

# Mengelompokkan Daerah Rawan Kecelakaan Di Sumatera Utara Dengan Algoritma Clustering

M. Sapriyaldi<sup>1</sup>, Dedy Hartama<sup>2</sup>

<sup>1</sup>STIKOM Tunas Bangsa Pematangsiantar

<sup>2</sup>STIKOM Tunas Bangsa Pematangsiantar

<sup>1</sup>msapriyaldi60@gmail.com, <sup>2</sup>dedyhartama@amiktunasbangsa.ac.id \*

## Abstract

*The large population has a very large need for motorized vehicles, both 2-wheeled and 4-wheeled, which most people consider to be a primary need, not a secondary need. The large number of vehicle users causes traffic congestion so that the number of accidents increases which can result in many fatalities, minor injuries and serious injuries. The aim of this research is to group accident-prone areas in North Sumatra using the clustering method. The data source used in the research is from BPS on the topic of traffic accidents in the North Sumatra region from 2015-2022. The method used to solve this problem is K-Means Data Mining. The results obtained from this search are 3 clusters with a DBI value of 0.384, cluster 1 contains 1 region, cluster 2 contains 16 regions, and cluster 3 contains 11 regions. Carrying out this research can provide knowledge input for further research regarding the development of the k-means clustering method and help the police, especially the traffic accident handling units in each region, in predicting accidents more easily and tracing possible cause's accidents in the area.*

*Keywords: data mining, DBI, k-means, traffic accident, vehicles.*

## Abstrak

Jumlah penduduk yang besar mempunyai kebutuhan yang sangat besar terhadap kendaraan bermotor, baik roda 2 maupun roda 4, yang oleh sebagian besar masyarakat dianggap sebagai kebutuhan primer, bukan kebutuhan sekunder. Banyaknya pengguna kendaraan menyebabkan padatnya lalu lintas sehingga angka kecelakaan meningkat sehingga dapat mengakibatkan banyak korban jiwa, luka ringan dan luka berat. Tujuan dari penelitian ini adalah untuk mengelompokkan daerah rawan kecelakaan di Sumatera Utara dengan menggunakan metode clustering. Sumber data yang digunakan dalam penelitian adalah dari BPS (<https://sumut.bps.go.id/indicator/17/277/1/nomor-kecelakaan-lalu-lintas-menurut-korban-dan-kabupaten-kota.html>) dengan topik kecelakaan lalu lintas di wilayah Sumatera Utara dari tahun 2015-2022. Metode yang digunakan untuk menyelesaikan permasalahan tersebut adalah K-Means Data Mining. Hasil yang diperoleh dari pencarian ini adalah 3 cluster dengan nilai DBI sebesar 0,384, cluster 1 berisi 1 wilayah, cluster 2 berisi 16 wilayah, dan cluster 3 berisi 11 wilayah. Dengan dilakukannya penelitian ini dapat memberikan masukan ilmu untuk penelitian selanjutnya mengenai pengembangan metode k-means clustering dan membantu pihak kepolisian khususnya unit penanganan kecelakaan lalu lintas masing-masing daerah dalam memprediksi kecelakaan dengan lebih mudah dan menelusuri kemungkinan penyebabnya kecelakaan di kawasan.

Kata kunci: data mining, DBI, k-means, kecelakaan lalu lintas, kendaraan.

©This work is licensed under a Creative Commons Attribution - ShareAlike 4.0 International License

## 1. Pendahuluan

Banyaknya penggunaan kendaraan membuat aktivitas di lalu lintas semakin padat dan banyaknya jalan berlubang di sepanjang jalan dapat mengakibatkan kecelakaan terjadi, kecelakaan antar kendaraan ataupun kecelakaan tunggal yang mengakibatkan banyak korban meninggal, luka ringan, luka berat maupun cacat permanen [1]. Kendaraan merupakan alat transportasi yang menjadi kebutuhan masyarakat luas, dengan adanya transportasi memudahkan aktivitas yang memiliki jarak cukup jauh dengan waktu yang lebih singkat.

Kecelakaan lalu lintas disebabkan dari peningkatan mobilitas transportasi yang cukup pesat [2]. Kecelakaan merupakan suatu tindakan maupun kejadian yang tidak terduga dan tidak dapat dikendalikan dimana kontak dan reaksi suatu benda, bahan maupun radiasi menyebabkan cedera ataupun kemungkinan besar menyebabkan kematian [3].

Peningkatan pertumbuhan penduduk di Sumatera Utara membuat persentase penggunaan kendaraan meningkat mengakibatkan semakin padatnya lalu lintas dapat mengakibatkan kecelakaan. Kecelakaan merupakan peristiwa yang tidak diinginkan dan tidak diharapkan yang dapat menyebabkan cedera atau bahkan kematian [4]. Tingginya angka kematian akibat kecelakaan lalu lintas dikalangan generasi muda disebabkan rendahnya kesadaran akan resiko bahaya yang besar yang ada di jalan raya [5].

K-Means merupakan salah satu metode *clustering* data berbasis jarak dengan cara membagi data menjadi beberapa kelompok dengan tipe data numerik [6].

Terdapat beberapa penelitian yang terkait dengan penggunaan metode K-Means seperti yang dilakukan oleh Kurniawan dan Jajuli pada tahun 2022 [7] yaitu mengelompokkan data kecelakaan lalu lintas di kecamatan Cileungi dengan menggunakan K-Means dengan memperoleh hasil penelitian berupa penggunaan evaluasi *silhouette coefficient* berupa 3

*cluster*, *cluster* 1 nilai evaluasi 0,35, *cluster* 2 dengan nilai evaluasi 0,22 dan *cluster* 3 dengan nilai evaluasi 0,38. Penelitian selanjutnya Gultom dan kawan-kawan pada tahun 2020 [8] yaitu untuk mengetahui tingkat kejahatan daerah Pematangsiantar dengan menggunakan Algoritma K-Means memperoleh 2 *cluster*, *cluster* 1 tingkat kejahatan tinggi sebanyak 4 kecamatan dan *cluster* 2 tingkat kejahatan rendah sebanyak 2 kecamatan.

Pada penelitian ini pengelompokan daerah rawan kecelakaan menggunakan algoritma K-Means. Banyak *cluster* yang akan digunakan yaitu sebanyak 3 *cluster*. Yang mana algoritma K-Means memiliki tingkat keakuratan yang cukup baik untuk pengelompokan data dalam jumlah yang besar, juga algoritma K-Means ini sangat mudah untuk dipahami dan mudah untuk diterapkan terhadap data [9].

Untuk mengetahui banyak *cluster* yang paling maksimal untuk digunakan untuk mengelompokkan daerah rawan kecelakaan menggunakan *Davies Bouldin Index* (DBI). DBI (*Davies Bouldin Index*) merupakan metode evaluasi *cluster* dengan memaksimalkan jarak antar *cluster* dan juga mencoba mengecilkan jarak antar titik dalam suatu *cluster* pada metode pengelompokan [10]. Semakin kecil nilai DBI banyak *cluster* yang digunakan maka semakin optimal pula banyak *cluster* itu untuk digunakan, dengan nilai DBI paling rendah mendekati 0 maka paling cocok untuk digunakan.

Tujuan penelitian ini yaitu untuk mengetahui daerah mana yang tingkat kecelakaannya yang tinggi, sedang dan rendah menggunakan algoritma K-Means dan penerapan DBI untuk memperoleh banyak *cluster* yang paling baik untuk digunakan. Hasil dari pengelompokan daerah rawan kecelakaan di Provinsi Sumatera Utara dapat membantu pemerintah agar dapat tahu di mana tingkat kecelakaannya tinggi sehingga dapat melakukan tindak pencegahan akan terjadinya kecelakaan pada tiap daerah berdasarkan pengclusteran daerah rawan kecelakaan.

## 2. Metode Penelitian

Penelitian ini menggunakan data yang diperoleh dari website Badan Pusat Statistik Sumatera Utara dengan topik kecelakaan lalu lintas di Sumatera Utara berdasarkan daerah dari tahun 2015-2022. Data akan dibagi menjadi 3 *cluster* yaitu *cluster* 1 tingkat kecelakaan tinggi, *cluster* 2 tingkat kecelakaan rendah dan *cluster* 3 dengan tingkat kecelakaan sedang..

### 2.1. Data Mining

Data mining bagian dari suatu proses *Knowledge Discovery in Database* (KDD) yang membantu menciptakan pola atau pola dalam data dengan menerapkan algoritma yang unik [11]. Data mining merupakan ilmu yang mempelajari pola data yang belum di ketahui secara pasti seperti pada penelitian [12-13].

### 2.2. K-Means

Pengelompokan K-Means adalah pengelompokan non hirarki untuk melakukan pengelompokan data dalam beberapa kelompok [14]. Langkah-langkah dalam penerapan algoritma k-means sebagai berikut [15] :

- 1) Menentukan banyak *cluster* yang akan digunakan.
- 2) Menentukan *centroid* (titik pusat) secara acak pada iterasi pertama, untuk iterasi selanjutnya menentukan *centroid* menggunakan persamaan (1) berikut :

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \quad (1)$$

Keterangan :

$V_{ij}$  = Rata-rata titik pusat kelompok ke-i untuk variabel ke-j

$N_i$  = Banyak data kelompok ke-i

i, k = Indikator dari kelompok

j = Indikator dari variable

$X_{kj}$  = nilai data ke-k variabel ke-j untuk kelompok

- 3) Setiap catatan dihitung jaraknya terhadap *centroid* menggunakan persamaan (2) berikut :

$$d(x, y) = \sqrt{\sum_{k=0}^{N_i} (x_i - y_i)^2} \quad (2)$$

- 4) Mengelompokkan data berdasarkan jarak *centroid* terdekat.
- 5) Mengulangi langkah ke 2 sampai ke 4 hingga memperoleh nilai *centroid* yang sama.

### 2.3. DBI (*Davies Bouldin Index*)

DBI (*Davies Bouldin Index*) merupakan cara memperoleh nilai validasi suatu kelompok yang menggunakan metode *clustering* [16]. Berikut langkah untuk mencari nilai DBI [17] :

- 1) Mencari nilai SSW (*Sum Of Square Between Cluster*) menggunakan rumus (3) :

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i) \quad (3)$$

Keterangan :

SSW = *Sum Of Square Within Cluster*

$m_i$  = Banyak data pada kelompok

$x_j$  = Data pada kelompok

$c_i$  = Titik pusat kelompok ke-i

$d(x_j, c_i)$  = Jarak data terhadap titik pusat

- 2) Mencari nilai SSB (*Sum Of Square Within Cluster*) menggunakan rumus (4) :

$$SSB_{ij} = d(c_i, c_j) = \sqrt{(x_{c_i} - x_{c_j})^2 + (y_{c_i} - y_{c_j})^2} \quad (4)$$

Keterangan :

SSB = *Sum Of Square Between Cluster*

D (ci , cj) = Jarak titik pusat satu dengan yang lain

- 3) Mencari nilai Rasio dengan persamaan (5) berikut :

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \quad (5)$$

Keterangan :

Rij = Nominal perbandingan kelompok i dan kelompok j

SSWi = *Sum Of Square Within Cluster* ke i

SSWj = *Sum Of Square Within Cluster* ke j

SSBij = *Sum Of Square Between Cluster* ke i dan j

- 4) Mencari nilai DBI (*Davies Bouldi Index*) dengan persamaan (6) berikut :

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} (R_{ij}) \quad (6)$$

Keterangan :

K = Banyak kelompok yang dipakai

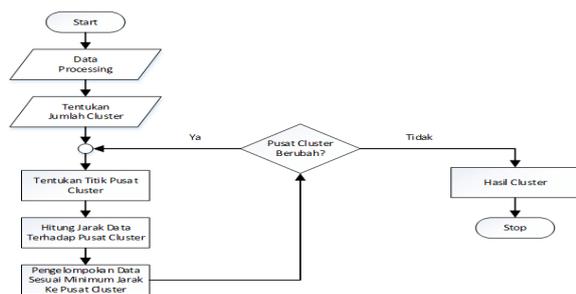
Rij = Rasio antar kelompok i dan kelompok j

Max = perolehan rasio terbesar antar kelompok

### 2.4. Flowchart K-Means

Flowchart merupakan diagram arahan yang melukiskan langkah-langkah untuk menyelesaikan masalah yang ada [18].

Flowchart K-Means dapat dilihat pada gambar 1 dibawah berikut ini :



Gambar 1. Flowchart K-Means

Gambar 1 diatas merupakan flowchart dari metode K-Means. Dimulai dari *start* lalu memproses data yang akan digunakan selanjutnya menentukan banyak *cluster* yang akan digunakan untuk mengelompokkan data, lalu menentukan titik pusat *cluster*, menghitung jarak data terhadap pusat *cluster*, mengelompokkan data sesuai dengan jarak yang paling kecil ke pusat *cluster*, apabila pusat *cluster* iterasi 1 dan iterasi 2 berbeda maka harus mencari iterasi ke 3 dengan cara mengulangi langkah-langkah sebelumnya mulai dari menentukan nilai *centroid* sampai mengelompokkan objek data, apabila pusat *cluster* dan posisi objek data

pada iterasi saat ini data iterasi sebelumnya sama, maka proses iterasi berhenti.

### 3. Hasil dan Pembahasan

Berikut pada Tabel 1. ini merupakan data asli yang dikelompokkan menggunakan algoritma k-means :

Tabel 1. Data Asli

	2015	2016	....	2020	2021	2022
Polisi Resort						
Nias	98	107	....	114	92	84
Mandailing Natal	97	93	....	84	77	58
Tapanuli Selatan	134	188	....	107	113	135
Tapanuli Tengah	66	107	....	129	129	141
Tapanuli Utara	105	134	....	134	142	154
Toba	98	92	....	89	105	141
Labuhanbatu	312	272	....	453	488	512
Asahan	390	453	....	332	295	339
Simalungun	385	252	....	349	253	277
Dairi	134	105	....	98	104	170
Karo	154	192	....	186	174	186
Deli Serdang	548	485	....	318	354	392
Langkat	344	358	....	294	260	262
Nias Selatan	13	26	....	35	28	29
Humbang Hasundutan	42	32	....	41	36	40
Pakpak Bharat	15	11	....	8	14	12
Samosir	25	38	....	55	44	75
Serdang Bedagai	330	297	....	289	282	300
Batubara	330	329	....	248	240	270
Padang Lawas	0	0	....	61	45	40
Sibolga	20	32	....	14	9	14
Tanjungbalai	68	43	....	69	57	51
Pematangsiantar	258	264	....	163	185	226
Tebing Tinggi	261	285	....	252	184	252
Binjai	195	186	....	280	266	299
Medan	1598	1574	....	1585	1345	1665
Padangsidempuan	62	58	....	35	39	32
Pel Belawan	314	263	....	261	256	309

Tabel 1 diatas merupakan data asli yang akan dilakukan pengelompokkan menggunakan metode K-Means.

Tabel 2. Centroid awal iterasi 1

Cluster	2015	2016	2017	2018	2019	2020	2021	2022
Cluster 1	1598	1574	1081	1212	1334	1585	1345	1665
Cluster 2	0	0	0	0	0	61	45	40
Cluster 3	195	186	163	176	253	280	266	299

Tabel 2 diatas merupakan nilai *centroid* awal iterasi ke 1 yang akan digunakan untuk menentukan pengelompokkan data berdasarkan jarak antara data terhadap *centroid*.

Tabel 3. Mengelompokkan Objek Pada Iterasi 1

	C1	C2	C3
Polisi Resort			
Nias		1	
Mandailing Natal		1	
Tapanuli Selatan			1
Tapanuli Tengah		1	
Tapanuli Utara			1

Polisi Resort	C1	C2	C3
Toba		1	
Labuhanbatu			1
Asahan			1
Simalungun			1
Dairi		1	
Karo			1
Deli Serdang			1
Langkat			1
Nias Selatan		1	
Humbang			1
Hasundutan		1	
Pakpak Bharat		1	
Samosir		1	
Serdang Bedagai			1
Batubara			1
Padang Lawas		1	
Sibolga		1	
Tanjungbalai		1	
Pematangsiantar			1
Tebing Tinggi			1
Binjai			1
Medan	1		
Padangsidempuan		1	
Pel Belawan			1

Tabel 3 diatas merupakan pengelompokkan objek data berdasarkan jarak antara centroid dan objek data yang paling dekat.

Tabel 4. Centroid iterasi 5

C	2015	2016	2017	2018	2019	2020	2021	2022
C 1	1598	1574	1081	1212	1334	1585	1345	1665
C 2	70,68	78,62	73,1	83,06	96,8	78,68	75,5	85,12
C 3	75	5	88	3	75	8		5
C 3	333,3	313,0	277,	313,5	330,	294,4	278,4	312,5
3	636	9	82	5	55	5	5	5

Tabel 4 diatas merupakan tabel centroid untuk iterasi terakhir yaitu iterasi ke 5.

Tabel 5. Mengelompokkan Objek Pada Iterasi 5

Polisi Resort	C1	C2	C3
Nias		1	
Mandailing Natal		1	
Tapanuli Selatan		1	
Tapanuli Tengah		1	
Tapanuli Utara		1	
Toba		1	
Labuhanbatu			1
Asahan			1
Simalungun			1
Dairi		1	
Karo		1	
Deli Serdang			1
Langkat			1
Nias Selatan		1	
Humbang			1
Hasundutan		1	

Polisi Resort	C1	C2	C3
Pakpak Bharat		1	
Samosir		1	
Serdang Bedagai			1
Batubara			1
Padang Lawas		1	
Sibolga		1	
Tanjungbalai		1	
Pematangsiantar			1
Tebing Tinggi			1
Binjai			1
Medan	1		
Padangsidempuan		1	
Pel Belawan			1

Tabel 5 diatas merupakan pengelompokkan data dari iterasi ke 5 yang mana proses iterasi berhenti sampai iterasi ke 5 dikarenakan posisi dan nilai centroid pada iterasi ke 4 dan iterasi ke 5 memperoleh hasil yang sama maka proses iterasi berhenti.

Untuk memperoleh nilai DBI (*Davies Bouldin Index*), diawali dengan memperoleh SSW (*Sum of Square within Cluster*), lalu mencari SSB (*Sum Of Square Between cluster*) dan mencari nilai Rasio.

- 1) Mencari nilai SSW (*Sum Of Square Between Cluster*) menggunakan rumus (3)

Menghitung nilai SSW untuk cluster 1

$$SSW_1 = \frac{1}{1} (0) = 0$$

Menghitung nilai SSW untuk cluster 2

$$SSW_2 = \frac{1}{16} (76,89+41,9155+226,227+153,226+170,502+76,6227+131,061+294,762+152,203+108,69+192,548+99,9289+189,996+171,726+81,1397+96,6335) = 141,505$$

Menghitung nilai SSW untuk cluster 3

$$SSW_3 = \frac{1}{11} (367,61+279,807+124,4+374,867+122,633+45,4635+133,243+242,932+148,666+271,955+176,074) = 207,968$$

- 2) Mencari nilai SSB (*Sum Of Square Between Cluster*)

Menghitung nilai SSB centroid 1 dan centroid 2

$$SSB_{1,2} = \sqrt{(1598 - 70,6875)^2 + (1574 - 78,625)^2 + (1081 - 73,188)^2 + (1212 - 83,063)^2 + (1334 - 96,875)^2 + (1585 - 78,688)^2 + (1345 - 75,5)^2 + (1665 - 85,125)^2} = 3842,67$$

Menghitung nilai SSB centroid 1 dengan centroid 3

$$SSB_{1,3} =$$

$$\sqrt{(1598 - 333,36)^2 + (1574 - 313,09)^2 + (1081 - 277,82)^2 + (1212 - 313,55)^2 + (1334 - 330,55)^2 + (1585 - 294,45)^2 + (1345 - 278,45)^2 + (1665 - 312,55)^2} = 3206,34$$

Menghitung nilai SSB centroid 2 dengan centroid 3  
 $SSB_{2,3} =$

$$\sqrt{(70,6875 - 333,36)^2 + (78,625 - 313,09)^2 + (73,188 - 277,82)^2 + (83,063 - 313,55)^2 + (96,875 - 330,55)^2 + (78,688 - 294,45)^2 + (278,45 - 278,45)^2 + (85,125 - 312,55)^2} = 642,68$$

Tabel 6 berikut merupakan matrik rasio yang diperoleh :

Tabel 6. Matrik SSB

Centroid	1	2	3	R Max
1	0	3842,67	3206,34	3206,34
2	3842,67	0	642,68	3842,67
3	3206,34	642,68	0	3206,34

Tabel 6 diatas merupakan tabel matrik rasio yang dibuat untuk memperoleh nilai maksimal DBI.

3) Mencari nilai Rasio

Menghitung nilai rasio cluster ke 1 dan cluster ke 2

$$R_{1,2} = \frac{0+141,505}{3842,67} = 0,03682$$

Menghitung nilai rasio cluster ke 1 dan cluster ke 3

$$R_{1,3} = \frac{0+207,968}{3206,34} = 0,06486$$

Menghitung nilai rasio cluster ke 2 dan cluster ke 3

$$R_{2,3} = \frac{141,505+207,968}{642,68} = 0,54377$$

Tabel 7 berikut merupakan matrik rasio yang diperoleh :

Tabel 7. Matrik Rasio

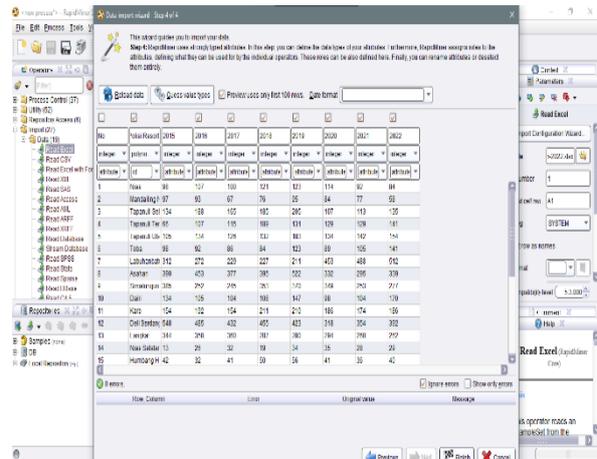
R	1	2	3	R Max
1	0	0,03682	0,06486	0,06486
2	0,03682	0	0,54377	0,54377
3	0,06486	0,54377	0	0,54377

Tabel 7 diatas merupakan tabel matrik rasio yang dibuat untuk memperoleh nilai maksimal rasio. Setelah memperoleh nilai Sum Of Square Within Cluster (SSW), Sum Of Square Between Cluster (SSB) dan Rasio maka dapat dilakukan pencarian nilai DBI.

4) Mencari nilai DBI

$$DBI = \frac{1}{3} (0,06486 + 0,54377 + 0,54377) = 0,38414$$

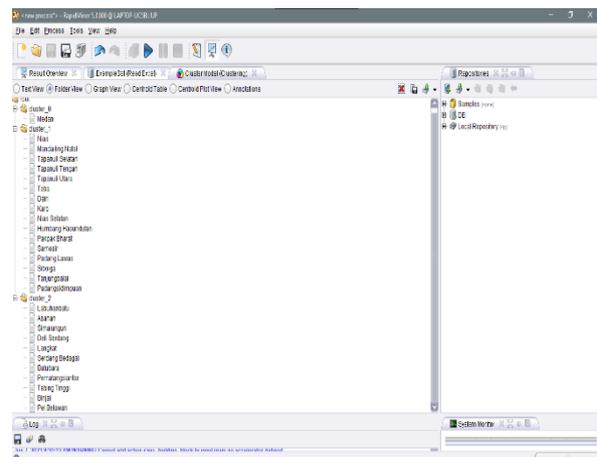
Berikut seperti pada Gambar 2 merupakan pemrosesan data dengan penggunaan aplikasi RapidMiner 5.3 dan hasil dari pengelompokan data.



Gambar 2. Pemrosesan data dengan penggunaan aplikasi RapidMiner 5.3 dengan algoritma K-Means

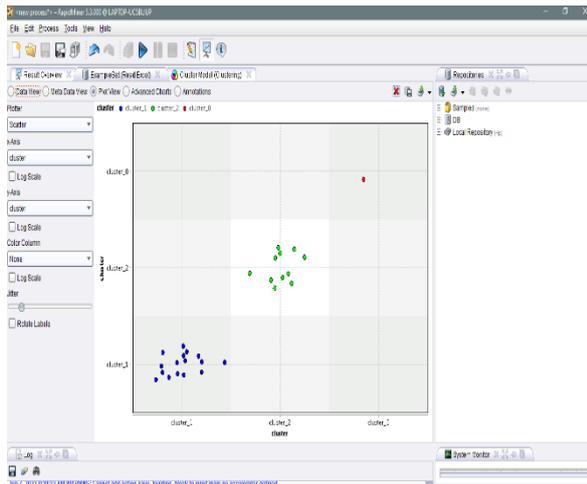
Pada Gambar 2 diatas penggunaan algoritma K-Means dalam mengelompokkan data terlebih dahulu dilakukan konfigurasi agar data yang akan dikelompokkan menjadi lebih akurat.

Hasil dari pengelompokkan dengan penggunaan algoritma K-Means pada gambar 3 berikut ini :



Gambar 3. Perolehan Dari Pengelompokkan dengan penggunaan Algoritma K-Means

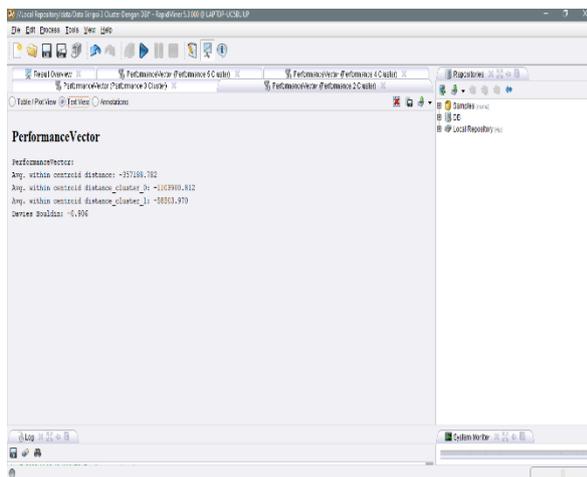
Dapat dilihat pada Gambar 3 hasil dari pengelompokkan yang menggunakan algoritma K-Means menghasilkan cluster 1 kecelakaan tinggi sebanyak 1 daerah, cluster 2 tingkat kecelakaan rendah sebanyak 16 daerah dan cluster 3 tingkat kecelakaan rendah sebanyak 11 daerah.



Gambar 4. Tampilan Cluster Berupa Plot View

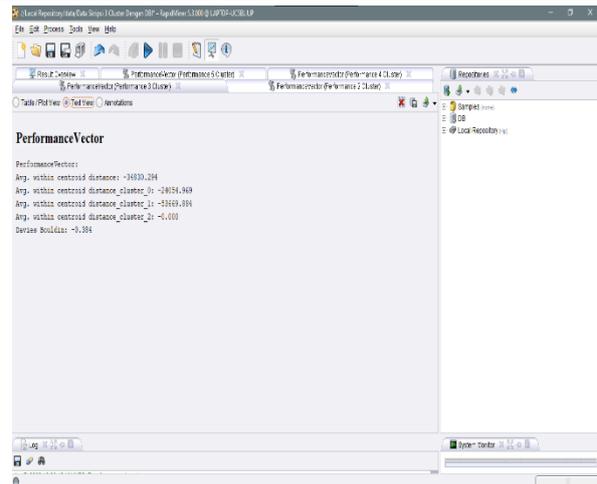
Dapat dilihat pada Gambar 4 merupakan tampilan sebaran data berdasarkan masing-masing *cluster*, pada titik berwarna merah merupakan anggota *cluster* 1 yang beranggotakan 1 daerah, titik warna hijau anggota dari *cluster* rendah dengan beranggotakan 11 daerah dan titik warna biru merupakan anggota dari *cluster* 3 yang beranggotakan 16 daerah.

Untuk mengetahui berapa banyak *cluster* yang dapat digunakan secara maksimal untuk melakukan pengelompokan data tersebut dengan mencari nilai DBI (*Davies Bouldin Index*), Tambah kecil nilai DBI (*Davies Bouldin Index*) yang didapat maka tambah optimal banyak kelompok yang dapat digunakan. Berikut merupakan nilai DBI yang diperoleh menggunakan aplikasi RapidMiner 5.3 sebagai berikut :



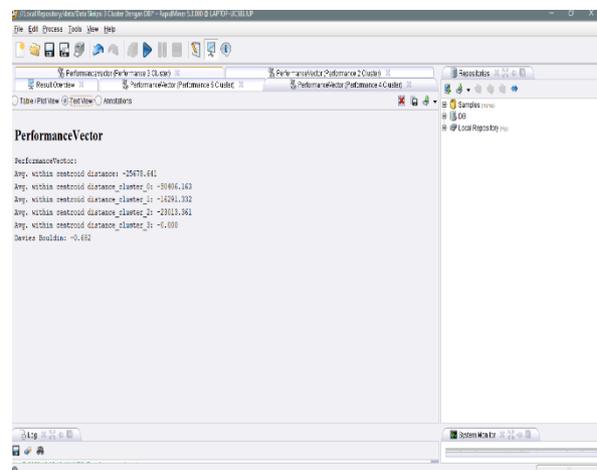
Gambar 5. Nilai DBI 2 Cluster

Dari Gambar 5 diatas bisa diperhatikan nilai DBI yang diperoleh untuk 2 kelompok sebesar 0,906.



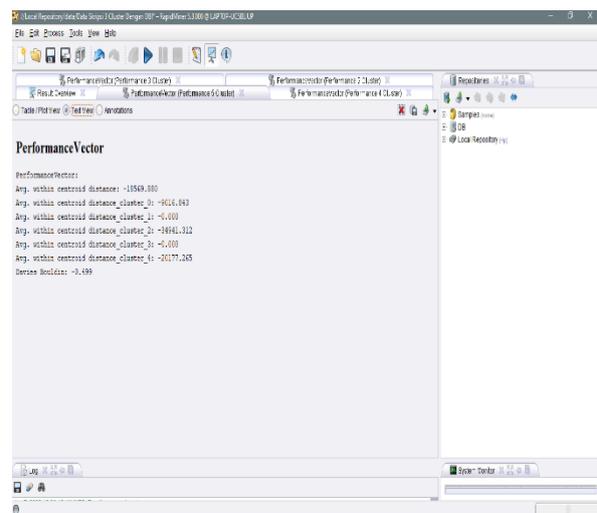
Gambar 6. Nilai DBI 3 cluster

Dari Gambar 6 diatas bisa diperhatikan nilai yang diperoleh untuk 3 kelompok sebesar 0,384.



Gambar 7. Nilai DBI 4 cluster

Dari Gambar 7 diatas bisa diperhatikan nilai yang diperoleh untuk 4 kelompok sebesar 0,602.



Gambar 8. Nilai DBI 5 cluster

Dari Gambar 8 diatas bisa diperhatikan nilai DBI untuk 5 kelompok sebesar 0,499. Banyak *cluster* yang

memiliki nilai DBI (*Davies Bouldin Index*) yang paling mendekati angka 0 merupakan banyak *cluster* yang paling cocok untuk digunakan dalam mengelompokkan data daerah rawan kecelakaan di Sumatera Utara tahun 2015-2022. Banyak *cluster* yang digunakan yaitu 3 kelompok dengan nilai DBI (*Davies Bouldin Index*) sebesar 0,384.

#### 4. Kesimpulan

Berdasarkan hasil dan pembahasan sebelumnya maka penulis dapat mengambil kesimpulan bahwa banyak *cluster* yang paling cocok untuk digunakan yaitu sebanyak 3 *cluster* dengan nilai DBI (*Davies Bouldin Index*) sebesar 0,385. Hasil pengelompokan yang diperoleh berupa *cluster* 1 dengan tingkat kecelakaan tinggi terdapat 1 daerah, *cluster* 2 dengan tingkat kecelakaan rendah terdapat 16 daerah dan *cluster* 3 dengan tingkat kecelakaan sedang terdapat 11 daerah.

Algoritma K-Means dapat diimplementasikan dengan baik dalam mengelompokkan daerah rawan kecelakaan di Sumatera Utara. Untuk penelitian selanjutnya diharapkan dapat menggunakan algoritma *clustering* yang lainnya sebagai perbandingan dengan algoritma K-Means.

#### Daftar Rujukan

- [1] N. F. Muto'in and A. Utami, "Analisis Tingkat Kecelakaan Lalu Lintas Menggunakan Metode Accident Rate Dan Equivalent Accident Number (EAN) Di Kota Magelang," *J. Rekayasa Sipil*, vol. 18, no. 1, p. 60, 2022, doi: 10.25077/jrs.18.1.60-67.2022.
- [2] Z. Siregar and I. Dewi, "Analisis Ruas Jalan Lintas Sumatera Kota Tebing Tinggi Dan Kisaran Sebagai Titik Rawan Kecelakaan Lalu Lintas," *J. MESIL (Mesin Elektro Sipil)*, vol. 1, no. 2, pp. 63–73, 2020, doi: 10.53695/jm.v1i2.88.
- [3] H. Hernawan, "ANALISIS Faktor PENYEBAB KECELAKAAN RUAS JALAN LIMBANGAN MALANGBONG KABUPATEN GARUT," *J. Tek. Sipil Cendekia*, vol. 3, no. 2, pp. 38–43, 2022, doi: 10.51988/jtsc.v3i2.53.
- [4] N. K. P. Widiastuti and I. M. S. Adiputra, "Gambaran Tingkat Pengetahuan Siswa tentang Pertolongan Pertama pada Kecelakaan di Sekolah Menengah Atas," *J. Akad. Baiturrahim Jambi*, vol. 11, no. 1, p. 23, 2022, doi: 10.36565/jab.v11i1.409.
- [5] Khairul Fahmi, "Faktor Penyebab Kecelakaan Lalu Lintas Dan Perilaku Berkendara Pada Siswa Sekolah Menengah Atas Di Pasir Pengaraian Riau," *J. Ilm. Cano Ekon.*, vol. 10, no. 1, pp. 1–10, 2021, doi: 10.30606/cano.v10i1.1084.
- [6] M. Fatkuroji, Fajrizal, Taslim, E. Sabna, and K. W. Ningsih, "Optimasi Nilai K Pada Algoritma K-Means Untuk Klasterisasi Data Pasien Covid-19," *Indones. J. Comput. Sci.*, vol. 11, no. 1, pp. 697–707, 2022.
- [7] Kurniawan Titus & Mohamad Jajuli, "Clustering Data Kecelakaan Lalu Lintas di Kecamatan Cileungsi Menggunakan Metode K-Means," *Gener. J.*, vol. 6, no. 1, pp. 1–12, 2022, doi: 10.29407/gj.v6i1.16103.
- [8] D. H. Devi Gultom, Hotma Dame Tampubolon, Luvita Yolanda Hutabarat, Fikrul Ilmi R H Zer, "Penerapan Algoritma K-Means Untuk Mengetahui Tingkat Tindak Kejahatan Daerah Pematangsiantar," *J. Teknol. Inf.*, vol. 4, no. 1, pp. 146–151, 2020, doi: 10.36294/jurti.v4i1.1263.
- [9] W. T. Saputro, M. Murhadi, and H. M. Jumasa, "Menemukan Pola Sebaran Vaksinasi Data Covid-19 di Indonesia Menggunakan Algoritma K-Means," *J.*

- Fasilkom*, vol. 13, no. 02, pp. 244–250, 2023, doi: 10.37859/jf.v13i02.5551.
- [10] M. Mughnyanti, S. Efendi, and M. Zarlis, "Analysis of determining centroid clustering x-means algorithm with davies-bouldin index evaluation," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 725, no. 1, 2020, doi: 10.1088/1757-899X/725/1/012128.
- [11] W. R. Fadillah *et al.*, "Implementasi Data Mining C4.5 Dalam Mengukur Tingkat Kepuasan Mahasiswa Terhadap Kinerja Asisten Laboratorium Komputer," *Pros. Semin. Nas. Ris. Dan Inf. Sci.*, vol. 2, pp. 403–414, 2020.
- [12] I. Rosyadi, F. A. Artanto, S. E. Rahmawati, and H. T. B. J. Pangestu, "Decision Tree Dalam Analisis Keputusan Pembelian Program Pada Perkumpulan Penggiat Programmer Indonesia," *J. Fasilkom*, vol. 12, no. 3, pp. 141–144, 2022, doi: 10.37859/jf.v12i3.3948.
- [13] Hidayatunnisa, Kusriani, and Kusnawi, "Perbandingan Kinerja Metode Naive Bayes dan Support Vector Machine dalam Analisis Soal," *J. Fasilkom*, vol. 13, no. 2, pp. 173–180, 2023.
- [14] A. Aditya, I. Jovian, and B. N. Sari, "Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019," *J. Media Inform. Budidarma*, vol. 4, no. 1, p. 51, 2020, doi: 10.30865/mib.v4i1.1784.
- [15] N. Zulfa, R. I. Auliya, and A. Zaenal, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [16] M. Fikri, "PENGELOMPOKKAN KABUPATEN / KOTA DI JAWA TIMUR BERDASARKAN BANYAKNYA SEKOLAH MENGGUNAKAN ALGORITMA PARTITIONING AROUND MEDOIDS (PAM)," *J. Apl. Sist. Inf. Dan Elektron.*, vol. 2, no. 2, pp. 99–106, 2020.
- [17] M. Wahyudi, L. Pujiastuti, and Solikhun, "Komparasi K-Means Clustering dan K-Medoids dalam Mengelompokkan Produksi Susu Segar di Indonesia Comparison of K-Means Clustering and K-Medoids in Clustering Fresh Milk Production in Indonesia," *J. Bumigora Inf. Technol.*, vol. 4, no. 2, pp. 243–254, 2022, doi: 10.30812/bite.v4i2.2104.
- [18] R. V. Siregar, E. Christian, A. C. Saputra, J. Y. Sudarso, and K. Tengah, "Rancang Bangun Media Promosi Oleh-Oleh Khas Sampit Berbasis Website," *JOINTECOMS (Journal Inf. Technol. Comput. Sci. p-ISSN 2798-284X)*, vol. 1, no. 2, pp. 2798–3862, 2021.