

PENGELOMPOKAN DATA KECELAKAAN LALU LINTAS DI KOTA TASIKMALAYA MENGGUNAKAN ALGORITMA K-MEANS

Iedam Fardian Anshori¹, Yeni Nuraini²

¹Universitas Adhirajasa Reswara Sanjaya
e-mail: iedam@ars.ac.id

²Universitas Adhirajasa Reswara Sanjaya
e-mail: yeninuraini96@gmail.com

Abstrak

Kenaikan jumlah penduduk mengakibatkan meningkatnya jumlah kebutuhan akan kendaraan bermotor sehingga memicu terjadinya kecelakaan lalu lintas. Pengelompokan data kecelakaan lalu lintas penting untuk dilakukan, karena akan memudahkan pihak polisi khususnya Unit Kecelakaan Lalu Lintas Polres Tasikmalaya Kota dalam mengurangi angka kecelakaan lalu lintas. Pada penelitian ini menggunakan Algoritma K-Means dengan tujuan untuk memperoleh informasi mengenai kelompok data kecelakaan lalu lintas berdasarkan waktu kejadian sehingga diketahui penyebab dari kecelakaan lalu lintas yang terjadi di Kota Tasikmalaya. Berdasarkan Validasi menggunakan metrik Davies Bouldin Index maka diperoleh 4 *cluster* yang dinilai dapat mengelompokan data dengan baik. *PerformanceVector* hasil evaluasi *cluster* yang dibentuk sebanyak 4 *cluster* dengan nilai sebesar 0,134. *Cluster* 1 dengan jumlah data kecelakaan sebanyak 74 kecelakaan yang paling banyak terjadi pada malam hari, *Cluster* 2 dengan jumlah kecelakaan sebanyak 16 kecelakaan yang paling banyak terjadi pada siang hari, *Cluster* 3 dengan jumlah kecelakaan sebanyak 6 kecelakaan yang paling banyak terjadi pada sore hari dan *Cluster* 4 dengan jumlah kecelakaan sebanyak 113 kecelakaan yang paling banyak terjadi pada pagi hari.

Kata Kunci: *Data Mining*, K-means, *Clustering*, Pelanggaran Lalu Lintas

Abstract

The increase in population has resulted in an increase in the need for motor vehicles, which has triggered traffic accidents. Traffic accident data grouping is important to do, because it will make it easier for the police, especially the Traffic Accident Unit of the Tasikmalaya City Police Precinct in reducing the number of traffic accidents. In this study using the K-Means Algorithm with the aim to obtain information about the group of traffic accident data based on the time of the incident so that the cause of the traffic accident that occurred in Tasikmalaya City is known. Based on Validation using the Davies Bouldin Index metric, 4 clusters are obtained which are considered to be able to group data well. *PerformanceVector* cluster evaluation results formed by 4 clusters with a value of 0.134. Cluster 1 with the number of accident data as many as 74 accidents that most occur at night, Cluster 2 with the number of accidents as many as 16 accidents most occur during the day, Cluster 3 with the number of accidents as many as 6 accidents most occur in the afternoon and Cluster 4 with the number of accidents as many as 113 accidents most occur in the morning.

Keywords: *Data Mining*, K-means, *Clustering*, *Traffic Violations*

1. Pendahuluan

Pertumbuhan jumlah penduduk di Indonesia sangatlah pesat, terbukti dengan terjadinya kenaikan jumlah penduduk pada tahun 2019 yang diproyeksikan mencapai 266,91 juta jiwa. Menurut jenis kelamin,

jumlah tersebut terdiri atas 134 juta jiwa laki-laki dan 132,91 juta jiwa perempuan (Supas, 2015).

Kenaikan jumlah penduduk mengakibatkan meningkatnya jumlah kebutuhan akan kendaraan bermotor

terutama kendaraan pribadi. Masyarakat lebih memilih menggunakan kendaraan pribadi karena dinilai lebih fleksibel dan juga lebih hemat. Peningkatan jumlah kendaraan mengakibatkan terjadinya permasalahan transportasi seperti pelanggaran lalu lintas yang cukup tinggi sehingga memicu terjadinya kecelakaan lalu lintas.

Kecelakaan lalu lintas masih menjadi salah satu penyumbang angka kematian yang besar di Indonesia. Laporan terbaru WHO menyebutkan, korban jiwa yang melayang akibat kecelakaan lalu lintas mencapai 1,35 juta jiwa pertahun. Direktorat Lalu Lintas Polda Metro Jaya mencatat jumlah kecelakaan yang terjadi selama periode Januari hingga November 2018 mencapai 5400 kejadian kecelakaan lalu lintas yang diantaranya 524 meninggal dunia, 804 korban luka berat, 5237 korban luka ringan. Namun pada kenyataan dilapangan data tersebut bisa saja bertambah karena seringkali masyarakat enggan untuk melaporkan kejadian kecelakaan lalu lintas ke pihak yang berwenang. Dari data tersebut dapat disimpulkan bahwa jalan raya masih menjadi tempat yang mematikan bagi pengguna jalan raya (Kesuma, 2018).

Data kecelakaan lalu lintas yang ada di pihak kepolisian didapat dari laporan polisi pada saat terjadinya kecelakaan lalu lintas, sehingga dari laporan tersebut menyebabkan menumpuknya data kecelakaan lalu lintas. Pengelompokan data kecelakaan lalu lintas sangat penting untuk dilakukan, karena akan memudahkan pihak polisi khususnya Unit LAKA LANTAS (Kecelakaan Lalu Lintas) Polres Tasikmalaya Kota dalam upaya penanganan yang tepat untuk mengurangi angka kecelakaan lalu lintas.

Algoritma k-means dapat mengklaster wilayah pada pelanggaran dalam berkendara berdasarkan banyak, sedang dan sedikit wilayah yang melakukan jenis pelanggaran dalam berkendara (Fajariyanti, 2017). Mengambil data-data penting yang tersembunyi dan dikelola kembali menjadi informasi yang akan dikelompokkan berdasarkan kriteria atau variabel tersebut. Sehingga dapat dihasilkan seluruh pengetahuan hubungan terdekat data-data pelanggaran berkendara tersebut (Ramadhani, Rahman, & Riskiyati, 2017). *Cluster* yang terbentuk menghasilkan pengetahuan terkait kasus perkara lalu lintas dan bisa digunakan sebagai kelas

target pada teknik klasifikasi (Delita & Fitri, 2014)

Penerapan data mining pada data pelanggaran lalu lintas untuk mengetahui pelanggaran dan daerah yang sering melanggar serta dilakukan analisa data tersebut dengan *clustering* dan algoritma k-means (Lestian & Ahmad, 2017). Analisis *cluster* berguna dalam meringkas data atau sejumlah variabel untuk menjadi lebih sedikit. Dalam melakukan proses data ini dapat dilakukan dengan mengelompokkan objek-objek berdasarkan kesamaan karakteristik tertentu diantara objek-objek yang hendak diteliti (Handoko, 2016).

Pada penelitian sebelumnya dengan algoritma K-Means dapat menampilkan pengelompokkan kemampuan siswa pada mata pelajaran Ujian Nasional, juga dapat digunakan untuk memantau perkembangan kemampuan setelah mengikuti pengayaan materi (Mulyani, Agustin, & Surgawi, 2018).

Berdasarkan permasalahan yang ada, maka pada kesempatan kali ini akan dilakukan penelitian terhadap data kecelakaan lalu lintas yang didapat dari Polres Tasikmalaya Kota. Data tersebut akan dikelompokkan dengan metode *clustering* menggunakan algoritma k-means dengan hasil yang diharapkan adalah berupa informasi mengenai kelompok data kecelakaan lalu lintas berdasarkan waktu kejadian sehingga diketahui penyebab dari kecelakaan yang terjadi di Kota Tasikmalaya.

2. Metode Penelitian Tahapan Penelitian

1. Mengidentifikasi Masalah

Tahapan pertama yaitu mengidentifikasi masalah yang ada, tahap ini adalah tahap yang paling penting dalam sebuah penelitian, karena semua jalannya penelitian akan dituntun oleh perumusan masalah. Dalam hal ini penulis merumuskan masalah yang berhubungan dengan data korban kecelakaan lalu lintas tahun 2018 yang terjadi di wilayah Kota Tasikmalaya.

2. Pengumpulan Data

Pengumpulan data merupakan kegiatan mencari data di lapangan yang akan digunakan untuk menjawab permasalahan penelitian. Validitas instrumen pengumpulan data serta kualifikasi pengumpul data sangat diperlukan untuk memperoleh data yang berkualitas. Penulis melakukan pengumpulan data langsung dari UNIT

LAKA LANTAS Polres wilayah Tasikmalaya Kota.

3. Mengolah data menggunakan algoritma *k-means*

Pada tahap ini penulis mengolah data yang telah didapat sesuai dengan tahapan metode yang penulis gunakan yaitu algoritma *k-means*.

4. Menguji Hasil Penelitian

Tahap ini adalah pengujian data yang telah diolah menggunakan aplikasi rapidminer kemudian dilakukan lagi pengolahan data secara manual untuk membuktikan bahwa algoritma *k-means* adalah algoritma yang paling tepat untuk mengolah data yang ada.

5. Kesimpulan dan Saran

Pada tahapan ini penulis menyimpulkan informasi yang didapat dari hasil pengolahan data menggunakan metode *k-means* dan memberikan saran atas kekurangan yang ada pada penelitian.

Metode Pengumpulan Data

Data yang digunakan merupakan data korban kecelakaan yang terjadi pada tahun 2018 di wilayah Kota Tasikmalaya yang penulis dapatkan pada saat melakukan pengamatan di unit LAKA LANTAS Polres Tasikmalaya Kota.

Populasi dan Sample Penelitian

Populasi dapat diartikan sebagai jumlah keseluruhan semua data kecelakaan lalu lintas yang diteliti (Doni Marlius, 2015:17). Jumlah keseluruhan data kecelakaan lalu lintas yang terjadi di Tasikmalaya pada tahun 2018 berjumlah 445 data kecelakaan kemudian data tersebut melalui tahap pembersihan data sehingga jumlahnya menjadi 438 data. Sedangkan Sampel merupakan bagian yang diambil dari populasi. Pengambilan sample pada penelitian ini berdasarkan pada rumus slovin dengan *margin of error* yang dikehendaki sebanyak 5% atau 0,005. Berikut merupakan perhitungan menggunakan rumus slovin :

$$n = \frac{N}{N(d)^2 + 1} \dots \dots \dots (3.1)$$

dimana :

n = Sampel

N = Populasi;

d = Nilai presisi 95% atau sig. = 0,05 (tingkat kesalahan yang dikehendaki adalah 5%)

jadi:

$$n = \frac{438}{438(0,05)^2 + 1} = 209 \text{ (dibulatkan)}$$

Metode Analisis Data

Pada tahap ini merupakan tahap untuk memastikan data korban kecelakaan yang dipilih telah layak untuk dilakukan proses pengolahan. Data awal yang penulis peroleh yaitu berjumlah 445 data kecelakaan lalu lintas dengan jumlah atribut sebanyak 9 atribut. Pembersihan Data

Tabel 1. Atribut Keseluruhan

No	Atribut	Keterangan
1.	TKP	Tempat terjadinya kecelakaan
2.	Waktu	Waktu terjadinya kecelakaan
3.	Usia	Usia korban kecelakaan
4.	Jenis Kelamin	Jenis kelamin korban kecelakaan
5.	Keadaan Korban	Keadaan korban kecelakaan
6.	Profesi	Profesi korban kecelakaan
7.	Jenis Kendaraan	Keadaan yang terlibat dalam kecelakaan
8	Modus	Modus yang dilakukan pengendara yang mengalrai kecelakaan
9.	Kerugian Material	Kerugian material dari kecelakaan yang terjadi

Tabel di atas merupakan atribut yang digunakan dalam penelitian ini.

1. Data *Cleanning*

Pada tahap *Data Cleaning* atau pembersihan data, dilakukan pemilihan atribut data yang akan digunakan dalam penelitian, maka untuk atribut data yang tidak digunakan modus dan kerugian material akan dihilangkan.

2. *Data Transformation*

Data yang berjenis nominal seperti TKP, waktu, usia, jenis kelamin, keadaan korban, profesi dan kendaraan yang terlibat harus dilakukan proses inialisasi data terlebih dahulu kedalam bentuk angka atau numerikal. Untuk melakukan inialisasi tersebut dapat dilakukan dengan cara pengurutan angka berdasarkan frekuensinya.

Berikut merupakan tabel inialisasi dari Tempat kejadian perkara yang terjadi kecelakaan di wilayah kota Tasikmalaya.

Tabel 2.
Tabel Inisialisasi TKP

No	TKP	Frekuensi	Inisialisasi
1	Jl. Raya Rajapolah	38	1
2	Jl. AH Nasution	25	2
3	Jl. Raya Jamanis	23	3
4	Jl. Raya Ciawi	21	4
5	Jl. Raya Kadipaten	21	5
6	Jl. Ir. H. Juanda	17	6
7	Jl. Syekh Abdul Muhyi	16	7
8	Jl. Raya Manonjaya	13	8
9	Jl. Letjen Ibrahim Adjie	12	9
10	Jl. Perintis kemerdekaan	12	10
11	Jl. Raya Cisayong	11	11

Berikut merupakan tabel inisialisasi waktu yang dibagi kedalam 4 waktu yaitu pagi, siang, sore dan malam.

Tabel 3.
Tabel inisialisasi Waktu

No	Waktu	Frekuensi	Inisialisasi
1	Malam	56	1
2	Sore	45	2
3	Siang	59	3
4	Pagi	49	4

Agar inisialisasi tidak terlalu panjang dan banyak, maka penulis menggunakan rumus jangkauan nilai untuk menginisialisasi umur. Selengkapnya dapat dilihat pada tabel berikut :

Tabel 4.
Tabel inisialisasi Umur

No	Umur	Frekuensi	Inisialisasi
1	15 – 26	71	1
2	27 – 38	83	2
3	39 – 50	50	3
4	51 – 62	3	4
5	63 – 92	2	5

Jenis kelamin merupakan jenis kelamin korban kecelakaan yang terlibat. Berikut merupakan tahap inisialisasi atribut jenis kelamin.

Tabel 5.
Tabel inisialisasi Jenis Kelamin

No	JK	Frekuensi	Inisialisasi
1	Laki-laki	154	1
2	Perempuan	55	2

Pada kecelakaan lalu lintas terdapat tiga jenis keadaan korban yaitu luka ringan, luka berat dan meninggal dunia. Berikut merupakan inisialisasi keadaan korban.

Tabel 6.
Tabel inisialisasi Keadaan Korban

No	Keadaan	Frekuensi	Inisialisasi
1	Luka Ringan	146	1
2	Meninggal Dunia	47	2
3	Luka Berat	16	3

Korban yang terlibat kecelakaan berasal dari kalangan yang berbeda – beda, berikut merupakan inisialisasi profesi korban kecelakaan lalu lintas di kota Tasikmalaya :

Tabel 7.
Tabel inisialisasi Profesi

No	Profesi	Frekuensi	Inisialisasi
1	Wiraswasta	49	1
2	Pelajar	30	2
3	Buruh	28	3
4	IRT	25	4
...
25	Security	1	25
26	Staff Desa	1	26
28	TNI	1	28

Banyaknya kendaraan yang melintas di jalan raya meningkatkan potensi terjadinya kecelakaan lalu lintas, berikut merupakan inisialisasi dari kendaraan yang terlibat kecelakaan lalu lintas di wilayah tasikmalaya :

Tabel 8.
Tabel inisialisasi Jenis Kendaraan

No	Kendaraan	Frekuensi	Inisialisasi
1	R2	140	1
2	R4	67	2
3	R6	2	3

Setelah semua data telah melalui tahap inisialisasi, kemudian dilakukan tahap transformasi data keseluruhan sehingga data tersebut tpat dihitung menggunakan

algoritma *k-means*. Berikut adalah transformasi data kecelakaan lalu lintas :

Tabel 9.
Tabel Contoh *Dataset* yang sudah dilakukan *Transformasi*

NO	T K P	U s i a	Jenis Kelami n	Ke ada an Kor ba n	PRO FESI	Jeni s Ken dara an
1	5	2	2	1	2	5
2	5	2	2	1	1	5
3	2	3	1	1	2	9
4	2	3	2	2	3	9
5	2	3	3	1	1	2
6	1	4	1	2	1	4
7	1	4	3	1	1	2
8	1	4	1	1	1	1
9	3	4	2	1	1	1
10	9	3	2	1	2	1
...
208	5	3	2	1	3	1
209	2	4	3	1	1	6

Tahap pertama dalam algoritma *k-means* adalah menentukan jumlah *cluster*, pada penelitian ini terdapat 4 *cluster* sesuai dengan pembuatan kelompok waktu terjadinya kecelakaan lalu lintas yaitu pagi, siang, sore dan malam. Penentuan centroid awal dipilih secara random dengan atribut TKP, waktu, usia, jenis kelamin, keadaan korban, profesi dan jenis kendaraan yang terlibat.

Berikut merupakan *Dataset* kecelakaan lalu lintas yang telah melalui tahapan *preprocessing* diantaranya tahap *Cleaning* atau pembersihan data, tahap inialisasi data yang bertipe nominal dan yang terakhir yaitu tahap transformasi semua data yang telah melewati tahap inialisasi.

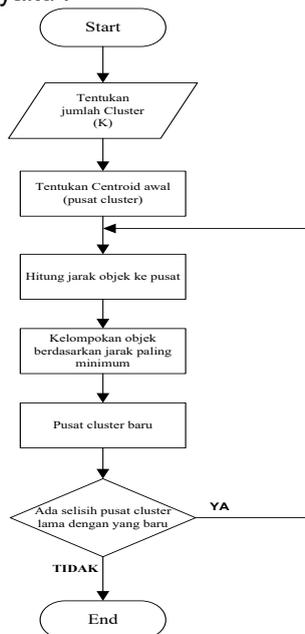
Tabel 10.
Dataset Hasil Preprocessing

N O	T K P	W ak tu	U s i a	Je nis Kel am in	Ke ada an Kor ba n	Pr of esi	Jeni s Ken dara an
1	5	2	2	1	2	5	1
2	5	2	2	1	1	5	3
3	2	3	1	1	2	9	1
4	2	3	2	2	3	9	2
5	2	3	3	1	1	2	1
6	1	4	1	2	1	4	1
7	1	4	3	1	1	2	1
8	1	4	1	1	1	1	1
9	3	4	2	1	1	1	2
10	9	3	2	1	2	1	1
...
208	5	3	3	1	3	2	2
209	2	4	1	2	1	5	2
210	8	4	1	2	1	5	2
211	6	2	1	1	2	1	2
212	7	3	3	1	2	2	2
213	1	4	2	1	2	1	1
214	4	4	3	1	2	2	2

3. Hasil dan Pembahasan

3.1 Perhitungan Algoritma K-means

Pada proses perhitungan menggunakan Algoritma *k-means* mempunyai beberapa tahapan, yaitu :



Gambar 1. Flowchart K-Means

206	5	1	3	1	1	2	1
207	2	2	1	1	2	22	1
208	5	3	2	1	3	1	1
209	2	4	3	1	1	6	1

Tahap selanjutnya yaitu menentukan titik pusat awal *cluster* (*centroid*) yang dipilih secara random. Pada penelitian kali ini diambil dari data ke-38, 9, 4 dan 57.

Tabel 11.
Centroid Awal

Centroid	N O	T K P	W a k t u	U s i a	J e n i s K e l a m i n	K e a d a a n K o r b a n	P r o f e s i	J e n i s K e n d a r a n
Cluster 0	38	8	3	1	2	1	4	2
Cluster 1	9	11	3	2	2	2	10	1
Cluster 2	4	1	1	2	1	1	26	1
Cluster 3	57	1	4	2	1	1	1	2

Setelah menentukan *centroid* awal, maka tahap selanjutnya yaitu menghitung jarak setiap data ke *centroid* terdekat untuk menentukan *cluster* yang diikuti oleh data tersebut menggunakan rumus *Euclidean distance* seperti pada persamaan (2.1).

Berikut contoh perhitungan jarak ke setiap *centroid*.

➤ Data ke - 1 ke pusat *cluster* ke - 0

$$d(x_1, c_0) = \sqrt{(5-8)^2 + (2-3)^2 + (2-1)^2 + (1-2)^2 + (2-1)^2 + (5-4)^2 + (1-2)^2} = \sqrt{9+1+1+1+1+1+1} = \sqrt{15} = 3,87$$

Dari perhitungan data ke-1 dengan *cluster* ke-0 maka dihasilkan jarak sebesar 3,87

➤ Data ke - 1 ke pusat *cluster* ke - 1

$$d(x_1, c_1) = \sqrt{((5-11)^2 + (2-3)^2 + (2-2)^2 + (1-2)^2 + (2-2)^2 + (5-10)^2 + (1-1)^2)} = \sqrt{36+1+0+1+0+25+0} = \sqrt{63} = 7,94$$

Dari perhitungan data ke-1 dengan *cluster* ke-1 maka dihasilkan jarak sebesar 7,94

➤ Data ke - 1 ke pusat *cluster* ke - 2

$$d(x_1, c_2) = \sqrt{((5-4)^2 + (2-1)^2 + (2-1)^2 + (1-2)^2 + (2-1)^2 + (5-26)^2 + (1-1)^2)} = \sqrt{1+1+1+1+1+441+0} = \sqrt{466} = 21,42$$

Dari perhitungan data ke-1 dengan *cluster* ke-2 maka dihasilkan jarak sebesar 21,42

➤ Data ke - 1 ke pusat *cluster* ke - 3

$$d(x_1, c_3) = \sqrt{((5-1)^2 + (2-4)^2 + (2-2)^2 + (1-1)^2 + (2-1)^2 + (5-1)^2 + (1-2)^2)} = \sqrt{16+4+0+0+1+16+1} = \sqrt{38} = 6,16$$

Dari perhitungan data ke-1 dengan *cluster* ke-3 maka dihasilkan jarak sebesar 6,16

Hasil perhitungan di atas yaitu jarak data pertama dengan dengan *cluster* ke - 0 sebesar 3,87, jarak data pertama dengan dengan *cluster* ke - 2 sebesar 7,94, jarak data pertama dengan dengan *cluster* ke - 3 sebesar 21,42 dan jarak data pertama dengan dengan *cluster* ke - 4 sebesar 6,16. Kemudian cari data yang memiliki jarak terdekat dengan *centroid*. Berdasarkan perhitungan, *Cluster* 0 adalah yang paling dekat dengan *Centroid*. Berikut merupakan perhitungan lengkap dari iterasi 1.

Tabel 12.
Hasil Perhitungan Menggunakan rumus *Euclidean* pada iterasi ke-1

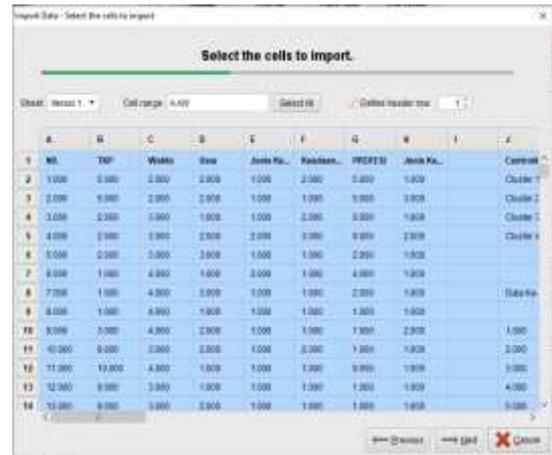
Data Ke-i	Jarak ke centroid				Terdekat	Cluster yang diikuti
	C0	C1	C2	C3		
1	3,87	7,94	21,42	6,16	3,87	C0
2	3,74	8,2	21,42	6,0	3,74	C0
3	8,00	9,1	17,0	8,3	8,00	C0
4	8,12	9,1	17,32	8,4	8,12	C0
5	6,78	12,17	24,12	2,24	2,24	C3

6	7,1 4	11, 79	22, 25	3,4 6	3,46	C3
7	7,7 5	12, 96	24, 21	1,7 3	1,73	C3
8	7,8 1	13, 60	25, 20	1,4 1	1,41	C3
9	6,0 8	12, 21	25, 28	2,0 0	2,00	C3
10	3,7 4	9,2 7	26, 34	8,1 9	3,74	C1
...
20 0	4,6 9	8,2 5	26, 19	10, 34	4,69	C0
20 1	1,4 1	6,1 6	22, 41	8,1 9	1,41	C0
20 2	4,0 0	10, 49	25, 57	5,5 7	4,00	C0
20 3	3,3 2	9,1 1	24, 88	6,3 2	3,32	C0
20 4	4,8 0	9,1 1	27, 11	10, 10	4,80	C0
20 5	5,2 0	10, 82	24, 43	3,4 6	3,46	C3
20 6	4,8 0	10, 34	24, 35	5,2 9	4,80	C0
20 7	19, 08	15, 10	4,4 7	21, 19	4,47	C2
20 8	5,0 0	10, 91	25, 48	4,6 9	4,69	C3
20 9	6,8 6	10, 05	20, 27	5,2 9	5,29	C3

Selanjutnya kelompokan data sesuai dengan *cluster* terdekat. Dari data tersebut akan didapat *centroid* baru berdasarkan hasil rata – rata dari setiap *clusternya*.

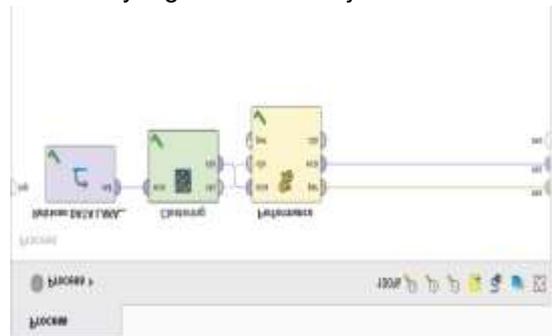
3.2. Perhitungan Algoritma K-Means Menggunakan Rapidminer

Pengujian dilakukan menggunakan aplikasi Rapidminer 9.0. Berikut merupakan tampilan *import dataset* kecelakaan lalu lintas yang telah melalui tahapan *preprocessing*.



Gambar 2. *Import Dataset* Kecelakaan Lalu Lintas

Berikut merupakan *design* pemodelan dataset kecelakaan menggunakan algoritma k-means yang dibentuk menjadi 4 *cluster*.

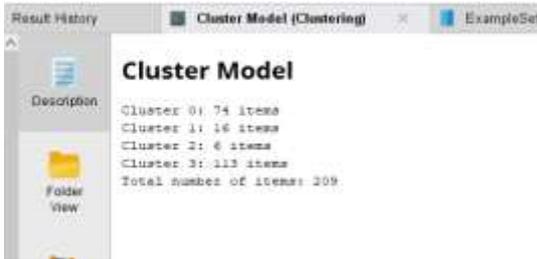


Gambar 3. *Design Pemodelan Data* Kecelakaan Lalu Lintas

Berikut merupakan data keseluruhan hasil *clustering*.

Gambar 4. Data Hasil *Clustering*

Berdasarkan pemodelan dari 4 *cluster* maka diperoleh data yang tergabung ke *cluster* 0 sebanyak 74 *items*, *cluster* 1 sebanyak 16 *items*, *cluster* 2 sebanyak 6 *items* dan *cluster* 3 sebanyak 113 *item*.

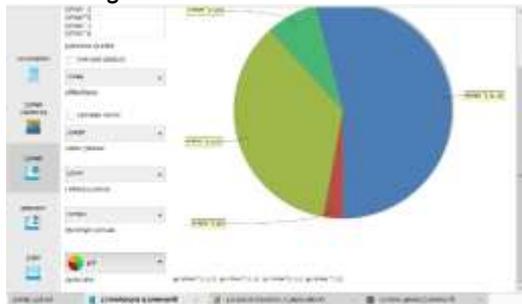


Gambar 5. Cluster Model

Dari 4 cluster diperoleh centroid akhir untuk masing-masing cluster seperti yang terdapat pada gambar 6 :

Gambar 6. Centroid Akhir

Berikut merupakan pie chart hasil clustering data kecelakaan lalu lintas.



Gambar 7. Pie Chart

3.3. Validasi Hasil

Berdasarkan validasi menggunakan metrik Davies Bouldin Index di aplikasi Rapidminer maka diperoleh 4 cluster yang dinilai dapat mengelompokan data dengan baik. PerformanceVector hasil evaluasi cluster yang dibentuk sebanyak 4 cluster dengan nilai sebesar 0,939 yang dapat dilihat pada gambar 8.



Gambar 8. PerformanceVector

Pada penentuan cluster menggunakan metrik Davies Bouldin Index yaitu menggunakan data centroid dari iterasi akhir pada proses clustering.

3.4. Pembahasan

Berdasarkan proses clustering dengan algoritma k-means menggunakan aplikasi Rapidminer, maka menghasilkan Informasi sebagai berikut :

Tabel 13.
Rekapitulasi Data Cluster 0-3

Variabel		C 0	C 1	C 2	C 3
Waktu	1 Malam	30	2	2	22
	2 Sore	17	3	2	23
	3 Siang	16	9	1	33
	4 Pagi	11	2	1	35
TKP	1 Jl. Raya Rajapolah	-	2	3	32
	2 Jl. AH Nasution	-	5	1	19
	3 Jl. Raya Jamanis	-	-	-	23
	4 Jl. Raya Ciawi	-	1	1	19
	5 Jl. Raya Kadipaten	-	1	-	20
	6 Jl. Ir. H. Juanda	16	1	-	-
	7 Jl. Syekh Abdul Muhyi	16	-	-	-
	8 Jl. Raya Manonjaya	11	1	1	-
	9 Jl. Letjen Ibrahim Adjie	10	2	-	-
	10 Jl. Perintis kemerdekaan	11	1	-	-

	11	Jl. Raya Cisayong	10	2	-	-
Usia	1	27 – 38	24	5	3	39
	2	15 – 26	28	8	3	44
	3	39 – 50	21	-	-	29
	4	51 – 62	1	1	-	1
	5	87 – 92	-	2	-	-
Jenis Kelamin	1	Laki-laki	58	11	6	79
	2	Perempuan	16	5	-	34
Keadaan Korban	1	Luka Ringan	55	8	5	78
	2	Meninggal Dunia	13	5	1	28
	3	Luka Berat	6	3	-	7
Profesi	1	Wiraswasta	22	-	-	27
	2	Pelajar	9	-	-	21
	3	Buruh	12	-	-	16
	4	IRT	9	-	-	16
	5	Pegawai Swasta	8	-	-	24
	6	Mahasiswa	11	-	-	4
	7	Belum bekerja	1	-	-	3
	8	Pedagog	2	-	-	2
	9	PNS	-	4	-	-
	10	Guru	-	4	-	-
	11	Pensiunan	-	2	-	-
	12	Petani	-	1	-	-
	13	Bidan	-	1	-	-
	14	Ojek	-	1	-	-
	15	Pengemudi	-	2	-	-

	17	DINAS PU	-	1	-	-
	18	Karyawan BUMN	-	-	1	-
	19	Kepala Desa	-	-	1	-
	22	Mekanik	-	-	1	-
	23	Pegawai BUMD	-	-	1	-
	26	Staff Desa	-	-	1	-
	28	TNI	-	-	1	-
Jenis Keadaan	1	R2	50	11	5	74
	2	R4	23	5	1	38
	3	R6	1	-	-	1

Berdasarkan uraian di atas, maka dapat disimpulkan bahwa pada *Cluster 0* kecelakaan banyak terjadi pada malam hari, salah satu penyebabnya yaitu intensitas cahaya atau penerangan yang kurang memadai. Pada *Cluster 1* kecelakaan banyak terjadi pada siang hari karena merupakan jam pulang sekolah. Pada *Cluster 2*, kecelakaan banyak terjadi pada sore hari, hal tersebut diakibatkan karena sore hari merupakan jam pulang kerja. Selanjutnya pada *Cluster 3*, kecelakaan banyak terjadi pada pagi hari, karena pagi hari adalah waktu untuk memulai aktivitas bagi pelajar, karyawan dan yang lainnya.

4. Kesimpulan

Berdasarkan hasil penelitian mengenai pengelompokan data kecelakaan menggunakan algoritma k-means maka diperoleh beberapa kesimpulan sebagai berikut : Berdasarkan validasi Davies Bouldin Index (DBI), dengan jumlah *sample* sebanyak 209 data terbentuk menjadi 4 *cluster* dengan nilai ketepatan sebesar 0,939; Dari pengelompokan data kecelakaan lalu lintas berdasarkan waktu terjadinya kecelakaan yang terbagi kedalam 4 *cluster* menghasilkan informasi pada *cluster 0* dengan jumlah data kecelakaan sebanyak 74 kecelakaan terjadi pada malam hari, *cluster 1* dengan jumlah kecelakaan sebanyak 16 kecelakaan yang terjadi pada siang hari, *cluster 2* dengan jumlah kecelakaan sebanyak 6 kecelakaan yang

terjadi pada sore hari dan *cluster* 3 dengan jumlah kecelakaan sebanyak 113 kecelakaan yang terjadi pada pagi hari; Berdasarkan informasi mengenai waktu terjadinya kecelakaan, maka dapat menjadi pengetahuan untuk pihak Unit Laka Lantas Polres Tasikmalaya Kota dengan cara menghubungkannya dengan tempat terjadinya kecelakaan lalu lintas guna melakukan penanganan yang tepat untuk mengurangi angka kecelakaan lalu lintas di kota Tasikmalaya.

penduduk-indonesia-2019-
mencapai-267-juta-jiwa

Referensi

- Delita, B., & Fitri, N. (2014). Clustering Pelanggaran Berkendaraan Menggunakan Algoritma K-means Pada Polres Binjai. *Jurnal STMIK Kaputama*.
- Fajariyanti, W. (2017). Clustering Wilayah Pada Pelanggaran Berkendaraan Menggunakan. *Sikmi-Techsain*, Vol 01, No.09.
- Handoko, K. (2016). Penerapan Data Mining dalam Meningkatkan Mutu Pembelajaran. *JURNAL SISTEM INFORMASI DAN TEKNOLOGI INFORMASI. Teknosi*, 31-40.
- Kesuma, D. P. (2018, Desember 20). *Polisi Akui Angka Kecelakaan Lalu Lintas Sepanjang 2018 Cenderung Naik*. Diambil kembali dari [Tribunnews.com](https://www.tribunnews.com): <https://www.tribunnews.com>
- Lestian, A. C., & Ahmad, Z. F. (2017). Implementasi Algoritma K-Means Pada Data pelanggaran Lalu Lintas Di Pengadilan Negeri Purwodadi. *Jurnal UDINUS*.
- Mulyani, E. D., Agustin, S., & Surgawi, N. (2018). Implementasi Algoritma K-Means dan Fp-Growth Untuk Rekomendasi Bimbingan Belajar Berdasarkan Segmentasi Akademik Siswa. *Informatic Technique Journal*, 160-173.
- Ramadhani, N., Rahman, A. F., & Riskiyati, D. (2017). Analisis Cluster Data Register Perkara lalu Lintas Menggunakan Algoritma K-Means. *SESINDO* 9.
- Supas. (2015, 01 01). *Jumlah Penduduk Indonesia 2019 Mencapai 267 Juta Jiwa*. Diambil kembali dari Kata data: <https://databoks.katadata.co.id/data-publish/2019/01/04/jumlah->