggi menunjukkan bahwa klaster telah terpisah dengan baik. Meskipun nilai *dunn index* pada klaster lima tidak mendekati satu namun nilai ini lebih tinggi di bandingkan nilai *dunn index* pada klaster lainnya. Sehingga dapat disimpulkan jumlah klaster lima adalah jumlah yang optimal untuk pengelompokan kabupaten/kota berdasarkan jenis kendaraan yang terlibat kecelakaan di jawa tengah. Untuk lebih memastikan agar pengklasteran yang diperoleh baik. Maka dilakukan validasi kembali menggunakan koefisien korelasi *cophenetic*.

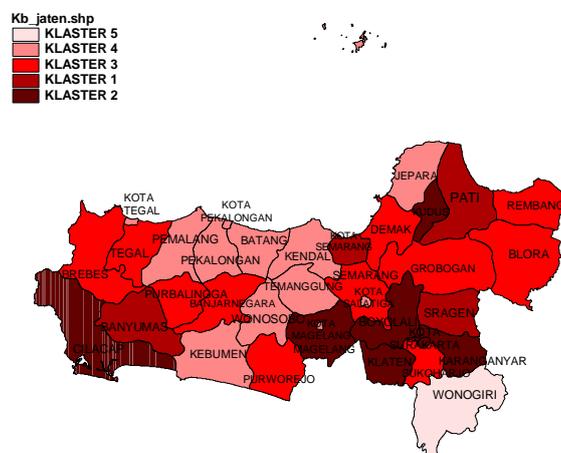




Gambar 3. Peta Pengklasteran Mobil Bus



Gambar 4. Peta Pengklasteran Mobil Barang



Gambar 5. Peta Pengklasteran Kendaraan Tidak Bermotor

Peta hasil pengklasteran kelima variabel menunjukkan bahwa wilayah pada klaster satu yaitu Kabupaten Banyumas, Kabupaten Pati, Kabupaten Sragen dan Kota Semarang terjadi kecelakaan lalu lintas sangat tinggi pada jenis kendaraan sepeda motor, mobil penumpang, mobil bus dan mobil barang dengan visualisasi wilayah berwarna merah tua yang dapat dilihat pada gambar 11 sampai gambar 14. Kecelakaan lalu lintas pada klaster dua tertinggi pada jenis kendaraan tidak bermotor dengan visualisasi wilayah berwarna merah tua yang dapat dilihat pada gambar 15. Klaster tiga merupakan wilayah dengan kecelakaan lalu lintas berdasarkan jenis kendaraan sepeda motor dan mobil penumpang cukup tinggi yang ditunjukkan pada gambar 11 dan 12, jenis kendaraan mobil barang dan mobil bus cukup rendah seperti pada gambar 13 dan 14, serta kecelakaan lalu lintas sedang pada jenis kendaraan tidak bermotor yang ditunjukkan seperti pada gambar 15.

Wilayah pada klaster empat merupakan wilayah dengan jumlah kecelakaan lalu lintas menurut jenis kendaraan sepeda motor, mobil penumpang, dan mobil bus paling rendah seperti pada gambar 11 sampai gambar 13, sedangkan pada jenis kendaraan mobil barang dan kendaraan tidak bermotor cukup rendah ditunjukkan pada gambar 14 dan gambar 15. Kabupaten Wonogiri merupakan anggota klaster lima, dengan jumlah kecelakaan lalu lintas berdasarkan jenis kendaraan mobil barang dan kendaraan tidak bermotor paling rendah, jenis kendaraan mobil bus cukup rendah, jenis kendaraan sepeda motor sedang, serta jenis kendaraan mobil penumpang cukup tinggi.

## 5. KESIMPULAN

- a. Berdasarkan pengklasteran menggunakan metode *ward* dan *average linkage* dengan jarak *squared euclidean* diperoleh hasil pengklasteran yang berbeda. Semakin tinggi nilai validasi *dunn index* dan koefisien korelasi *cophenetic* yang diperoleh, maka klaster yang dihasilkan semakin heterogen. Metode *average linkage* memiliki nilai validasi yang lebih besar sehingga menghasilkan klaster yang heterogen dibandingkan dengan metode *ward*.
- b. Pengklasteran menggunakan metode *ward* dan *average linkage* dengan  $q = 3, 4, \text{ dan } 5$  diperoleh nilai *dunn index* tertinggi pada metode *average linkage* untuk jumlah 5 klaster dengan nilai *dunn index* sebesar 0,08571196 dan nilai  $r_{\text{coph}}$  yaitu 0,687458.
- c. Berdasarkan hasil pengklasteran diketahui bahwa variabel sepeda motor merupakan jenis kendaraan yang mengalami kecelakaan lalu lintas tertinggi pada klaster 1 hingga klaster 4 dan variabel mobil penumpang merupakan jenis kendaraan yang mengalami kecelakaan lalu lintas tertinggi pada klaster 5, untuk itu maka perlu dilakukan sosialisasi kepada pengendara sepeda motor dan mobil penumpang agar dapat meningkatkan kesadaran berkendara dengan baik dan benar.

## DAFTAR PUSTAKA

- Ansari, Z., Babu, A.V., Azeem, M.F. dan Ahmed, W. 2011. *Quantitive Evaluation of Performance and Validity Indices for Clustering the Web Navigational Sessions*. World of Computer Science and Information Technology Journal (WCSIT) Vol. 1, No. 5: Hal. 217-226
- Gudono. 2014. *Analisis Data Multivariat*. Yogyakarta: BPFE
- Gujarati, D. 2009. *Dasar-dasar Ekonometrika Jilid 2*. Jakarta : Erlangga
- Hair, J.F.Jr., R.E. Anderson, R.L. Thatham dan Black, W.C. 2010. *Multivariat Data Analysis Seventh Edition*. New Jersey: Pearson Education, Inc
- Johnson, R. A. dan Wincern, D. W. 2007. *Applied Multivariate Statistical Analysis Sixth Edition*. New Jersey: Prentice Hal, Inc
- Saber, G.A.F. 2004. *Multivariate Observation*. New Jersey: Wiley Interscience
- Saracli, S., Dogan, N., dan Dogan, I. 2013. *Comparison of Hierarchical Cluster Analysis Methods by Cophenetic Correlation*. Journal of Inequalities and Applications, Hal. 203.
- Silva, A.R. dan Dias, C.T. 2013. *A Cophenetic Correlation Coefficient for Tocher's Method*. Pesq Agropec. Brasilia Vol. 48, No. 6: Hal. 589-596.
- Supranto, J. 2004. *Analisis Multivariat Arti dan Interpretasi*. Jakarta: Rineka Cipta
- Uyun, S. dan Subanar. 2007. *Stock Data Clustering of Food and Beverage Company*. IJCCS Vol. 1, No. 2: Hal. 123-130
- Widarjono, A. 2010. *Analisis Statistika Multivariat Terapan*. Edisi pertama. Yogyakarta: UPP STIM YKPN