

## ***K-MEANS CLUSTERING UNTUK DATA KECELAKAAN LALU LINTAS JALAN RAYA DI KECAMATAN PELAIHARI***

Winda Aprianti<sup>1</sup>, Jaka Permadi<sup>2</sup>

<sup>1,2</sup>Jurusan Teknik Informatika, Politeknik Negeri Tanah Laut  
Email: <sup>1</sup>winda@politala.ac.id, <sup>2</sup>jakapermadi.88@politala.ac.id

(Naskah masuk: 10 Oktober 2018, diterima untuk diterbitkan: 24 Oktober 2018)

### **Abstrak**

Kecelakaan lalu lintas di jalan raya masih menjadi penyumbang tingginya angka kematian di Indonesia, sehingga menjadi perhatian khusus bagi kepolisian di negara ini. Termasuk Kepolisian Resor (Polres) Tanah Laut, yang telah membuktikan perhatian tersebut dengan membentuk komunitas korban kecelakaan lalu lintas dan Pelatihan Pertolongan Pertama Gawat Darurat (PPGD). Tahapan awal pencegahan kecelakaan lalu lintas adalah dengan mengetahui faktor-faktor penyebab kecelakaan lalu lintas yang diperoleh melalui analisa data kecelakaan. Analisa tersebut dapat dilakukan dengan data mining, yaitu *K-Means Clustering*. *K-Means Clustering* mengelompokkan data menjadi beberapa *cluster* sesuai karakteristik data tersebut. Data kecelakaan lalu lintas dibagi menjadi 2 dataset, yakni dataset 1 dan dataset 2. Hasil *cluster* penerapan *K-means clustering* terhadap dataset 1 dan dataset 2 kemudian dilakukan pengujian *silhouette coefficient* untuk mencari hasil *cluster* dengan kualitas terbaik. Pengujian *silhouette coefficient* secara berurutan menghasilkan *distance measure* paling optimal yakni *clustering* dengan 4 *cluster* untuk dataset 1 dan *clustering* dengan 2 *cluster* untuk dataset 2. Selain memperoleh *cluster* dengan kualitas terbaik, penganalisaan data juga menghasilkan beberapa informasi kecelakaan lalu lintas yang sering terjadi, yakni faktor penyebab dan korban kecelakaan adalah pengemudi, umur korban adalah 9 sampai 28 tahun, dan keadaan korban kecelakaan adalah luka ringan.

**Kata kunci:** *k-means, clustering, kecelakaan, lalu lintas, silhouette coefficient*

## ***K-MEANS CLUSTERING FOR HIGHWAY TRAFFIC ACCIDENT DATA IN PELAIHARI SUB DISTRICT***

### **Abstract**

Traffic accidents on the highway are still contribute to the high mortality rate in Indonesia, which are becoming a special concern for the police. Including the Police of Tanah Laut Resort where prove themselves by established The Community of Traffic Accident Victims and Emergency First Aid Training. The first prevention of traffic accidents is knowing the factors causing traffic accidents which is obtained through the analysis of traffic accident's data. It can be done through data mining, i.e. *K-Means Clustering*, which is clustering data into clusters according to characteristics of the data. Traffic accident data is divided into two datasets, namely dataset 1 and dataset 2. After obtaining the cluster results, the next step is to calculate *silhouette coefficient* which is used to find the best quality cluster result. The result of testing *silhouette coefficient* are clustering with 4 clusters for dataset 1 and clustering with 2 clusters for dataset 2. Analyzing data in this research also produces some information on traffic accidents that often occur, namely the causes and victims of accidents are drivers, the age of the victims is between 9 and 28 years old, and the circumstance of the accidents victims are minor injuries.

**Keywords:** *k-means, clustering, accident, traffic, silhouette coefficient*

### **1. PENDAHULUAN**

Kecelakaan lalu lintas di jalan raya masih menjadi penyumbang angka kematian yang relatif tinggi di Indonesia. Hal ini didukung dengan data yang dikeluarkan oleh Korps Lalu Lintas Kepolisian Republik Indonesia (Korlantas Polri), trend kecelakaan lalu lintas di Indonesia mengalami

penurunan menjadi 95.906 kasus pada tahun 2014 dan kembali mengalami peningkatan menjadi 105.374 kasus dengan 25.859 korban meninggal pada tahun 2016 (Maulana, 2017). Kecelakaan lalu lintas juga menjadi permasalahan di Provinsi Kalimantan Selatan. Data yang bersumber dari PT Jasa Raharja Cabang Kalimantan Selatan

menyatakan bahwa kecelakaan yang terjadi pada tahun 2014 berjumlah 963 kasus. Pada tahun 2016 periode Januari sampai dengan Agustus telah terjadi 7 kasus kecelakaan lalu lintas di Kabupaten Tanah Laut (Wahid, 2016). Kecelakaan lalu lintas merupakan hal yang menjadi perhatian khusus oleh Kepolisian Resor (Polres) Tanah Laut dibuktikan dengan dibentuknya komunitas korban kecelakaan lalu lintas dan pelatihan pertolongan pertama gawat darurat (PPGD) sebagai salah satu upaya dalam mengurangi angka kecelakaan dan fatalitas korban (Humas Polres Tanah Laut, 2017).

Berdasarkan data yang telah diuraikan maka perlu analisis lebih lanjut terhadap permasalahan kecelakaan lalu lintas untuk mengetahui faktor-faktor penyebab kecelakaan lalu lintas. Analisa terhadap data kecelakaan dapat dilakukan menggunakan data mining, yakni *K-Means clustering* yang dapat mengklasifikasikan data menjadi beberapa *cluster* sesuai karakteristik data. Penerapan *K-Means clustering* untuk analisa daa kecelakaan telah dilakukan oleh Iswari (2015), Fajar (2015), serta Rahmat dkk (2017). Iswari (2015) memanfaatkan algoritma *K-Means* untuk pemetaan daerah rawan kecelakaan di Sleman, Daerah Istimewa Yogyakarta. Rahmat dkk (2017) mengimplementasikan *K-Means Clustering* untuk menganalisis frekuensi tingkat kecelakaan di tiap lokasi yang berpotensi terjadi kasus kecelakaan di Kota Kendari. Iswari dan Rahmat mengklasifikasikan data kecelakaan lalu lintas menjadi beberapa cluster berdasarkan daerah rawan kecelakaan. Fajar (2015) dalam penelitiannya juga menggunakan *K-Means* untuk mengklasifikasikan data kecelakaan di Semarang menjadi beberapa cluster kategori tingkat kecelakaan berdasarkan umur korban kecelakaan.

Penerapan *K-Means Clustering* pada data kecelakaan lalu lintas pada penelitian ini akan menyediakan fitur penginputan data kecelakaan, pemilihan kategori klasifikasi, penentuan *centroid* awal, pemrosesan *K-Means Clustering*, dan menampilkan *cluster* yang dihasilkan. Kemudian akan dilakukan analisa terhadap *cluster* yang dihasilkan untuk membantu pengurangan resiko terjadinya kecelakaan di Kecamatan Pelaihari Kabupaten Tanah Laut.

Tujuan yang akan dicapai pada penelitian ini adalah mengetahui faktor penyebab, lokasi yang menjadi daerah rawan kecelakaan, korban, usia dan keadaan korban yang paling sering mengalami kecelakaan lalu lintas, serta menerapkan *K-Means Clustering* pada data kecelakaan lalu lintas sehingga dapat memprediksi waktu terjadinya kecelakaan lalu lintas di Kecamatan Pelaihari Kabupaten Tanah Laut.

## 2. METODE PENELITIAN

Tahap pertama pada penelitian ini adalah pengumpulan data, yakni data sekunder dari Polisi

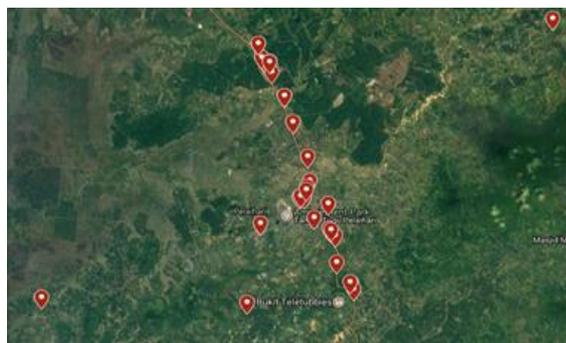
Resort Kabupaten Tanah Laut pada Tahun 2016 dan 2017 Adapun data yang dijadikan objek penelitian ada 9 atribut, yakni lokasi kecelakaan; usia korban kecelakaan; jenis kendaraan meliputi kendaraan roda 2, kendaraan roda 4, dan kendaraan roda lebih dari 4; penyebab terjadinya kecelakaan meliputi pengemudi, lingkungan, jalan, kendaraan; keadaan korban kecelakaan meliputi meninggal, luka berat, dan luka ringan; jumlah kendaraan yang terlibat; korban kecelakaan meliputi pengemudi, penumpang, dan pejalan kaki; hari terjadinya kecelakaan meliputi hari kerja, hari libur, dan akhir minggu; serta waktu terjadinya kecelakaan meliputi dini hari, pagi hari, siang hari, sore hari dan malam hari.

Setelah pengumpulan data dilakukan, tahapan selanjutnya adalah *preprocessing data* meliputi *selection data* dan *transformation data*. *Selection data* merupakan proses pemilihan beberapa faktor yang memuat informasi utama untuk dijadikan atribut *clustering*. Sedangkan tranformation data merupakan proses pengubahan data dari bentuk deskriptif ke bentuk tipe data nominal. Data hasil *preprocessing* kemudian dikelompokkan menggunakan *K-Means Clustering* dengan penentuan *centroid awal* menggunakan metode *simple random sampling*. Setelah memperoleh *cluster* dari data kecelakaan, maka akan dihitung *Silhouette Coefficient* untuk mengetahui kualitas dan kekuatan *cluster* (Anggara dkk, 2016 dan Anggodo dkk, 2017). Tahap terakhir adalah melakukan analisa terhadap *cluster* yang dihasilkan.

## 3. HASIL DAN PEMBAHASAN

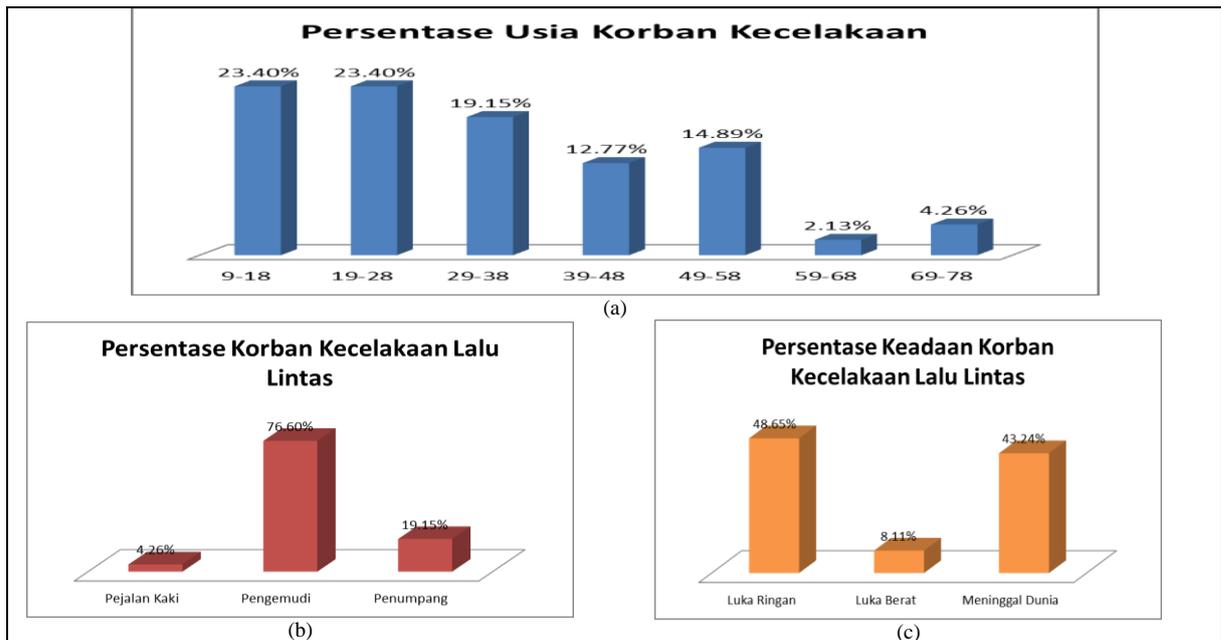
Data kecelakaan yang diperoleh dari Polisi Resort Kabupaten Tanah Laut pada Tahun 2016 dan 2017 berjumlah 32 data dimana setiap data memuat 9 atribut. Berdasarkan data sekunder tersebut diperoleh bahwa penyebab kecelakaan lalu lintas adalah pengemudi. Selain itu, diperoleh juga informasi mengenai korban, usia korban, dan keadaan yang paling sering dialami korban pada kecelakaan lalu lintas di Kecamatan Pelaihari.

Visualisasi data kecelakaan awal disajikan pada Gambar 1.



Gambar 1. Visualisasi Data Kecelakaan Lalu Lintas

Sedangkan informasi mengenai hal yang disebutkan di atas disajikan pada Gambar 2.



Gambar 2. Grafik Persentase: (a) Usia Korban Kecelakaan, (b) Korban Kecelakaan Lalu Lintas, (c) Keadaan Korban Kecelakaan Lalu Lintas

Gambar 2 menunjukkan kecelakaan lalu lintas sering terjadi pada usia kelompok 9-18 tahun dan 19-28 tahun, korban kecelakaan sering kali adalah pengemudi kendaraan, dan keadaan yang dialami korban kecelakaan yang paling sering luka ringan tetapi perbedaan persentase dengan keadaan korban yang meninggal dunia tidak berbeda jauh, yakni hanya berbeda 5.41%.

Guna menggali informasi lebih lanjut, maka diterapkan algoritma *K-Means Clustering* terhadap data yang telah dilakukan *preprocessing*. Tahap awal *preprocessing* adalah *selection data*. Data yang

dikumpulkan menunjukkan faktor penyebab terjadinya kecelakaan hanya disebabkan pengemudi maka faktor ini dihapuskan dari atribut *clustering* maka pada tahapan selection data diperoleh 8 atribut *clustering*. Data ini kemudian ditransformasi ke bentuk nominal sehingga diperoleh dataset 1 yang ditampilkan pada Tabel 1 dan dataset 2 yang ditampilkan pada Tabel 2.

Tabel 1 menyajikan dataset 1 yang memuat data lokasi kejadian dalam koordinat *latitude* dan *longitude*.

Tabel 1. Dataset Awal untuk Lokasi Kejadian

ID	Latitude	Longitude	ID	Latitude	Longitude
lap_01	-3.81597	114.78542	lap_17	-3.71626	114.75082
lap_02	-3.72637	114.75423	lap_18	-3.79004	114.77582
lap_03	-3.73156	114.75729	lap_19	-3.80225	114.78481
lap_04	-3.81925	114.78816	lap_20	-3.79884	114.77066
lap_05	-3.74428	114.76335	lap_21	-3.81646	114.78601
lap_06	-3.84506	114.7962	lap_22	-3.81273	114.75173
lap_07	-3.74445	114.76345	lap_23	-3.77673	114.77485
lap_08	-3.85549	114.74524	lap_24	-3.79298	114.77462
lap_09	-3.72397	114.75256	lap_25	-3.85256	114.64546
lap_10	-3.7988	114.77327	lap_26	-3.79807	114.77154
lap_11	-3.7029	114.89416	lap_27	-3.72608	114.75615
lap_12	-3.81555	114.785	lap_28	-3.83383	114.78884
lap_13	-3.91452	114.63109	lap_29	-3.84828	114.79734
lap_14	-3.72577	114.75623	lap_30	-3.79463	114.77399
lap_15	-3.80972	114.77807	lap_31	-3.84438	114.79537
lap_16	-3.71626	114.75082	lap_32	-3.75838	114.76751

Tabel 2 menyajikan dataset 2 yang memuat 7 atribut, yakni umur korban (A); jumlah kendaraan yang terlibat (B); jenis kendaraan meliputi kendaraan roda 2 (C1), kendaraan roda 4 (C2), dan kendaraan roda lebih dari 4 (C3); korban kecelakaan meliputi pejalan kaki (D1), pengemudi (D2), dan penumpang (D3); hari terjadinya kecelakaan

meliputi hari kerja (E1), hari libur (E2), dan akhir minggu (E3); waktu terjadinya kecelakaan meliputi pagi hari (F1), siang hari (F2), sore hari (F3), malam hari (F4), dan dini hari (F5); serta keadaan korban kecelakaan meliputi luka ringan (G1), luka berat (G2), dan meninggal dunia (G3).

Tabel 2. Dataset Awal untuk 7 Atribut

No	A	B	C1	C2	C3	D1	D2	D3	E1	E2	E3	F1	F2	F3	F4	F5	G1	G2	G3
1	9	1	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	1
2	14	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
3	17	5	5	0	0	0	3	2	0	0	1	1	0	0	0	0	2	2	1
4	18	4	4	0	0	0	3	1	3	0	0	1	0	1	1	0	2	0	2
5	22	2	2	0	0	0	2	0	2	0	0	2	0	0	0	0	2	0	0
6	24	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0
7	25	3	2	0	1	0	2	1	1	0	0	1	0	0	0	0	2	0	1
8	26	2	1	1	0	0	2	0	2	0	0	0	0	2	0	0	2	0	0
9	27	2	0	2	0	0	0	2	0	0	0	0	0	0	0	0	2	0	0
10	28	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
11	29	2	2	0	0	0	2	0	1	0	0	0	0	0	1	0	0	0	2
12	30	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
13	32	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
14	35	1	1	0	0	0	1	0	0	0	1	1	0	0	0	0	1	0	0
15	36	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
16	37	2	2	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	2
17	38	1	0	1	0	0	1	0	1	0	0	1	0	0	0	0	1	0	0
18	41	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
19	43	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
20	45	2	2	0	0	0	1	1	0	0	1	1	0	0	0	0	1	0	1
21	47	1	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
22	48	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
23	49	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
24	50	2	2	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0	2
25	51	2	2	0	0	0	2	0	0	0	0	0	0	0	0	0	0	1	1
26	53	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
27	56	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
28	60	1	1	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	1
29	70	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
30	74	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1

Hasil clustering dataset 1 divisualisasikan pada Gambar 3 sampai dengan Gambar 7.

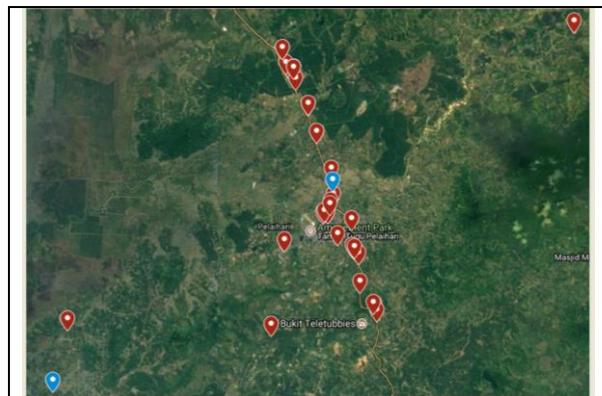
Selanjutnya adalah penerapan *K-Means Clustering* dan perhitungan *silhouette coefficient*.

### 3.1. Hasil *K-Means Clustering* dan *silhouette coefficient* pada dataset lokasi kejadian

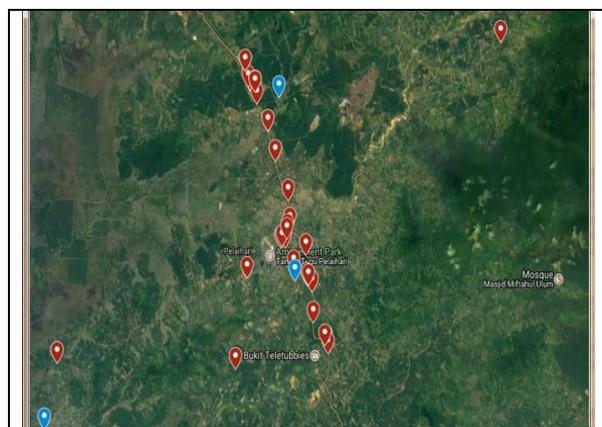
Tabel 3 adalah jumlah anggota hasil clustering dari Tabel 1 secara berurutan untuk jumlah cluster sebanyak 2, 3, 4, 5, dan 6.

Tabel 3. Jumlah Anggota Hasil Clustering Dataset 1

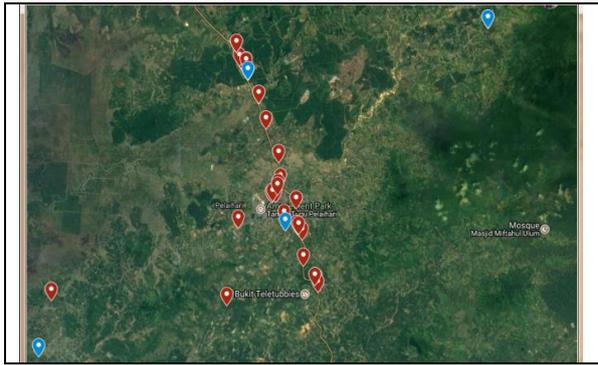
Cluster	Pusat Cluster		Jumlah Anggota
	Lattitude	Longitude	
<b>Jumlah cluster = 2</b>			
1	-3.78284	114.7755	30
2	-3.88354	114.6383	2
<b>Jumlah cluster = 3</b>			
1	-3.88354	114.6383	2
2	-3.72875	114.7697	11
3	-3.81416	114.7788	19
<b>Jumlah cluster = 4</b>			
1	-3.7029	114.8942	1
2	-3.88354	114.6383	2
3	-3.81416	114.7788	19
4	-3.73134	114.7572	10
<b>Jumlah cluster = 5</b>			
1	-3.803	114.7767	14
2	-3.84541	114.7846	5
3	-3.73134	114.7572	10
4	-3.7029	114.8942	1
5	-3.88354	114.6383	2
<b>Jumlah cluster = 6</b>			
1	-3.73134	114.7572	10
2	-3.84541	114.7846	5
3	-3.7029	114.8942	1
4	-3.803	114.7767	14
5	-3.85256	114.6455	1
6	-3.91452	114.6311	1



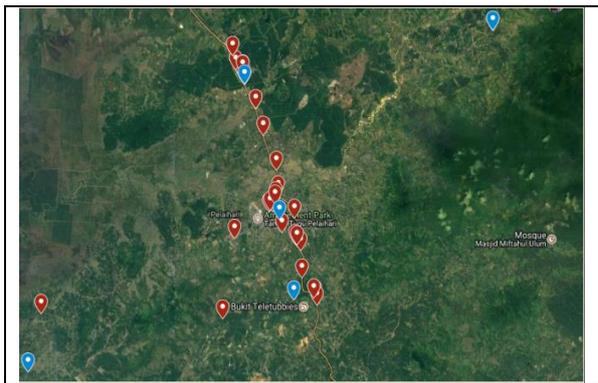
Gambar 3. Visualisasi Hasil Clustering Dataset 1 dengan 2 cluster



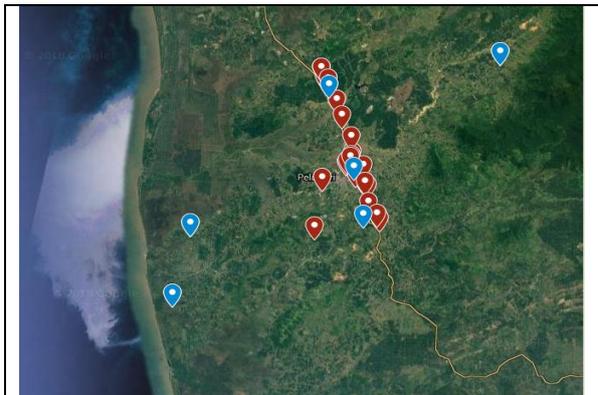
Gambar 4. Visualisasi Hasil Clustering Dataset 1 dengan 3 cluster



Gambar 5. Visualisasi Hasil *Clustering* Dataset 1 dengan 4 cluster



Gambar 6. Visualisasi Hasil *Clustering* Dataset 1 dengan 5 cluster



Gambar 7. Visualisasi Hasil *Clustering* Dataset 1 dengan 6 cluster

Setelah memperoleh *cluster* dari penerapan *K-Means*, dilakukan pengujian *silhouette coefficient* pada hasil *clustering* setiap *distance measure* untuk menentukan kualitas dan kekuatan *cluster*. Hasil pengujian dapat dilihat pada Tabel 4.

Tabel 4. Hasil Pengujian *Silhouette Coefficient* pada Dataset 1

Jumlah Cluster	<i>Silhouette Coefficients</i>
2	0.610674694
3	0.638989160
4	0.685351106
5	0.633353438
6	0.664788553

Hasil pengujian pada Tabel 4 menunjukkan bahwa *distance measure* paling optimal adalah untuk jumlah *cluster* sebanyak 4. Hasil *clustering* akhir untuk jumlah *cluster* sebanyak 4 untuk dataset 1 disajikan pada Tabel 5.

Tabel 5. Hasil *Clustering* Dataset 1 dengan 4 Cluster

ID	Latitude	Longitude	C-1	C-2	C-3	C-4
lap_01	-3.81597	114.78542			x	
lap_02	-3.72637	114.75423				x
lap_03	-3.73156	114.75729				x
lap_04	-3.81925	114.78816			x	
lap_05	-3.74428	114.76335				x
lap_06	-3.84506	114.7962			x	
lap_07	-3.74445	114.76345				x
lap_08	-3.85549	114.74524			x	
lap_09	-3.72397	114.75256				x
lap_10	-3.7988	114.77327			x	
lap_11	-3.7029	114.89416	x			
lap_12	-3.81555	114.785			x	
lap_13	-3.91452	114.63109		x		
lap_14	-3.72577	114.75623				x
lap_15	-3.80972	114.77807			x	
lap_16	-3.71626	114.75082				x
lap_17	-3.71626	114.75082				x
lap_18	-3.79004	114.77582			x	
lap_19	-3.80225	114.78481			x	
lap_20	-3.79884	114.77066			x	
lap_21	-3.81646	114.78601			x	
lap_22	-3.81273	114.75173			x	
lap_23	-3.77673	114.77485			x	
lap_24	-3.79298	114.77462			x	
lap_25	-3.85256	114.64546		x		
lap_26	-3.79807	114.77154			x	
lap_27	-3.72608	114.75615				x
lap_28	-3.83383	114.78884			x	
lap_29	-3.84828	114.79734			x	
lap_30	-3.79463	114.77399			x	
lap_31	-3.84438	114.79537			x	
lap_32	-3.75838	114.76751				x

Keterangan:

- C-1 = Cluster 1
- C-2 = Cluster 2
- C-3 = Cluster 3
- C-4 = Cluster 4

### 3.2. Hasil *K-Means Clustering* dan *silhouette coefficient* pada Dataset 2

*K-means clustering* diterapkan pada Dataset 2 secara berurutan untuk jumlah *kluster* sebanyak 2, 3, 4, 5, 6, dan 7. Kemudian dilakukan pengujian *silhouette coefficient* pada hasil *clustering* setiap *distance measure* pada jumlah *cluster* sebanyak 2, 3, 4, 5, 6, dan 7. Hasil pengujian *Silhouette Coefficient* pada Dataset 2 dapat dilihat pada Tabel 6.

Tabel 6. Hasil Pengujian *Silhouette Coefficient* pada Dataset 2

Jumlah Cluster	<i>Silhouette Coefficients</i>
2	0.383686134
3	0.316937195
4	0.236770018
5	0.247250522
6	0.277495919
7	0.308527071

Tabel 6 menunjukkan bahwa *Silhouette Coefficient* paling besar adalah pada jumlah *cluster* 2, yakni 0.383686134. Dengan kata lain, *distance measure* paling optimal adalah untuk jumlah *cluster* sebanyak 2. Hasil *clustering* akhir untuk jumlah *cluster* sebanyak 2 disajikan pada Tabel 7.

Tabel 7. Hasil *Clustering* Dataset 2 dengan 2 *Cluster*

No	A	B	C1	C2	C3	D1	D2	D3	E1	E2	E3	F1	F2	F3	F4	F5	G1	G2	G3	Cluster 1	Cluster 2
1	9	1	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	1	x	
2	14	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	x	
3	17	5	5	0	0	0	3	2	0	0	1	1	0	0	0	0	2	2	1		x
4	18	4	4	0	0	0	3	1	3	0	0	1	0	1	1	0	2	0	2		x
5	22	2	2	0	0	0	2	0	2	0	0	2	0	0	0	0	2	0	0		x
6	24	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0		x
7	25	3	2	0	1	0	2	1	1	0	0	1	0	0	0	0	2	0	1		x
8	26	2	1	1	0	0	2	0	2	0	0	0	0	2	0	0	2	0	0		x
9	27	2	0	2	0	0	0	2	0	0	0	0	0	0	0	0	2	0	0		x
10	28	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0		x
11	29	2	2	0	0	0	2	0	1	0	0	0	0	0	1	0	0	0	2		x
12	30	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0		x
13	32	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0		x
14	35	1	1	0	0	0	1	0	0	0	1	1	0	0	0	0	1	0	0		x
15	36	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0		x
16	37	2	2	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	2		x
17	38	1	0	1	0	0	1	0	1	0	0	1	0	0	0	0	1	0	0		x
18	41	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0		x
19	43	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0		x
20	45	2	2	0	0	0	1	1	0	0	1	1	0	0	0	0	1	0	1		x
21	47	1	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1		x
22	48	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0		x
23	49	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0		x
24	50	2	2	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0	2		x
25	51	2	2	0	0	0	2	0	0	0	0	0	0	0	0	0	0	1	1		x
26	53	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0		x
27	56	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0		x
28	60	1	1	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	1		x
29	70	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1		x
30	74	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1		x

Tabel 7 menunjukkan bahwa baik pada *cluster 1* maupun *cluster 2*, kecelakaan lalu lintas sering terjadi pada kendaraan roda 2, korbannya adalah pengemudi, terjadi pada hari kerja, dan terjadi di waktu pagi hari. Namun untuk keadaan korban, pada *cluster 1* korban dengan persentase terbanyak adalah meninggal dunia, yaitu 50%, sedangkan pada *cluster 2* korban dengan persentase terbanyak adalah meninggal dunia, yaitu 44.44%.

Berdasarkan hasil dan pembahasan yang diuraikan, pencegahan terjadinya kecelakaan lalu lintas dapat dilakukan melalui:

1. Sosialisasi kepada kelompok usia 9 sampai 16 tahun dan orang tuanya bahwa kelompok usia tersebut tidak diperbolehkan menjadi pengemudi kendaraan dimana hal ini perlu ditekankan tentang pentingnya kesadaran orang tua untuk memberikan pemahaman dan pengawasan terhadap anak-anaknya. Kelompok usia 9 sampai 16 tahun sebagai penumpang juga perlu diberikan sosialisasi mengenai kelengkapan dan keamanan sebagai penumpang.
2. Sosialisasi kepada kelompok usia 17 sampai 28 tahun, terutama umur 17 dan 18 tahun mengenai perlunya kesadaran berkendara secara aman dan sesuai dengan peraturan lalu lintas.
3. Perlunya pemberian marka jalan secara jelas dan tanda peringatan di daerah yang merupakan daerah rawan kecelakaan lalu lintas.
4. Perlunya pengaturan lalu lintas setiap pagi di hari kerja karena berdasarkan hasil *cluster* pada Tabel 7, waktu terjadinya kecelakaan lalu lintas adalah di hari kerja dan di pagi hari. Hal ini sesuai

dengan fakta bahwa hari kerja dan pagi hari merupakan waktu dengan kegiatan yang padat.

#### 4. KESIMPULAN DAN SARAN

Berdasarkan hasil dan pembahasan terhadap penerapan *K-Means Clustering* pada data kecelakaan lalu lintas dapat disimpulkan bahwa:

1. Penyebab kecelakaan lalu lintas pada 32 kecelakaan pada dataset adalah karena kesalahan pengemudi.
2. Korban kecelakaan yang paling sering adalah pengemudi dan usia yang paling sering menjadi korban kecelakaan adalah usia 9 sampai dengan 28 tahun. Sedangkan keadaan yang paling sering dialami korban kecelakaan lalu lintas adalah luka ringan.
3. Penerapan *K-means clustering* telah berhasil dilakukan pada dataset 1 yang divisualisasikan dalam bentuk peta lokasi kejadian.
4. Pengujian *silhouette coefficient* pada setiap *distance measure* untuk dataset lokasi kejadian menunjukkan bahwa *distance measure* paling optimal adalah jumlah *cluster* sebanyak 4, yakni 0.685351106.
5. Pengujian *silhouette coefficient* pada setiap *distance measure* untuk dataset yang memuat 7 atribut menunjukkan bahwa *distance measure* paling optimal adalah jumlah *cluster* sebanyak 2.
6. Penerapan *K-means clustering* telah berhasil dilakukan pada dataset 2, dimana kedua *cluster* yang dihasilkan menunjukkan bahwa kecelakaan lalu lintas sering terjadi pada kendaraan roda 2, korbannya adalah pengemudi, terjadi pada hari

kerja, dan terjadi di waktu pagi hari. Namun untuk keadaan korban di *cluster* 1 didominasi dengan keadaan meninggal dunia, sedangkan keadaan korban di *cluster* 1 didominasi dengan keadaan luka ringan.

7. Waktu terjadinya kecelakaan dengan peluang terbesar adalah saat hari kerja dan di pagi hari.

Penelitian selanjutnya diperlukan pengumpulan data yang memuat kejadian kecelakaan yang lebih banyak dan mencakup daerah yang lebih luas agar memperoleh keberagaman data sehingga proses *clustering* lebih optimal. Selain itu, pemilihan *centroid* awal juga perlu ditentukan menggunakan metode selain *simple random sampling*.

## 5. UCAPAN TERMA KASIH

Penulis mengucapkan terima kasih kepada Kementerian Riset, Teknologi, dan Pendidikan Tinggi atas bantuan biaya pada skema Penelitian Dosen Pemula Tahun Anggaran 2018.

## DAFTAR PUSTAKA

- ANGGARA, M., SUJIANI, H., dan NASUTION, H., 2016. Pemilihan *Distance Measure* pada *K-Means Clustering* untuk Pengelompokkan *Member* di *Alvaro Fitness*. *Jurnal Sistem dan Teknologi Informasi*, Vol 1, pp. 1-6.
- ANGGODO, Y. P., CAHYANINGRUM, W., FAUZIYAH, A. N., KHOIRIYAH, I. L., KARTIKASARI, O., dan CHOLISSODIN, I., 2017. *Hybrid K-Means dan Particle Swarm Optimization* untuk *Clustering* Nasabah Kredit. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 4, no. 2, pp. 104-110.
- FAJAR, M. S., 2015. Analisis Kecelakaan Lalu Lintas Jalan Raya di Kota Semarang Menggunakan Metode *K-Means Clustering*. Skripsi. Semarang: Jurusan Teknik Elektro Fakultas Teknik Universitas Negeri Semarang.
- HAN, J., KAMBER, M., dan PEI, J., 2011. *Data Mining: Concepts and Techniques* (3rd ed.). USA: Morgan Kaufmann.
- HUMAS POLRES TANAH LAUT, 2017. Polres Tanah Laut Bentuk Komunitas Korban Laka Lantas. Tersedia di <<http://humas-restala.blogspot.co.id/2017/03/polres-tanah-laut-bentuk-komunitas.html>> [Diakses 7 Juni 2017]
- ISWARI, L. dan AYU, E. G., 2015. Pemanfaatan Algoritma *K-Means* untuk Pemetaan Hasil Klusterisasi Data Kecelakaan Lalu Lintas, *Teknoin*, Volume 21, 1-11.
- MAULANA, A., 2017. Angka Kecelakaan Lalu Lintas Tahun Lalu Naik. *Kompas*. Tersedia di <<https://sains.kompas.com/read/2017/01/25/180500230/angka.kecelakaan.lalu.lintas.tahun.lalu.naik>> [Diakses 6 Juni 2017]
- RAHMAT, B., GAFAR, A.A, FAJRIANI, N., RAMDANI, U., UYUN, F.R., PURNAMASARI, Y., dan RANSI, N., 2017. Prosiding Seminar Nasional Riset Kuantitatif Terapan 2017: Implementasi *K-Means Clustering* pada RapidMiner untuk Analisis Daerah Rawan Kecelakaan. Kendari: Lembaga Pengembangan Sistem Informasi Universitas Halu Oleo.
- WAHID, M., 2016. Sudah 47 Nyawa Melayang Akibat Kecelakaan Lalu Lintas di Tanah Laut. Tersedia melalui: *Banjarmasin Post* <<http://banjarmasin.tribunnews.com/2016/08/22/sudah-47-nyawa-melayang-akibat-kecelakaan-lalulintas-di-tanahlaut>> [Diakses 6 Juni 2017]

*Halaman ini sengaja dikosongkan*