

## **PEMANFAATAN ALGORITMA K-MEANS UNTUK PEMETAAN HASIL KLASTERISASI DATA KECELAKAAN LALU LINTAS**

**Lizda Iswari<sup>1</sup>, Ervina Gita Ayu<sup>2</sup>**

*Jurusan Teknik Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia  
Jalan Kaliurang Km.14,5 Sleman, Yogyakarta 55501  
[lizdaiswari@gmail.com](mailto:lizdaiswari@gmail.com)<sup>1</sup>, [ervina.gita@gmail.com](mailto:ervina.gita@gmail.com)<sup>2</sup>*

### **ABSTRACT**

*It is a vital importance to analyze road traffic accidents in order to improve traffic security management. Currently, most of the traffic information analysis is limited to general statistical analysis, which is hard to explore the rules hiding in its dataset and also difficult to find the spatial distribution characteristics. This paper aims to analyze the road traffic accidents dataset based on data mining method of K-Means clustering and visualize the result as a map. Firstly, data are extracted for clustering road segments based on similar characteristics that lies on the dataset, i.e. the number of accidents, the number vehicles involved, and the number accidents' victims. Secondly, the result of clustering are presented as a map that aims to assist the police officer in identifying and evaluating some black spot areas (accident prone areas) in a monthly period, hence monitoring the safety of highways users can be anticipated earlier.*

*Keywords: Prone Areas, Traffic accidents, K-Means, Mapping*

### **1. PENDAHULUAN**

Berita tentang kecelakaan lalu lintas adalah salah satu topik yang senantiasa menjadi pokok pembicaraan di masyarakat. Secara statistik, angka kecelakaan lalu lintas dominan meningkat dari tahun ke tahun. Data Kepolisian RI menyebutkan pada tahun 2012 terjadi 117.949 kecelakaan lalu lintas di Indonesia dengan 29.544 orang tewas karenanya. Angka kecelakaan ini meningkat 8,51% dibandingkan tahun sebelumnya. Bahkan jika ditinjau dari perbandingan dua tahun sebelumnya terdapat peningkatan jumlah kecelakaan lalu lintas sebesar 63% (BPS, 2014).

Hal ini jelas perlu mendapat perhatian dan penanganan efektif karena berkaitan dengan kebijakan yang diterapkan oleh Direktorat Lalu Lintas (Ditlantas) pada masing-masing wilayah Kepolisian Daerah (Polda). Kebijakan yang diambil tentunya haruslah memiliki relevansi dan didukung oleh pengetahuan yang berasal dari data yang tersedia. Informasi utama yang direkam pada suatu kecelakaan lalu lintas adalah dimana, kapan, dan bagaimana kecelakaan terjadi. Hampir seluruh Ditlantas memiliki sistem atau mekanisme untuk mengarsipkan kejadian kecelakaan lalu lintas

dari waktu ke waktu. Berdasarkan rekaman data inilah secara rutin pihak Ditlantas merekapitulasi jumlah kecelakaan, jumlah korban, dan total kerugian materiil di suatu wilayah untuk kemudian dianalisis dalam menentukan daerah yang rawan terhadap kecelakaan lalu lintas.

Umumnya analisis data kecelakaan lalu lintas disajikan sebagai sebuah informasi hasil olahan statistik deskripsi (Wedasana, 2011; Depkimpraswil, 2004; Bolla, Messah, & Koreh, 2013), yang meliputi: (a) distribusi frekuensi, (b) data berkala yaitu data yang disusun berdasarkan urutan waktu sehingga terbentuk tren, (c) pembobotan yaitu suatu nilai yang digunakan untuk menghitung indeks kecelakaan berdasarkan karakteristik masing-masing kecelakaan, (d) teknik *z-score* yaitu analisis yang didasarkan pada bilangan baku, atau (e) teknik cumulative summary yaitu suatu prosedur yang digunakan untuk mengidentifikasi lokasi kecelakaan, ataupun (f) analisis diagram tongkat yang digunakan untuk mengelompokkan tipe kecelakaan yang sejenis. Pengolahan data dengan teknik-teknik demikian baru mengungkap sebagian kecil informasi yang tersembunyi di dalam basis data. Informasi penting yang mendukung dalam pengambilan

kebijakan untuk mengurangi bahkan mencegah terjadinya kecelakaan lalu lintas, seperti pola-pola penyebab kecelakaan dan tren yang berkembang akibat kecelakaan belumlah dapat disajikan (Li & Song, 2012). Selain itu, keterbatasan lain juga terletak pada ketidakmampuan untuk menunjukkan aturan-aturan sebab akibat antar parameter ataupun untuk mengenali kesamaan fenomena yang mungkin tersembunyi di dalam data (Chen, Chun, Wu, & Sun, 2011).

Dalam dunia ilmu komputer, dikenal luas data mining sebagai teknik untuk meringkas data dengan cara yang berbeda dengan yang biasa diterapkan, menemukan hubungan yang tidak diduga, menemukan pola yang dapat dipahami dan bermanfaat bagi pemilik data (Larose, 2005). Sebelumnya telah dilakukan sejumlah penelitian berkaitan dengan pemanfaatan data mining dalam mengolah dataset kecelakaan lalu lintas di Indonesia, seperti pemanfaatan teknik Apriori (Harahap, 2013) dan Naïve Bayes (Yunanto, Hariadi, & Purnomo, 2012) untuk memprediksi kecelakaan lalu lintas, serta pemanfaatan teknik Asosiasi (Ransi, 2014) yang juga berkaitan dengan prediksi kecelakaan berdasarkan aturan kejadian yang terdapat di dalam dataset. Namun, penelitian yang berkaitan dengan klasterisasi data sebagai dasar untuk mengenali daerah yang rawan kecelakaan lalu lintas masih sedikit sehingga diperlukan penelitian berkaitan dengan hal tersebut.

Data kecelakaan lalu lintas sangat berhubungan erat dengan data yang bersifat spasial temporal, yaitu memiliki informasi utama berupa lokasi dan waktu kejadian. Data yang bersifat demikian akan sangat efektif jika disajikan sebagai sebuah peta (Kraak & Ormeling, 2010). Peta merupakan media grafis yang mampu menggambarkan hubungan keruangan dan fenomena georafis dalam bentuk simbol-simbol dua dimensi. Terutama dalam kajian masalah kecelakaan lalu lintas, jalan raya dapat direpresentasikan dengan simbol garis dan warna untuk menggambarkan fenomena tingkat kerawanan lalu lintas yang melekat pada ruas jalan tersebut.

Penelitian ini mengangkat kolaborasi tema antara penggalian dataset kecelakaan lalu lintas

dan penyajian informasi sebagai peta dua dimensi. Penggalian data ditujukan untuk mengelompokkan (klasterisasi) ruas-ruas jalan berdasarkan faktor kesamaan karakteristik yang melekat di dataset tersebut, yaitu jumlah kecelakaan, jumlah kendaraan yang terlibat, dan jumlah korban kecelakaan dalam suatu rentang waktu tertentu. Salah satu metode klasterisasi yang populer dan dapat digunakan untuk mencapai tujuan ini adalah algoritma K-Means. Adapun obyek penelitian adalah sejumlah ruas jalan utama di Kabupaten Sleman, DIY. Hasil klastering yang telah dianalisis oleh seorang pakar lalu lintas kemudian disajikan dalam bentuk peta yang bertujuan untuk membantu pihak Ditlantas dalam mengidentifikasi dan mengevaluasi wilayah-wilayah yang memiliki kemiripan karakteristik kejadian kecelakaan lalu lintas. Diharapkan melalui penelitian ini, monitoring terhadap keselamatan pengendara di jalan raya dapat diantisipasi sejak dini.

## **2. DAERAH RAWAN KECELAKAAN LALU LINTAS**

Kecelakaan lalu lintas didefinisikan sebagai suatu peristiwa di jalan raya yang tidak diduga dan tidak disengaja melibatkan kendaraan dengan atau tanpa pemakai jalan lainnya, yang mengakibatkan korban manusia atau kerugian harta benda. Terdapat lima jenis dan bentuk kecelakaan lalu lintas, yaitu: berdasarkan korban kecelakaan, berdasarkan lokasi kejadian, berdasarkan waktu terjadinya, berdasarkan posisi kecelakaan, dan berdasarkan jumlah kendaraan yang terlibat (Depkimpraswil, 2004).

Di Indonesia, pihak berwenang untuk merekam kejadian kecelakaan lalu lintas adalah Satuan Lalu Lintas (Satlantas) Direktorat Lalu Lintas Kepolisian RI. Terdapat tujuh informasi utama yang diarsipkan sebagai dokumen digital seperti tampak pada Gambar 1, yang meliputi:

- a. Nomor berkas perkara
- b. Hari, waktu, dan tempat kejadian
- c. Jenis, merk, dan nomor plat kendaraan yang terlibat kecelakaan (Antara)
- d. Uraian singkat kejadian
- e. Kondisi korban kecelakaan yang terbagi 3 jenis, yaitu: meninggal (MB), luka berat (LB), atau luka ringan (LR).

- f. Kerugian materiil (Kermat)
- g. Tindakan yang telah dilakukan oleh pihak Kepolisian (Giat / CB)

NO	HARI/TGL /JAM /LOKASI /NO LP	A N T A R A	URAIAN SINGKAT KEJADIAN	KORBAN			KERMAT	GIAT / CB
				MD	LB	LR		
1	2	3	4	5	6	7	8	9
1	Rabu, 30 Nov 2011 sekira jam 10.00 Wib tkp Jl Magelang km 12 Dsn Wadas, Tridadi, Sleman (dilaporkan 30 Nov 2011) Lp/1505/XII/2011/Sleman -7.657319 110.326166	Spm Honda Kirana AB 3090 IZ X Spm Honda vario AB 2137 AW	----- Semula Spm Honda Vario dan Sepeda Angin sama-sama melaju dari timur ke barat dengan posisi Spm Honda Vario dibelakang sesampainya di tkp bersamaan Spm Honda Kirana melaju dari utara menyeberang ke selatan hendak ke barat karena jarak sudah dekat sehingga kedua SPm bersenggolan dan terpentak membentur Sepeda ANgin yang melaju didepannya dan terjadilah kecelakaan lalu lintas <u>Pengemudi SPm Honda Kirana AB 3090 IZ :</u> Nama : SURATNO, 39 Th, Lk, Swasta alamat Jetis 01/30 Jogopaten, Pandowoharjo, Sleman dan memiliki Sim C <u>Pengemudi SPm Honda Vario AB 2137 AW :</u> Nama : WARTANA, 36 Th, Lk, PNS alamat Karangduwet I Rt 16 Karangrejek, Wonosari, GK mengalami LB RS Bethesda dan memiliki Sim C <u>Pengayuh Sepeda Angin :</u> Nama : WAHYU DEWANTO, 36 Th, LK, Swasta alamat Dukuh, Tridadi, Sleman	-	1	-	300.000	Mendatangi tkp, mengamankan barang bukti, tersangka, mencatat saksi saksi, mengecek korban, membuat Laporan Polisi dan sket gambar tkp.

Gambar 6. Rekaman Data Kecelakaan Lalu Lintas  
(Sumber: Ditlantas Polres Sleman, DIY)

Dari data tersebut, pihak Ditlantas setiap bulan merekapitulasi jumlah kecelakaan, jumlah korban yang meninggal dunia, luka berat, luka ringan, dan total kerugian secara materiil. Hasil rekapitulasi ini merupakan salah satu informasi utama untuk mengidentifikasi daerah rawan kecelakaan lalu lintas. Menurut Austroads (1992) dalam Wedasana (2011), penentuan daerah rawan kecelakaan idealnya memperhitungkan data historis, sehingga pihak Ditlantas biasanya juga berpedoman pada jumlah kecelakaan dalam beberapa tahun terakhir.

Daerah rawan kecelakaan adalah suatu lokasi dimana angka kecelakaan tinggi dengan kejadian kecelakaan berulang dalam suatu ruang dan rentang waktu yang relatif sama, diakibatkan oleh suatu penyebab tertentu. Suatu daerah dinyatakan sebagai lokasi rawan kecelakaan lalu lintas apabila memiliki kriteria sebagai berikut (Depkimpraswil, 2004):

1. Memiliki angka kecelakaan yang tinggi;
2. Lokasi kejadian kecelakaan relatif menumpuk;
3. Lokasi kecelakaan berupa simpangan atau segmen ruas jalan sepanjang 100-300 m untuk jalan perkotaan, ruas jalan sepanjang 1 km untuk jalan antar kota;
4. Kecelakaan terjadi dalam ruang dan rentang waktu yang relatif sama;
5. Memiliki penyebab kecelakaan dengan faktor yang spesifik.

Untuk mengidentifikasi daerah rawan kecelakaan ada dua tahapan yang harus dilewati (Wedasana, 2011), yaitu:

- a. Mempelajari sejarah kecelakaan (*accident history*) dari seluruh wilayah studi untuk kemudian memilih lokasi yang dianggap rawan kecelakaan.
- b. Mempelajari secara detil lokasi terpilih untuk menemukan penanganan yang dapat dilakukan.

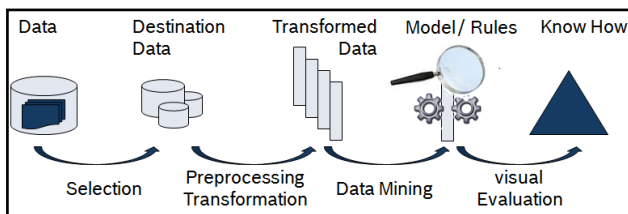
### 3. KLASTERISASI DATA KECELAKAAN LALU LINTAS

Pencarian pengetahuan dalam data, dikenal juga sebagai *Knowledge discovery in Database* (KDD), didefinisikan sebagai ekstraksi data yang memiliki potensi informasi berharga yang implisit dan tidak dikenali sebelumnya. Terdapat sejumlah tahapan dalam proses KDD, Namun, pada dasarnya ada tiga tahapan utama (Fayyad, 1996) seperti tampak pada Gambar 1, yaitu:

- a. Pra-prosesing, berkaitan dengan pengumpulan dan pengambilan data (*data collection*), pembersihan data (*data cleaning*), dan pemilihan dan transformasi data (*data selection and tranformation*).
- b. Data mining, berkaitan dengan proses eksplorasi data untuk menemukan pola atau aturan yang belum dikenali sebelumnya,

dapat diinterpretasi, dan memiliki potensi untuk dimanfaatkan (Han & Kamber, 2005). Terdapat sejumlah model atau teknik yang dapat digunakan untuk menemukan pola tersebut, seperti: *anomaly detection*, *association rule learning*, *clustering*, *classification*, *regression*, dan *summarization*.

- c. Pos-prosesing, berkaitan dengan evaluasi hasil dan memvisualisasikannya dalam bentuk yang mudah dipahami pengguna.



Gambar 1. Tahapan *Knowledge Discovery in Database*

Berdasarkan tahapan pencarian pengetahuan tersebut, maka proses penggalian data kecelakaan lalu lintas juga dilaksanakan dalam tiga tahap sebagai berikut:

### 3.1 Tahap Pra-Prosesing

Hal yang dilakukan dalam tahap pra-prosesing adalah pemilihan kriteria (*selection*) dan normalisasi data (*transformation*) sehingga data siap untuk diklasterisasi.

Dari data kecelakaan lalu lintas, terdapat sejumlah informasi utama yang dipilih untuk dijadikan sebagai parameter atau kriteria klastering, yaitu: waktu dan lokasi kejadian, kendaraan yang terlibat, serta kondisi korban akibat kecelakaan. Informasi tersebut umumnya disimpan dalam bentuk uraian/deskripsi seperti tampak pada Gambar 1.

Untuk kebutuhan komputasi, informasi yang bersifat deskriptif (nominal) perlu diubah dalam bentuk tipe data ordinal, interval, atau rasio. Dalam hal ini berlaku proses simplifikasi dan generalisasi data untuk mendapatkan informasi yang dapat diolah lanjut.

Adapun penerapan proses tersebut adalah sebagai berikut:

- a. Waktu kejadian disimplifikasi (informasi utama yang diambil) adalah bulan dan tahun kejadian.
- b. Lokasi kejadian digeneralisasi untuk beberapa jalan utama.
  - i. Dalam penelitian ini lokasi kejadian kecelakaan di Kabupaten Sleman DIY yang dikenali adalah jalan kolektor, yaitu: Jalan Ring Road Utara, Jalan Ring Road Selatan, Jalan Magelang, Jalan Kaliurang, Jalan Jogja Solo, Jalan Wates, Jalan Palagan, dan seterusnya yang dibatasi untuk 24 objek jalan utama di Kabupaten Sleman.
  - ii. Untuk sebuah jalan yang memiliki panjang lebih dari 10 km akan dibagi menjadi beberapa segmen atau dipecah menjadi beberapa objek jalan. Contoh Jalan Kaliurang memiliki panjang kurang lebih 20 kilometer, sehingga jalan perlu dibagi menjadi 4 segmen, yaitu Jalan Kaliurang kilometer 0-5, Jalan Kaliurang kilometer 5-10, dan Jalan Kaliurang kilometer 15-20.
- c. Kendaraan yang terlibat kecelakaan disimplifikasi sebagai total jumlah kendaraan yang terlibat dalam kecelakaan lalu lintas.
- d. Kondisi korban akibat kecelakaan disimplifikasi sebagai total jumlah korban yang terlibat. Untuk kriteria ini tidak dibedakan kondisi korban meninggal, luka berat atau luka ringan.

Berikut adalah contoh simplifikasi dan generalisasi data dari laporan kecelakaan Ditlantas Polres Sleman, DIY. Dari uraian kecelakaan yang tampak pada Gambar 1 diambil sejumlah informasi sebagai berikut:

- a. Waktu kejadian: Bulan Nopember Tahun 2011.
- b. Lokasi kejadian: Jalan Magelang km 10-15.
- c. Jumlah kendaraan yang terlibat: 2 buah.
- d. Jumlah korban: 1 orang.

Contoh hasil rekapitulasi kecelakaan lalu lintas pada 10 segmen jalan dapat dilihat pada Tabel 1. Setiap jalan memiliki tiga parameter, yaitu jumlah kecelakaan (kejadian), jumlah kendaraan yang terlibat, dan jumlah korban.

Setelah melalui tahapan simplifikasi dan generalisasi, langkah selanjutnya adalah transformasi data, yaitu proses data dinormalisasi ke dalam rentang nilai 0-1. Normalisasi data sangat dibutuhkan sebelum proses data mining, agar tidak ada parameter yang mendominasi dalam perhitungan jarak antar data (Han & Kamber, 2005). Adapun rumus untuk normalisasi dapat dilihat pada rumus (1) dan contoh hasil normalisasi data dapat dilihat pada Tabel 2.

$$N_{ij} = \left( \frac{x_{ij} - \min_{ij}}{\max_j - \min_j} \right) \quad (1)$$

Dengan:

$X_{ij}$ : Nilai awal data

$N_{ij}$ : Hasil normalisasi

$\min_j$ : Nilai minimum untuk kriteria (j)

$\max_j$ : Nilai maksimum untuk kriteria (j)

Sebagai contoh untuk parameter pertama berupa jumlah kecelakaan memiliki nilai minimum = 0 dan nilai maksimum = 9. Berdasarkan rumus (1), hasil normalisasi untuk jalan Ring Road Utara yang memiliki jumlah kecelakaan = 7 adalah:

$$N_{11} = \left( \frac{7 - 0}{9 - 0} \right) = 0.78$$

### 3.2 Tahap Data Mining

Dalam penelitian ini digunakan teknik klusterisasi dengan algoritma K-Means yang bertujuan mengelompokkan data berdasarkan kemiripan karakteristik yang dimiliki oleh data tersebut. *K-Means* merupakan salah satu metode data *klustering partitionial* (non hirarki) yang mempartisi data ke dalam klaster/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan dalam satu klaster. Selain itu, K-Means juga termasuk sebagai teknik klusterisasi yang terawasi sehingga pada awal komputasi telah ditetapkan terlebih dahulu jumlah klaster yang diinginkan (Larose D. , 2005).

Algoritma K-Means secara umum memiliki tahapan sebagai berikut (Han & Kamber, 2005):

1. Pada awal iterasi, ada pemilihan pusat klaster secara bebas. Dalam hal ini dapat dipilih secara random sejumlah anggota dari dataset sesuai dengan jumlah klaster (k) yang telah ditetapkan.
2. Menghitung jarak setiap data dengan setiap pusat klaster. Untuk melakukan penghitungan jarak dapat digunakan formula Euclidean (Larose D. , 2005) seperti tampak pada rumus (2):

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2} \quad (2)$$

Dengan:

$d_{ik}$  = jarak antara data ke-i dengan titik pusat klaster ke-k

m = jumlah atribut

$x_j$  = data ke-i

$c_k$  = data pusat klaster ke-k.

3. Suatu data akan menjadi anggota suatu klaster dengan ketentuan memiliki jarak paling minimum diantara semua klaster yang ada.
4. Mengelompokkan data yang menjadi anggota pada setiap klaster.
5. Memperbaharui nilai pusat klaster yang dapat dihitung dengan cara mencari nilai rata-rata sesuai jumlah anggota masing-masing klaster sesuai dengan rumus (3):

$$c_{kj} = \frac{\sum_{h=1}^p y_{hj}}{p}; y_{hj} = x_{ij} \in \text{cluster ke-k} \quad (3)$$

Dengan:

$C_{kj}$  = Pusat *Cluster*

p = Jumlah semua anggota *cluster*

h = Jumlah awal anggota *cluster*

$Y_{hj}$  = Jumlah data

6. Perulangan langkah 2-5 hingga sudah tidak ada lagi data yang berpindah ke klaster yang lain.

Adapun model klusterisasi data sesuai dengan algoritma K-Means untuk membagi data dalam tiga klaster adalah sebagai berikut:

1. Menentukan pusat kluster awal secara random/acak. Misal terpilih 3 obyek seperti tampak pada Tabel 3:
  - a. Pusat kluster pertama (C1) adalah Ring Road Utara dengan data yang telah dinormalisasi (0.78 ; 0.67 ; 1.00).
  - b. Pusat kluster kedua (C2) adalah Jalan Magelang km 5-10 dengan data yang telah dinormalisasi (1.00 ; 1.00 ; 0.93).
  - c. Pusat kluster ketiga (C3) adalah Jalan Wates km 5-10 dengan data yang telah dinormalisasi (0.67 ; 0.62 ; 1.00).
2. Menghitung jarak setiap data (jalan) terhadap setiap pusat kluster menggunakan perhitungan jarak Euclidean seperti terlihat pada rumus (2).  
 Contoh untuk menghitung jarak data pertama (Ring Road Utara) dengan setiap pusat kluster:
  - a. Jarak data pertama dengan C1:  

$$d_{11} = \frac{0.00}{\sqrt{(0.78 - 0.78)^2 + (0.67 - 0.67)^2 + (1.00 - 1.00)^2}}$$
  - b. Jarak data pertama dengan C2:  

$$d_{12} = \frac{0.41}{\sqrt{(0.78 - 1.00)^2 + (0.67 - 1.00)^2 + (1.00 - 0.93)^2}}$$
  - c. Jarak data pertama dengan C3:  

$$d_{13} = \frac{0.12}{\sqrt{(0.78 - 0.67)^2 + (0.67 - 0.62)^2 + (1.00 - 1.00)^2}}$$
3. Menentukan anggota kluster. Suatu data akan menjadi anggota dari suatu kluster yang memiliki jarak terpendek/terkecil terhadap pusat klasternya. Contoh untuk data pertama, jarak terkecil diperoleh pada C3, sehingga data pertama menjadi anggota C3. Perhitungan jarak setiap data terhadap setiap kluster dapat dilihat pada Tabel 4.
4. Mengelompokkan data per kluster. Berdasarkan hasil di Tabel 4 dapat dilihat bahwa:
  - a. C1 terdiri atas 1 anggota: data ke-1.
  - b. C2 terdiri atas 1 anggota: data ke-3
  - c. C3 terdiri atas 8 anggota: data ke-2, 4, 5, 6, 7, 8, 9, dan 10.
5. Memperbaharui pusat kluster yang baru sesuai dengan jumlah anggota tiap kluster

seperti tampak pada Tabel 5. Sebagai contoh adalah perhitungan nilai pusat kluster C3 sesuai Rumus (3):

- a. Pusat kluster untuk jumlah kecelakaan:  

$$C_{31} = \frac{(0+0.33+0.56+0.44+0.33+0.67+0.00+0.44)}{8} = 0.35$$
- b. Pusat kluster untuk jumlah kendaraan  

$$C_{32} = \frac{(0.00+0.29+0.48+0.38+0.29+0.62+0.00+0.43)}{8} = 0.31$$
- c. Pusat kluster untuk jumlah korban:  

$$C_{33} = \frac{(0.00+0.64+0.57+0.36+0.50+1.00+0.00+0.36)}{8} = 0.43$$

6. Iterasi akan berhenti ketika posisi data sudah tidak mengalami perubahan. Untuk data kecelakaan yang terjadi di 10 segmen jalan dibutuhkan tiga kali iterasi untuk mendapatkan hasil klusterisasi akhir. Hasil ini dapat dilihat pada Tabel 6. Berdasarkan hasil akhir ini, maka nilai rata-rata setiap kluster dapat dijadikan sebagai informasi untuk mengukur nilai kemiripan (karakteristik) setiap data yang berkumpul dalam satu kluster seperti tampak pada Tabel 7.

### 3.3 Tahap Pos-Prosesing

Setelah proses klusterisasi telah selesai dilaksanakan maka masuk pada tahap ketiga pos-prosesing data berupa analisis dan visualisasi hasil. Linoff dan Berry (2011) menegaskan bahwa hasil data mining tidak dapat diserahkan sepenuhnya kepada mesin komputer, tetap diperlukan campur tangan manusia untuk menganalisis hasilnya. Dengan kata lain dibutuhkan seorang pakar yang mengenal baik dataset yang di-mining, serta memahami statistik dan struktur model matematika yang mendasari kerja perangkat lunak (Linoff & Berry, 2011).

Dalam penelitian ini, dibutuhkan peran seorang pakar yang memahami dengan baik karakteristik data kecelakaan lalu lintas untuk membandingkan hasil klustering dan menentukan status dari objek-objek jalan di Kabupaten Sleman, DIY.

Informasi nilai rata-rata tiap klaster, seperti tampak pada Tabel 7, dapat digunakan pakar untuk menganalisis dan kemudian mengklasifikasikan status atau tingkat kerawanan jalan raya. Sebagai contoh dapat dilihat pada Tabel 7, jika pusat klaster (nilai rata-rata) untuk semua parameter dari ketiga klaster dibandingkan, tampak secara umum nilai rata-rata C2 memiliki nilai paling besar untuk ketiga parameter. Hal ini berarti obyek-obyek jalan yang menjadi anggota pada klaster C2 memiliki kesamaan karakteristik pada jumlah kecelakaan, jumlah korban, dan jumlah kendaraan yang terlibat paling banyak diantara dua klaster lainnya. Berdasarkan hal ini, pakar dapat menentukan status klaster, misal C2 termasuk sebagai jalan dengan status Sangat Rawan.

#### **4. PERANCANGAN SISTEM**

Sistem pemetaan klusterisasi data kecelakaan lalu lintas dirancang berdasarkan analisis terhadap kebutuhan sistem dan perancangan basis data yang melekat di dalam sistem.

##### **4.1 Analisis Kebutuhan Sistem**

Analisis kebutuhan sistem dilakukan untuk mengetahui kebutuhan data input, kebutuhan proses, dan definisi keluaran sistem. Di awal perancangan didefinisikan terlebih dahulu jenis pengguna sistem. Berdasarkan hasil wawancara dan observasi, sistem untuk mengklusterisasi data kecelakaan lalu lintas ditujukan untuk 4 jenis pengguna, yaitu:

- a. Admin: merupakan user dengan hak akses tertinggi untuk mengelola master data, menentukan hak akses pengguna, dan aktivasi data yang telah diinputkan oleh jenis pengguna lainnya.
- b. Operator: merupakan user yang berperan untuk mengelola data rutin sistem, yaitu data kecelakaan lalu lintas berdasarkan waktu, lokasi, jumlah korban, dan jumlah kendaraan yang terlibat.
- c. Pakar: merupakan user yang berperan untuk menjalankan fungsi klusterisasi

data, memberikan analisis hasil klustering, dan menetapkan kategori klaster. Terdapat tiga jenis kategori klaster yang berkaitan dengan tingkat kerawanan jalan, yaitu Tidak Rawan (Aman), Rawan, dan Sangat Rawan.

- d. Pengunjung: merupakan user umum yang dapat melihat hasil klusterisasi data beserta kategori tingkat kerawanan jalan.

Adapun kebutuhan data input sistem terdiri atas:

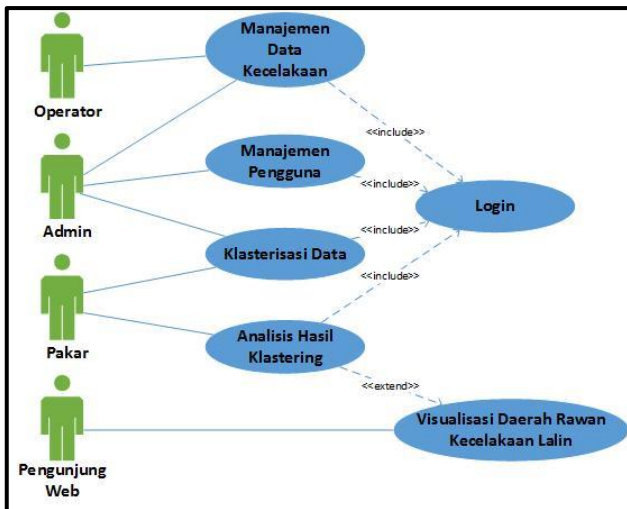
- a. Data kecelakaan lalu lintas yang meliputi: waktu, lokasi, jumlah kendaraan yang terlibat, dan jumlah korban.
- b. Data spasial berupa batas administrasi wilayah Kabupaten Sleman dan data jalan utama di Kabupaten Sleman.
- c. Data pelengkap: tahun, username, dan password.

Keluaran sistem yang diharapkan adalah peta klusterisasi atau pengelompokkan jalan yang memiliki karakteristik kecelakaan yang sama. Untuk mencapai keluaran tersebut dibutuhkan sejumlah proses pengolahan data sebagai berikut:

- a. Manajemen Data Kecelakaan Lalu Lintas, adalah fungsi yang digunakan untuk mengelola data kecelakaan lalu lintas berdasarkan waktu, lokasi, jumlah kendaraan yang terlibat, dan jumlah korban.
- b. Manajemen Pengguna, adalah fungsi untuk mengelola akun dan hak akses pengguna sistem.
- c. Klusterisasi Data, adalah fungsi untuk mengelompokkan data kecelakaan per bulan menggunakan algoritma K-Means.
- d. Analisis Hasil Klustering, adalah fungsi yang dapat dijalankan oleh user Pakar untuk memasukkan hasil analisisnya dan menentukan kategori tingkat kerawanan jalan.
- e. Visualisasi Hasil, adalah fungsi penghubung antara data spasial jalan dengan hasil klusterisasi data yang divisualisasikan dalam bentuk peta.



Gambaran interaksi antara pengguna terhadap sistem divisualisasikan dalam bentuk Use Case Diagram seperti tampak pada Gambar 3. User Admin memiliki hak penuh untuk mengakses seluruh fungsi. User Operator memiliki hak akses pada fungsi manajemen data kecelakaan, yaitu mengelola data mentah seperti yang didefinisikan dalam tahap pra-prosesing. User Pakar memiliki hak akses pada fungsi Klasterisasi Data dan Analisis Hasil Klastering. Dalam hal ini, fungsi Analisis Klastering hanya dapat dijalankan jika Pakar telah menjalankan fungsi Klasterisasi Data untuk waktu (bulan dan tahun) yang sama. Output dari fungsi Analisis Hasil Klastering inilah yang akan tampil di peta dalam bentuk persebaran jalan dengan status tingkat kerawanan kecelakaan lalu lintas.



Gambar 3. Perancangan Use Case Diagram

#### 4.2 Perancangan Basis Data

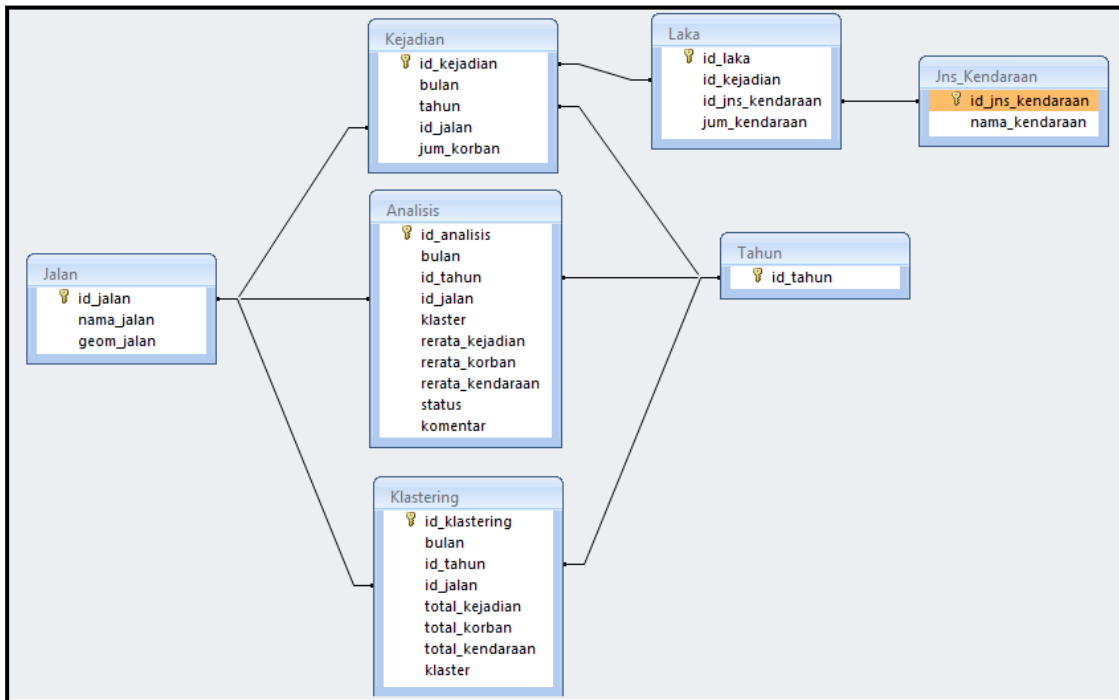
Dalam pengelolaan data dibutuhkan sejumlah tabel yang berfungsi untuk repositori data. Berikut adalah penjelasan singkat tabel-tabel yang dibutuhkan:

- a. Tabel Login untuk menyimpan data akun, password, dan jenis user.
- b. Tabel Jalan untuk menyimpan data spasial jalan, yaitu: ID Jalan, nama jalan, dan geometri jalan.

- c. Tabel Tahun untuk menyimpan nama tahun. Hal ini dibutuhkan sehingga data dapat disimpan untuk berbagai tahun di dalam sistem.
- d. Tabel Kejadian untuk menyimpan data kecelakaan lalu lintas, berupa: waktu kejadian (bulan dan tahun), lokasi kejadian (ID jalan), dan jumlah korban. Tabel ini memiliki relasi terhadap tabel Jalan dan Tabel Tahun.
- e. Tabel Laka berfungsi sebagai penjelas kejadian kecelakaan lalu lintas, yaitu untuk menyimpan informasi jenis-jenis kendaraan dan jumlah kendaraan yang terlibat dalam kecelakaan lalu lintas. Di sistem ini dikenali beberapa jenis kendaraan, seperti: motor, mobil, sepeda, truk, dan bus. Tabel ini memiliki relasi terhadap tabel Kejadian dan Tabel Jenis Kendaraan.
- f. Tabel Jenis Kendaraan berfungsi untuk menyimpan nama jenis kendaraan. Hal ini dibutuhkan untuk mengakomodasi jika jenis kendaraan bertambah.
- g. Tabel Klasterisasi berfungsi untuk menyimpan rekapitulasi kejadian kecelakaan, total korban, dan total kendaraan yang terlibat kecelakaan pada suatu jalan, bulan, dan tahun. Selain itu, tabel ini juga menyimpan hasil perhitungan K-Means berupa posisi kluster terakhir.
- h. Tabel Analisis berfungsi untuk menyimpan analisis dari user Pakar yang berupa penetapan status untuk setiap kluster di suatu bulan dan tahun. Di tabel ini juga menyimpan nilai pusat tiap kluster (rerata setiap parameter klastering).

Rancangan basis data beserta daftar atribut dan relasi yang terbentuk dapat dilihat pada Gambar 4.





Gambar 4. Relasi Antar Tabel

## 5. IMPLEMENTASI RANCANGAN

Berdasarkan hasil perancangan, maka secara umum pemetaan klasterisasi data kecelakaan lalu lintas ini memiliki memiliki tiga antarmuka utama, sebagai berikut:

### a. Halaman Manajemen Data Kecelakaan

Pada halaman manajemen data kecelakaan, user Operator dapat memasukkan data nama bulan, tahun, nama jalan, jumlah korban, dan menentukan jenis dan jumlah kendaraan yang terlibat kecelakaan. Output dari aktivitas ini akan masuk ke halaman List Data yang berisi rekapitulasi data kecelakaan pada bulan, tahun, dan jalan tertentu seperti tampak di Gambar 5.

### b. Halaman Klasterisasi Data

Pada halaman klasterisasi data, user Pakar dapat menjalankan proses klasterisasi dengan cara memilih data bulan dan tahun tertentu seperti tampak pada Gambar 6.

Pada halaman tersebut tersedia 2 tombol, yaitu: Cari dan Simpan. Fungsi tombol Cari adalah untuk menjalankan proses klasterisasi data berdasarkan algoritma K-Means sehingga akan tampil hasil klasterisasi berupa nama jalan dan jenis klasternya. Setiap objek jalan hanya melekat pada satu klaster yang ditandai dengan angka 1. Sedangkan tombol Simpan berfungsi untuk menyimpan data ke Tabel Klasterisasi.

**Tambah Data Kejadian**

Bulan : januari

Tahun : 2011

Nama jalan : Afandi

Jumlah Korban :

Laka	Jenis Kendaraan	Jumlah kendaraan
	motor	<input type="text"/>
	mobil	<input type="text"/>
	sepeda	<input type="text"/>
	truk	<input type="text"/>
	bus	<input type="text"/>

Gambar 5. Antarmuka Tambah Data

Pilih Bulan :

Pilih Tahun :

nama_jalan	c1	c2	c3
Ring Road Utara	1	0	0
Ring Road Selatan	0	1	0
Magelang KM 0-5	0	1	0
Magelang KM 5,1-10	1	0	0
Magelang KM 10,1-15	0	0	1
Magelang KM 15,1-20	0	0	1
Kaliurang KM 0-5	0	1	0
Kaliurang KM 5,1-10	1	0	0
Kaliurang KM 10,1-15	0	0	1
Jogja-Solo KM 0-5	0	1	0
Jogja-Solo KM 5,1-10	0	0	1
Jogja-Solo KM 10,1-15	1	0	0
Jogja-Solo KM 15,1-20	0	1	0
Laksda Adi Sucipto KM 0-5	0	0	1

Gambar 6. Antarmuka Klasterisasi Data

**c. Halaman Analisis Hasil Klustering**

Setelah proses klustering dijalankan dan disimpan, selanjutnya user Pakar dapat memasukkan analisisnya terkait hasil klustering pada halaman analisis hasil klustering seperti tampak pada Gambar 7. Pada halaman tersebut Pakar dapat melihat informasi nilai minimum, nilai maksimum, dan pusat klaster (nilai rata-rata klaster) untuk setiap indikator.

Berdasarkan informasi pusat klaster inilah seorang Pakar diharapkan dapat menentukan jenis status klaster dan memasukkan komentarnya. Adapun output dari hasil analisis ini akan tampil sebagai peta yang menunjukkan pengelompokkan jalan beserta dengan status tingkat kerawanan seperti tampak di Gambar 8.

Status klaster terbagi 3, yaitu: Tidak Rawan, Rawan, dan Sangat Rawan. Penentuan jenis status akan mempengaruhi warna objek jalan, dalam hal ini: status Tidak Rawan berasosiasi dengan warna hijau, status Rawan berasosiasi dengan warna kuning, dan status Sangat Rawan berasosiasi dengan warna Merah

Cluster	Jumlah Korban	Jumlah Kendaraan	Kejadian	Status	Komentar
c1	Min : 8 Max : 15 P Cluster : 11,14	Min : 16 Max : 33 P Cluster : 23,71	Min : 11 Max : 26 P Cluster : 19,29	Rawan	jalan ring road utara sangat rawan terjadi kecelakaan
c2	Min : 0 Max : 2 P Cluster : 0,77	Min : 0 Max : 4 P Cluster : 1,46	Min : 0 Max : 4 P Cluster : 1,31	Tidak Rawan	rawan
c3	Min : 3 Max : 6 P Cluster : 4,56	Min : 6 Max : 12 P Cluster : 9,44	Min : 5 Max : 14 P Cluster : 8,33	Sangat Rawan	sangat rawaan

Gambar 7. Antarmuka Untuk Analisis Hasil Klustering



Gambar 8. Visualisasi Hasil Klustering

**6. KESIMPULAN DAN SARAN**

**6.1 Kesimpulan**

Berdasarkan pemetaan hasil klasterisasi data kecelakaan lalu lintas yang telah dilakukan maka terdapat beberapa hal yang dapat disimpulkan, sebagai berikut:

- Sistem klasterisasi data kecelakaan lalu lintas dengan metode K-Means dapat digunakan untuk mengelompokkan obyek jalan berdasarkan kesamaan karakteristik pada jumlah korban, jumlah kendaraan yang terlibat, dan jumlah kecelakaan yang terjadi dalam suatu rentang waktu tertentu.
- Sistem klasterisasi data kecelakaan lalu lintas membutuhkan peran seorang pakar lalu lintas untuk menganalisis hasil klustering dan menentukan klasifikasi berupa status tingkat kerawanan jalan raya.
- Visualisasi hasil klustering sebagai sebuah peta merupakan media yang sangat efektif untuk menunjukkan jalan-

jalan yang memiliki tingkat kerawanan yang sama.

## 6.2 Saran

Sistem pemetaan hasil klasterisasi data kecelakaan lalu lintas ini masih memiliki sejumlah kekurang, sehingga mengacu dari kekurangan inilah maka terdapat sejumlah saran untuk pengembangan sistem di penelitian selanjutnya. Adapun saran yang diajukan adalah sebagai berikut:

- a. Perlunya integrasi terhadap sistem perekapan kejadian kecelakaan yang telah digunakan secara rutin oleh pihak Ditlantas. Berkaitan dengan hal ini, maka sangat disarankan untuk mengembangkan sistem *retrieval* yang mampu mendeteksi kata kunci yang dibutuhkan dari deskripsi kejadian perkara kecelakaan untuk menjadi input parameter klustering.
- b. Data temporal yang digunakan dalam penelitian ini masih terbatas pada periode bulanan, sehingga terdapat peluang untuk mengembangkan sistem agar mampu menggabungkan (agregasi) data dalam periode yang dinamis, misal: harian, mingguan, bulanan, tahunan, bahkan mampu mengolah waktu kejadian yang lebih spesifik, misal: pagi, siang, sore, atau malam hari.
- c. Perlunya mengembangkan fungsi pembandingan data antar waktu sehingga dapat diketahui tren kejadian kecelakaan dalam berbagai satuan waktu seperti tercantum pada poin b.
- d. Parameter klasterisasi yang digunakan dalam penelitian ini masih bersumber pada data deskripsi kejadian, belum melibatkan data geometri (kondisi fisik) jalan raya. Sehingga disarankan untuk turut mempertimbangkan data geometri jalan, berupa (Depkimpraswil, 2004) tingkat kelandaian, kinerja jalan, ataupun tingkat hambatan sampling.

## DAFTAR PUSTAKA

- Bolla, M. E., Messah, Y., & Koreh, M. B. (2013). Analisis Daerah Rawan Kecelakaan Lalu Lintas (Studi Kasus Ruas Jalan Timor Raya Kota Kupang). *Jurnal Teknik Sipil Vol. II, No. 2*, 147-156.
- BPS. (2014). *Angka Kecelakaan Lalu Lintas*. Retrieved Oktober 1, 2014, from Badan Pusat Statistik: [http://bps.go.id/tab\\_sub/view.php?kat=2&tabel=1&daftar=1&id\\_subyek=17&notab=14](http://bps.go.id/tab_sub/view.php?kat=2&tabel=1&daftar=1&id_subyek=17&notab=14)
- Chen, Y., Chun, L., Wu, H., & Sun, W. (2011). Identification of Black Spot on Traffic Accidents and its Spatial Association Analysis Based on Geographic Information System. *Seventh International Conference on Natural Computation* (pp. 143-150). IEEE.
- Depkimpraswil. (2004). *Penanganan Lokasi Rawan Kecelakaan Lalu Lintas*. Jakarta: Departemen Permukiman dan Prasarana Wilayah.
- Fayyad, U. P.-S. (1996). *Advances in Knowledge Discovery and Data Mining*. MIT Press.
- Han, J., & Kamber, M. (2005). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.
- Harahap, C. S. (2013). *Perancangan Aplikasi Prediksi Kecelakaan Lalu Lintas di Percut Sei Tuan Menggunakan Algoritma Apriori*. Medan: Pelita Informatika Budi Darma, Volume : V, Nomor: 3.
- Kraak, M. J., & Ormeling, F. (2010). *Cartography: Visualization of Geospatial Data, 3rd Ed*. The Guilford Press.
- Larose, D. T. (2005). *Discovering Knowledge in Data*. New Jersey: John Wiley & Sons, Inc.
- Li, A. Z., & Song, X. H. (2012). Traffic Accident Characteristics Analysis Based on Fuzzy Clustering. *IEEE Symposium on Electrical & Electronics Engineering* (pp. 468-470). IEEE.
- Linoff, G. S., & Berry, M. J. (2011). *Data Mining Technique: For Marketing, Sales, and Customer Relationship Management, 3rd Ed*. Wiley and Sons.

Ransi, N. (2014). *Pengaplikasian Algoritma Classification Based On Predictive Association Rules Untuk Analisa Karakteristik Kecelakaan Lalu Lintas*. Yogyakarta: Program Magister Ilmu Komputer UGM.

Wedasana, A. S. (2011). *Analisis Daerah Rawan Kecelakaan dan Penyusunan Database Berbasis Sistem Informasi Geografis (Studi Kasus Kota Denpasar)*.

Denpasar: Program Magister Teknik Sipil Universitas Udayana.

Yunanto, W., Hariadi, M., & Purnomo, M. H. (2012). *Visualisasi Cerdas Informasi Geospasial dan Timeline Hybrid Berbasis Oracle Database Menggunakan Naive Bayes*. Surabaya: Program Magister Bidang Keahlian Telematika ITS.

### LAMPIRAN

Tabel 1. Hasil Simplikasi dan Generalisasi Data Kecelakaan Lalu Lintas

No	Nama Jalan	Jumlah Kecelakaan	Jumlah Kendaraan	Jumlah Korban
1	Ring Road Utara	7	14	14
2	Ring Road Selatan	0	0	0
3	Magelang km 5-10	9	21	13
4	Kaliurang km 10-15	3	6	9
5	Jogja-Solo km 10-15	5	10	8
6	Adi Sucipto km 5-10	4	8	5
7	Godean km 5-10	3	6	7
8	Wates km 5-10	6	13	14
9	Palagan	0	0	0
10	Affandi	4	9	5

Tabel 2. Hasil Normalisasi Data

No	Nama Jalan	Jumlah Kecelakaan	Jumlah Kendaraan	Jumlah Korban
1	Ring Road Utara	0.78	0.67	1.00
2	Ring Road Selatan	0.00	0.00	0.00
3	Magelang km 5-10	1.00	1.00	0.93
4	Kaliurang km 10-15	0.33	0.29	0.64
5	Jogja-Solo km 10-15	0.56	0.48	0.57
6	Adi Sucipto km 5-10	0.44	0.38	0.36
7	Godean km 5-10	0.33	0.29	0.50
8	Wates km 5-10	0.67	0.62	1.00
9	Palagan	0.00	0.00	0.00
10	Affandi	0.44	0.43	0.36

Tabel 3. Titik Pusat Klaster Untuk Iterasi Pertama

Klaster	Nama Jalan	Jumlah Kecelakaan	Jumlah Kendaraan	Jumlah Korban
C1	Ring Road Utara	0.78	0.67	1.00
C2	Magelang km 5-10	1.00	1.00	0.93
C3	Wates km 5-10	0.67	0.62	1.00

Tabel 4. Perhitungan Jarak & Penentuan Anggota Klaster

No	Nama Jalan	C1	C2	C3	Klaster
1	Ring Road Utara	0.00	0.41	0.12	C1
2	Ring Road Selatan	1.43	1.69	1.35	C3
3	Magelang km 5-10	0.41	0.00	0.51	C2
4	Kaliurang km 10-15	0.69	1.02	0.59	C3
5	Jogja-Solo km 10-15	0.52	0.77	0.47	C3
6	Adi Sucipto km 5-10	0.78	1.01	0.72	C3
7	Godean km 5-10	0.77	1.07	0.69	C3
8	Wates km 5-10	0.12	0.51	0.00	C3
9	Palagan	1.43	1.69	1.35	C3
10	Affandi	0.76	0.98	0.71	C3

Tabel 5. Pusat Klaster Hasil Baru

Klaster	Jumlah Kecelakaan	Jumlah Kendaraan	Jumlah Korban
C1	0.78	0.67	1.00
C2	1.00	1.00	0.93
C3	0.35	0.31	0.43

Tabel 6. Informasi Pusat Klaster Akhir

Klaster	Jumlah Kecelakaan	Jumlah Kendaraan	Jumlah Korban	Anggota
C1	0.72	0.64	1.00	Data ke-1, 8
C2	1.00	1.00	0.93	Data ke-3
C3	0.30	0.27	0.35	Data ke-2, 4, 5, 6, 7, 9, 10

Tabel 7. Perbandingan Nilai dan Penentuan Status Klaster

Klaster	Jumlah Kecelakaan	Jumlah Kendaraan	Jumlah Korban	Status	Asosiasi Warna Obyek
C1	Min: 0.67 Maks: 0.78 Rerata: 0.72	Min: 0.62 Maks: 0.67 Rerata: 0.64	Min: 1.00 Maks: 1.00 Rerata: 1.00	Rawan	Kuning
C2	Min: 1.00 Maks: 1.00 Rerata: 1.00	Min: 1.00 Maks: 1.00 Rerata: 1.00	Min: 0.93 Maks: 0.93 Rerata: 0.93	Sangat Rawan	Merah
C3	Min: 0.00 Maks: 0.56 Rerata: 0.30	Min: 0.00 Maks: 0.48 Rerata: 0.27	Min: 0.00 Maks: 0.64 Rerata: 0.35	Tidak Rawan	Hijau