

## **Penerapan *K-Means Clustering* dalam Pemetaan Titik Kecelakaan Lalu Lintas di Kecamatan Bawen**

**Christian Alessandro Doke<sup>1\*</sup>, Hanna Prillysca Chernovita<sup>2</sup>**  
 Sistem Informasi, Universitas Kristen Satya Wacana, Salatiga, Indonesia  
 \*e-mail *Corresponding Author*: 682019137@student.uksw.edu

### **Abstract**

*With the high traffic flow in Bawen sub-district, traffic accidents are one of the unavoidable events. With so many accident cases in Bawen sub-district, the use of accident data in identifying areas based on traffic accident vulnerability is still very minimal. Therefore, this study was conducted to group and map the Bawen sub-district area based on the level of vulnerability to traffic accidents using the k-means clustering algorithm. The results of this study divide the areas in Bawen sub-district into three clusters, namely very accident-prone, accident-prone and somewhat accident-prone. To validate the use of K-means clustering, the Davies Boulder index method was used. From the use of this method, it is found that the use of K-means clustering with the number of 3 clusters is more optimal than the number of other clusters. With this research, it is hoped that it can help the local government and police in making policies for driving in the area.*

**Keywords:** *Accidents hotspots; Geographic Information System; K-Means Clustering*

### **Abstrak**

Dengan tingginya arus lalu lintas pada kecamatan Bawen, kecelakaan lalu lintas adalah salah satu kejadian yang tidak dapat dipisahkan. Dengan banyaknya kasus kecelakaan pada kecamatan bawen, penggunaan data kecelakaan dalam melakukan identifikasi daerah berdasarkan kerawanan kecelakaan lalu lintas masih sangat minim. Oleh karena itu, penelitian ini dilakukan untuk melakukan pengelompokan dan pemetaan daerah kecamatan Bawen berdasarkan tingkat rawan kecelakaan lalu lintas menggunakan algoritma *k-means clustering*. Hasil dari penelitian ini membagi daerah-daerah pada kecamatan Bawen menjadi tiga klaster, yaitu sangat rawan kecelakaan, rawan kecelakaan dan agak rawan kecelakaan. Untuk memvalidasi penggunaan *K-means Clustering*, digunakan metode *Davies Boulder index*. Dari penggunaan metode ini, didapatkan hasil bahwa penggunaan *K-means Clustering* dengan jumlah 3 klaster lebih optimal dibandingkan jumlah klaster yang lainnya. Dengan adanya penelitian ini, diharapkan dapat membantu pemerintah dan kepolisian setempat dalam membuat kebijakan dalam berkendara pada area tersebut

**Kata Kunci:** *K-Means Clustering; Sistem Informasi Geografis; Titik rawan kecelakaan*

### **1. Pendahuluan**

Keselamatan dalam berlalu lintas adalah salah satu faktor prioritas dalam manajemen transportasi, khususnya dalam daerah dengan kepadatan arus lalu lintas yang sangat tinggi. Kecelakaan lalu lintas merupakan salah satu penyumbang kerugian secara materi, kecelakaan juga memberi dampak buruk yang besar terhadap aktivitas sosial seperti gangguan terhadap aktivitas Masyarakat dan juga korban jiwa. Dalam menangani kasus kecelakaan lalu lintas, dibutuhkan pemahaman yang mendalam terhadap factor-faktor penyebab kecelakaan lalu lintas, salah satu faktor yang berpengaruh dalam melakukan penanganan kecelakaan lalu lintas adalah Lokasi terjadinya kejadian kecelakaan lalu lintas. Oleh karena itu, pemetaan titik rawan kecelakaan berdasarkan data sangat berguna dalam mendukung perencanaan kebijakan yang lebih efisien.

Sebagai kecamatan dengan Lokasi yang strategis, kecamatan Bawen yang terletak pada kabupaten Semarang, dimana kecamatan Bawen menghubungkan beberapa kota/kabupaten memiliki resiko besar dalam kecelakaan lalu lintas. Hal ini disebabkan oleh padatnya alur lalu lintas pada kecamatan bawen. Selain itu, kondisi geografis kecamatan

bawen yang penuh jalan berkelok, tanjakan yang panjang dan juga kondisi jalan yang berbeda-beda pada tiap ruas jalannya menjadi salah satu faktor yang menyebabkan meningkatnya resiko kejadian kecelakaan lalu lintas pada kecamatan Bawen. Berdasarkan laporan dan data kecelakaan lalu lintas yang ada, terdapat titik-titik pada kecamatan bawen yang menjadi Lokasi rawan kecelakaan lalu lintas. Namun, Upaya dalam mencari pola distribusi lokasi titik kecelakaan lalu lintas masih sangat terbatas pada metode manual, dimana metode ini sering mendapatkan hasil yang tidak akurat dan penerapannya yang sangat terbatas. Hal ini mengakibatkan upaya dalam melakukan mitigasi tidak dapat dilakukan sepenuhnya.

Dalam mengatasi masalah ini, penggunaan pendekatan *K-means Clustering* dianggap sebagai salah satu jalur yang efektif. Terdapat beberapa penelitian yang dapat membuktikan keefektifan penggunaan *K-Means Clustering* dalam pemetaan titik rawan kecelakaan. Berdasarkan penelitian Herry Wahyono, dkk pada tahun 2024 bahwa *k-means clustering* sangat berperan penting dalam mengidentifikasi Tingkat kerawanan kecelakaan suatu daerah berdasarkan jumlah kejadian yang terjadi[1]. Selanjutnya dari penelitian Diah Puspitasari, dkk pada tahun 2020 juga menyatakan hal yang sama. Bahwa dengan penggunaan *K-Means Clustering* dalam pemetaan titik-titik rawan kecelakaan sangat cocok dalam mengetahui Tingkat kerentanan jalur lalu lintas pada suatu daerah[2]. Penelitian terakhir berasal dari Christopher Sinclair pada tahun 2021 menyatakan bahwa penggunaan *K-means clustering* sangat cocok digunakan dalam analisis data kecelakaan, terkhususnya jika *dataset* kecelakaannya sangat besar. Penggunaan *K-means Clustering* juga dapat membantu mendukung bukti penyebab kecelakaan lalu lintas yang sering terjadi di daerah yang diteliti[3].

Penulisan penelitian ini bertujuan untuk melakukan pemetaan terhadap titik-titik rawan kecelakaan pada kecamatan bawen dengan menerapkan metode *K-means Clustering*. Dengan penggunaan *K-means Clustering*, proses pengidentifikasian kelompok Lokasi yang rentan kecelakaan berdasarkan data lebih mudah untuk dilakukan. Hasil dari penelitian ini diharapkan dapat berdistribusi dalam pengambilan keputusan dan pembuatan kebijakan dan perimeter yang dilakukan oleh kepolisian, pemerintah kecamatan Bawen maupun pihak-pihak yang ikut serta dalamnya. Selain itu, dengan adanya penelitian ini, diharapkan dapat membantu dalam bidang transportasi dan keselamatan lalu lintas dalam hal ilmu data.

## 2. Tinjauan Pustaka

Penelitian yang menjadi sumber peninjauan untuk penelitian ini diambil dari beberapa sumber. Penelitian pertama diambil dari hasil penelitian milik Bias Arbi Fauzan, dkk yang berjudul *Implementation of K-means Clustering algorithm for grouping traffics violation levels in Siak*. Dalam penelitian ini, mereka menggunakan *k-means clustering* dengan tujuan untuk mendapatkan informasi mengenai kecelakaan lalu lintas dengan mengelompokkannya ke dalam 4 klaster berdasarkan waktu kejadian[4].

Selanjutnya, mengikuti kajian dari hasil penelitian milik Hussien Kamh, dkk yang berjudul *Exploring road traffics accident hotspots using clustering algorithms and gis-based spatial analysis*. Pada penelitian ini, peneliti dalam mengidentifikasi dan mempelajari pola, dan melakukan pemetaan titik rawan kecelakaan pada kota Najran, Arab Saudi maka digunakan *clustering algorithm*. Teknik klasterisasi yang digunakan dalam penelitian ini sebanyak 5 model, yaitu *DBSCAN*, *Hierarchical Agglomerative Clustering*, *GIS-based density analysis*, *proximity analysis* dan *spatial interpolation*[5].

Penelitian berikutnya yang diambil menjadi sumber kajian adalah hasil penelitian milik Aryo Pratama, dkk yang berjudul *Implementation of K-Means Clustering in recognizing crimes and traffic issues through GIS*. Pada penelitian ini, dalam melakukan pemetaan berdasarkan aktivitas kriminal dan gangguan lalu lintas ( kemacetan dan kecelakaan lalu lintas ) pada distrik batu bara, Sumatera utara maka digunakan model *K-means Clustering* dan hasilnya dikategorikan menjadi tiga klaster dimana, C1 merupakan area yang jarang rawan, C2 merupakan area rawan dan C3 merupakan area sangat rawan[6].

Penelitian berikutnya diambil dari hasil penelitian milik Khopipah Parawansah Siregar, dkk dengan judul penelitian *Implementation of purity K-means algorithm in accident data clustering in north padang lawas district*. Pada penelitian ini, peneliti menggunakan *K-means Clustering* untuk mempelajari dan mengelompokkan daerah-daerah pada kabupaten Padang lawas utara berdasarkan tingkat kecelakaan lalu lintas dan dalam menentukan Tingkat kecelakaan pada daerah-daerah didalam wilayah penelitian, data-data seperti jumlah kecelakaan, korban jiwa, jumlah korban luka berat dan lingkungan menjadi titik acuan[7].

Penelitian berikutnya adalah hasil penelitian dari Wan Fairoz Wan Yacoob, dkk berjudul *Spatio-Temporal clustering of road accidents in Kelantan, Malaysia*. dalam melakukan analisa dan pemetaan terhadap titik–titik rawan kecelakaan pada Kelantan, Malaysia para peneliti menggunakan metode *K-means Clustering* untuk membagi daerah–daerah berdasarkan intensitas kecelakaan lalu lintas yang terjadi. Data yang digunakan diambil dari data kecelakaan dari tanggal 1 januari, 2019 hingga 11 februari, 2019 dengan area cakupan 10 distrik didalam wilayah Kelantan, Malaysia. Jenis–jenis data yang diambil sebagai acuan dalam menentukan Tingkat kerawanan kecelakaan lalu lintas sebuah distrik adalah jumlah korban kecelakaan, waktu kecelakaan terjadi, Lokasi terjadinya kecelakaan, longitude dan latitude[8].

Penelitian selanjutnya adalah hasil penelitian dari Lulu Lutfi Latifah, Sahid Agustian Hudjimartu dan Iksal Hanuarsyah. Pada penelitian, penulis mengungkapkan keresahannya mengenai minimnya informasi area rawan kecelakaan di kota Bogor yang diakibatkan oleh ketidakefektifan pemanfaatan data kecelakaan lalu lintas. Oleh karena itu, perlu adanya pembuatan webGIS menggunakan algoritma *k-means clustering* untuk membagi titik kecelakaan berdasarkan daerah daerah di bogor berdasarkan kerawanan dalam kecelakaan lalu lintas. Dalam penelitian ini, para peneliti membagi Tingkat kecelakaan lalu lintas menjadi 3 klaster, yaitu daerah cukup aman, daerah rawan dan daerah sangat rawan[9].

Berdasarkan penelitian dari Nurul Rahmadani, dkk berjudul *K-means clustering areas prone to traffic accidents in Asahan regency*. Dalam penelitian ini, para peneliti menggunakan *K-means Clustering* untuk membagi 25 kelurahan/kecamatan pada kabupaten asahan berdasarkan Tingkat kecelakaan yang terjadi dengan acuan data dari data kecelakaan lalu lintas pada tahun 2015–2018 pada kabupaten Asahan, dimana data tersebut tersusun dari jumlah kecelakaan, jumlah korban jiwa, jumlah korban luka berat dan luka ringan[10].

Dari penelitian milik Arninda Agnes Veranda, dkk berjudul Penerapan metode k-means clustering untuk pemetaan daerah rawan kecelakaan lalu lintas di kota malang berbasis website. Peneliti menggunakan *k-means clustering* untuk melakukan pengelompokkan terhadap lima kecamatan di kota malang dan enam ruas jalan pada setiap kecamatan di kota malang dan hasil akhir dari penelitian ini berupa pembagian lima kecamatan kedalam tiga klaster, dimana klaster ini mewakili kerawanan suatu daerah atas kasus kecalakaan lalu lintas dan dari hasil klasterisasi ini dijadikan kedalam sebuah peta *online* yang diwujudkan dengan penggunaan QGIS untuk pembuatan petanya dan untuk websitenya menggunakan *framework* CSS Bootstrap dan MySQL sebagai databasenya[11].

Selanjutnya dari penelitian milik Valdi Irawan, dkk dengan judul Penerapan algoritma *k-mean clustering* pada kecelakaan lalu lintas berdasarkan kabupaten/kota di Provinsi Jawa Tengah tahun 2020. Penelitian ini membahas mengenai penerapan *K-means Clustering* dalam memperoleh informasi pembagian kelompok rawan kecelakaan pada tiap kabupaten/kota berdasarkan jumlah kejadian kecelakaan, jumlah korban jiwa, jumlah korban luka berat dan ringan. Hasil akhir dari penelitian ini adalah pembagian daerah kedalam 3 klaster yaitu klaster pertama mewakili daerah dengan Tingkat kecelakaan rendah, klaster kedua mewakili Tingkat kecelakaan sedang dan klaster ketiga mewakili Tingkat kecelakaan tinggi [12].

Pada akhirnya dapat ditarik Kesimpulan dari sebelas penelitian yang dipaparkan sebelumnya bahwa penggunaan *K-means Clustering* sangat mempermudah untuk melakukan pengelompokkan data, terutama melakukan pengelompokkan daerah berdasarkan Tingkat rawan kecelakaan. Pada penelitian ini, yang membuat berbeda dari kesebelas hasil penelitian sebelumnya adalah daerah yang dijadikan studi kasus, dimana dalam penelitian ini adalah kecamatan bawen, dimana daerah ini belum pernah dijadikan studi kasus sebelumnya. Selain daerah penelitian yang jarang disorot sebagai area penelitian, hasil dari penelitian ini dirancang untuk dijadikan sebagai dasar untuk pemerintah dan kepolisian setempat dalam pembuatan kebijakan dalam keselamatan berkendara di kecamatan Bawen.

### 3. Metodologi

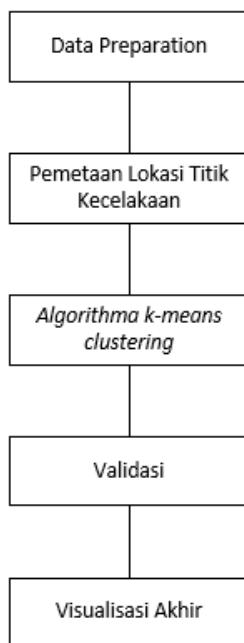
#### 3.1. Jenis dan Sumber Data

Dalam penelitian ini, data yang digunakan bersifat kuantitatif. Data yang digunakan sebagai syarat pembuatan klaster dalam *K-Mean Clustering* adalah data jumlah kejadian kecelakaan, jumlah korban meninggal dunia, jumlah korban luka berat dan jumlah korban luka ringan dan dari data ini, klaster dibagi menjadi 3 yaitu Sangat Rawan Kecelakaan (C1), Rawan Kecelakaan (C2) dan Agak Rawan Kecelakaan (C3). Data yang dijadikan acuan dalam

penelitian ini adalah data rekapitulasi kecelakaan lalu lintas tahun 2021–2023 pada kecamatan Bawen, kabupaten Semarang sebanyak 184 atribut.

### 3.2 Teknik Pengolahan Data

Berikut merupakan *flowchart* yang menggambarkan proses metode penelitian pada penelitian ini. Untuk penjelasan mengenai *flowchart* akan dipaparkan setelah *flowchart*.



Gambar 1. *Flowchart* Metode Penelitian

#### 1) **Data Preparation**

Pada tahap ini, data kecelakaan lalu lintas yang didapatkan dari Dinas Perhubungan Kabupaten Semarang dan Satlantas Kabupaten Semarang masih dalam cakupan kabupaten Semarang secara keseluruhan. Oleh karena itu perlu adanya *data selection* dan *data cleaning*. Hal ini dilakukan dalam rangka mengerucutkan data yang diperlukan kedalam jangkauan kecamatan Bawen. Selain menseleksi data berdasarkan kecamatan bawen, data ini disertakan dengan data koordinat. *data cleaning* diperlukan karena koordinat yang diberikan mempunyai kemungkinan tidak masuk kedalam area penilitian dan juga adanya salah input berdasarkan desa atau kelurahan tempat kejadian kecelakaan.

#### 2) **Pemetaan Lokasi Titik Kecelakaan**

Pada tahap ini, data koordinat yang sudah diolah akan dimasukkan kedalam *tools* Sistem Informasi Geografis (dalam penelitian ini menggunakan *QGIS*). Setelah memasukkan data koordinat kedalam *QGIS*, maka *output* yang akan dihasilkan akan ditampilkan dalam bentuk peta.

#### 3) **Algoritma K-Means Clustering**

*K-Means Clustering* adalah teknik data mining atau cara dalam menganalisis sekumpulan data yang menerapkan proses pemodelan tanpa *supervise* (*unsupervised*) dan *K-means* melakukan pengelompokan data dengan sistem partisi[13]. Tujuan dari *K-means clustering* adalah untuk memaksimalkan kemiripan antar data didalam sebuah klaster dan mengurangi kemiripan data antar klaster data dengan cara mengelompokkan data–data tersebut ke tiap klaster berdasarkan titik pusatnya[14]

Berikut adalah langkah–langkah dalam menghasilkan *K-Means Clustering*:

- 1) Menentukan jumlah *cluster* yang diinginkan dengan cara menghitung nilai *k*
- 2) Nilai *k* diinisiasikan sebagai pusat *cluster*
- 3) Untuk menghitung jarak antara masing–masing *centroid* dalam setiap data, gunakan persamaan *Euclidian Distance* dengan rumus sebagai berikut:

$$D(i, j) = \sqrt{\sum (X_{ki} - X_{kj})^2} \dots\dots\dots 1$$

Dimana:

- $D(i, j)$  = jarak geometris dari data  $i$  ke pusat kelompok  $j$
- $X_{ki}$  = objek atau data ke  $i$  pada atribut  $k$
- $X_{kj}$  = centroid atau titik pusat data  $j$  pada atribut  $k$

- 4) Kelompokkan setiap data menurut jarak terdekat antara *centroid* dan data lainnya
- 5) Tentukan posisi baru ( $k$ ) / *centroid* rata-rata di setiap kelompok
- 6) Jika hasil *centroid* baru dan *centroid* lama tidak sama, Kembali ke Langkah 3. Namun, jika hasil *centroid* baru dan *centroid* lama sama, maka perhitungan dihentikan[11]

#### 4) Validasi

Setelah melakukan penghitungan menggunakan algoritma *k-means clustering*, data akhir akan divalidasi. Dalam penelitian ini, metode validasi yang digunakan adalah metode *Davies Bouldin Index*. *Davies Bouldin Index* atau DBI adalah sebuah metode pengukuran yang digunakan dalam melakukan evaluasi terhadap kinerja pengelompokan[15]. Dalam pengukuran *Davies bouldin index*, semakin kecil hasil perbandingan yang didapatkan maka semakin optimal model yang digunakan pada *K-means Clustering*[16].

#### 5) Visualisasi Data

Setelah melakukan perhitungan menggunakan algoritma *k-means clustering*, data-data yang telah dibagi ke dalam klaster-klaster akan dimasukkan lagi ke dalam *QGIS* untuk melakukan pemetaan berdasarkan klaster yang telah ditentukan

### 4. Hasil dan Pembahasan

#### 4.1 Data Preparation

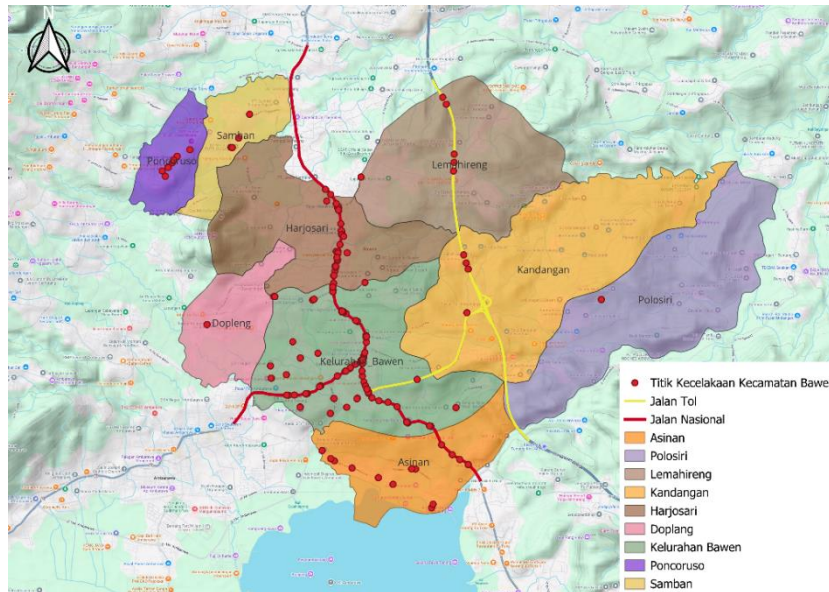
Berdasarkan data kecelakaan dari tahun 2021–2023 pada kecamatan Bawen, didapati bahwa dalam rentang waktu tersebut telah terjadi 184 kecelakaan lalu lintas yang terjadi di kecamatan Bawen. Data yang didapatkan dapat dilihat pada tabel 1 dibawah ini

Tabel 1. Data Kecelakaan Lalu Lintas Tahun 2021 – 2023 di Kecamatan Bawen

No	Kode Kecelakaan	Latitude	Longitude	Desa/Kelurahan	Meninggal Dunia	Luka Berat	Luka Ringan
1	Bwn_001	-7.240379	110.435217	Bawen	0	0	1
2	Bwn_002	-7.224396	110.431261	Harjosari	3	0	2
3	Bwn_003	-7.248004	110.428055	Bawen	0	0	1
4	Bwn_004	-7.23413	110.420667	Bawen	0	0	1
5	Bwn_005	-7.227665	110.450474	Kandangan	2	0	0
6	Bwn_006	-7.234153	110.420836	Bawen	0	0	1
7	Bwn_007	-7.228569	110.430641	Harjosari	0	0	2
8	Bwn_008	-7.236485	110.431418	Bawen	1	0	0
.	.	.	.	.	.	.	.
182	Bwn_182	-7.219705	110.430463	Harjosari	0	0	1
183	Bwn_183	-7.236421	110.431386	Bawen	0	0	1
184	Bwn_184	-7.229872	110.451167	Lemahireng	0	0	5

#### 4.2 Pemetaan Titik Lokasi

Berdasarkan tabel 1, dibuatlah peta persebaran titik kecelakaan lalu linta mengikuti data latitude dan longitude setiap kecelakaan yang lalu lintas di dalam tabel 1. Gambar 2 adalah hasil dari pemetaan menggunakan data pada tabel 1.



Gambar 2. Peta Titik Kecelakaan Kecamatan Bawen Tahun 2021 – 2023

### 4.3 Implementasi algoritma *K-means Clustering*

Setelah itu, data dari tabel 1 akan dikelompokkan berdasarkan desa atau kelurahan tempat kecelakaan lalu lintas terjadi sebelum melakukan *K-means Clustering*. Data dibagi menjadi sembilan bagian berdasarkan desa dan kelurahan yang ada di kecamatan Bawen. Untuk pengelompokan data dapat dilihat pada Tabel 2.

Tabel 2. Pengelompokan data berdasarkan Desa/Kelurahan

No	Desa/Kelurahan	Jumlah Kejadian	Meninggal Dunia	Luka Berat	Luka Ringan
1	Asinan	14	4	7	15
2	Bawen	102	33	14	110
3	Dopleng	1	0	0	1
4	Harjosari	44	16	5	48
5	Kandangan	3	3	0	1
6	Lemahireng	7	2	3	18
7	Polosiri	1	0	0	1
8	Poncoruso	8	1	1	9
9	Samban	4	0	2	3

Berdasarkan dari Tabel 2, akan dilakukan penghitungan menggunakan *K-means Clustering*. Langkah pertama yang akan menentukan jumlah *cluster* dan nilai *centroid* dan pada langkah ini telah ditentukan akan menggunakan 3 *cluster*, dan data yang akan digunakan sebagai nilai *centroid* pada iterasi 1 adalah data ke-2 sebagai sangat rawan kecelakaan, data ke-4 sebagai rawan kecelakaan dan data ke-6 sebagai aman.

Tabel 3. *Centroid* Awal

No	Cluster	Jumlah Kejadian	Meninggal Dunia	Luka Berat	Luka Ringan
1	C1	102	33	14	110
2	C2	44	16	5	48
3	C3	7	2	3	18

Setelah menentukan jumlah dan nilai pada tiap *cluster* yang akan digunakan, langkah berikut yang akan dilakukan adalah menghitung jarak tiap data yang ada pada tabel 2 menuju centroid yang telah ditentukan. Dalam menghitung jarak, akan menggunakan rumus *Euclidian Distance*. Data ke-1 akan dihitung jaraknya menggunakan rumus *Euclidian Distance*.

Jarak pusat cluster 1 :

$$D_{11} = \sqrt{(14-102)^2 + (4-33)^2 + (7-14)^2 + (15-14)^2} = 132.8871702$$

Jarak pusat cluster 2 :

$$D_{12} = \sqrt{(14-44)^2 + (4-1)^2 + (7-5)^2 + (15-21)^2} = 30.8058436$$

Jarak Pusat Cluster 3:

$$D_{13} = \sqrt{(14-7)^2 + (4-2)^2 + (7-3)^2 + (15-18)^2} = 8.831760866$$

Dengan penggunaan rumus *Euclidian Distance* pada data ke-1, didapatkan hasil 132.8871702 untuk C1, 30.8058436 untuk C2 dan 8.831760866 untuk C3, lalu dari perhitungan jarak untuk ketiga centroid tersebut, data akan dikelompokkan kedalam ketiga *Cluster* berdasarkan hasil dengan nilai terkecil dan data ke-1 termasuk kedalam kelompok C3 atau daerah aman kecelakaan. Untuk hasil perhitungan untuk seluruh data akan ditampilkan pada tabel 4.

Tabel 4. Hasil perhitungan iterasi 1

Desa/Kelurahan	C1	C2	C3	Nilai Terdekat	Cluster
Asinan	132.8871702	46.22769733	8.831760866	8.831760866	3
Bawen	0	87.05170877	136.2754563	0	1
Dopleng	152.8626835	65.87108622	18.38477631	18.38477631	3
Harjosari	87.05170877	0	49.689033	0	2
Kandangan	150.9238218	63.90618123	17.74823935	17.74823935	3
Lemahireng	136.2754563	49.689033	0	0	3
Polosiri	152.8626835	65.87108622	18.38477631	18.38477631	3
Poncoruso	142.2322045	55.29918625	9.327379053	9.327379053	3
Samban	149.2849624	62.36986452	15.45962483	15.45962483	3

Setelah melakukan perhitungan iterasi 1, maka akan dilanjutkan ke perhitungan iterasi 2. Perhitungan ini berfungsi untuk mengetahui apakah ada perubahan *centroid* pada iterasi 1 dan iterasi 2. Perhitungan iterasi akan dilakukan berulang, jika hasil *centroid* pada iterasi sebelum dan sesudahnya ada perubahan dan akan berhenti disaat tidak ada perubahan *centroid* lagi.

Sebelum melakukan perhitungan untuk iterasi 2, centroid baru akan ditentukan untuk perhitungan pada iterasi kedua. Untuk centroid baru akan ditampilkan pada tabel 5.

Tabel 5. Nilai *Centroid* Iterasi 2

<i>Centroid</i>	Jumlah Kejadian	Meninggal Dunia	Luka Berat	Luka Ringan
C1	102	33	14	110
C2	44	16	5	48
C3	5.428571429	1.428571429	1.857142857	6.857142857

Setelah menentukan centroid baru untuk iterasi 2, perhitungan pada iterasi 2 sama seperti perhitungan pada iterasi 1. Hasil dari perhitungan iterasi 2 akan ditampilkan pada Tabel 6.

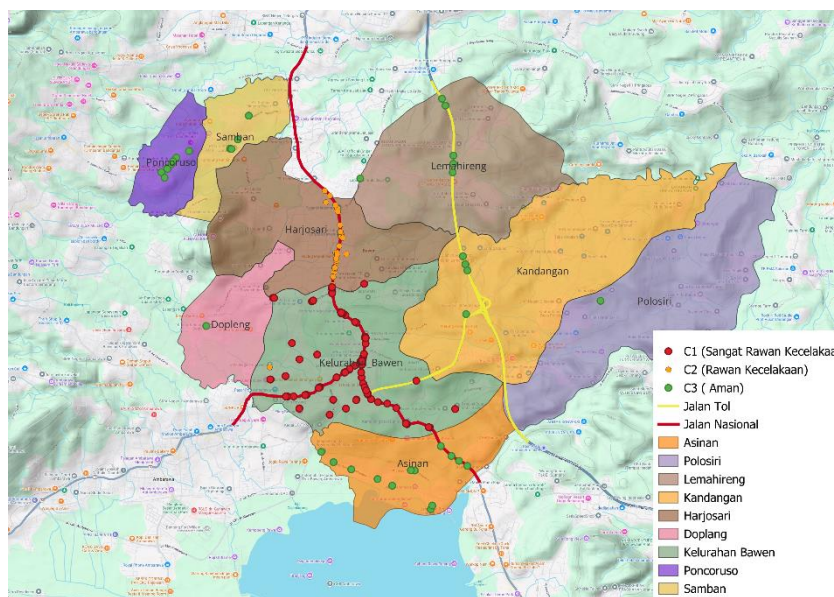
Tabel 6. Hasil Perhitungan Iterasi 2

No	Desa/Kelurahan	C1	C2	C3	Nilai Terdekat	Cluster
1	Asinan	132.8871702	46.22769733	13.14673856	13.14673856	3
2	Bawen	0	87.05170877	145.2883129	0	1
3	Dopleng	152.8626835	65.87108622	7.70766912	7.70766912	3
4	Harjosari	87.05170877	0	58.33261418	0	2
5	Kandangan	150.9238218	63.90618123	6.791351042	6.791351042	3
6	Lemahireng	136.2754563	49.689033	11.32542741	11.32542741	3
7	Polosiri	152.8626835	65.87108622	7.70766912	7.70766912	3
8	Poncoruso	142.2322045	55.29918625	3.481730745	3.481730745	3
9	Samban	149.2849624	62.36986452	4.356557338	4.356557338	3

Setelah mendapatkan hasil perhitungan untuk iterasi 2, maka akan dilakukan perbandingan hasil centroid pada iterasi 1 dan iterasi 2. Pada iterasi 2, tidak terjadi perubahan centroid sehingga perhitungan pun dihentikan.

Pada iterasi 2, data ke-2 masuk kedalam C1 (sangat rawan kecelakaan), data ke-4 masuk kedalam C2 (Rawan Kecelakaan) dan data ke-1,3,5,6,7,8,dan 9 masuk kedalam C3(Agak rawan kecelakaan).

#### 4.4 Visualisasi hasil K-means clustering kedalam peta



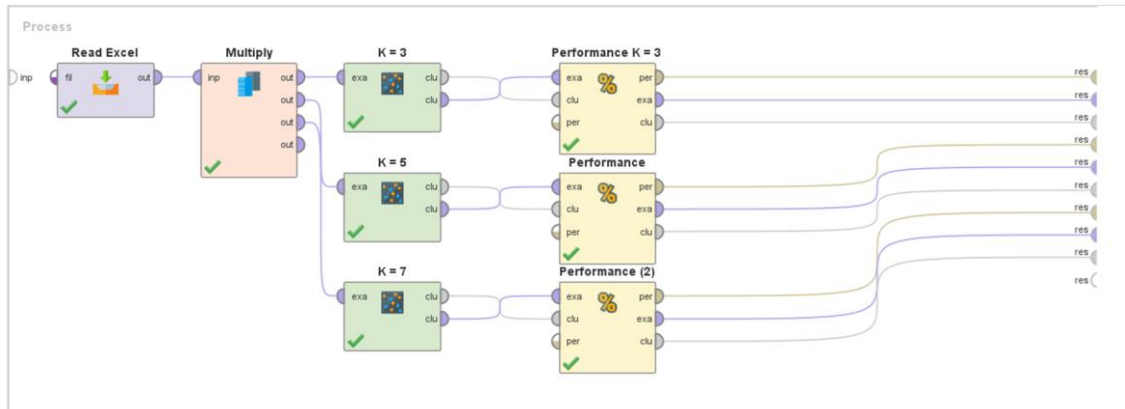
Gambar 3. Pemetaan titik rawan kecelakaan berdasarkan cluster

Setelah mendapatkan hasil pada iterasi 2, selanjutnya akan dilakukan pengelompokan titik kecelakaan berdasarkan klaster masing masing yang akan dijadikan kedalam sebuah peta. Gambar 3 akan menunjukkan hasil pemetaan berdasarkan klaster.

#### 4.5 Uji Validasi

Untuk membuktikan bahwa penggunaan model *k-means clustering* dengan jumlah klaster sebanyak tiga merupakan jumlah klaster dengan performa paling baik, maka akan dilakukan validasi model. Pada penelitian ini, model validasi yang dipakai adalah *Davies boulder index*. Untuk mengetahui apakah model pembagian klaster menjadi tiga adalah model yang optimal, maka akan dilakukan perbandingan jumlah klaster. Pada tahap validasi ini, model dengan jumlah klaster tiga akan dibandingkan dengan jumlah klaster lima, tujuh dan Sembilan. Alat yang dipakai dalam menguji validasi model *K-means Clustering* adalah *Rapidminer*. Proses validasi akan digambarkan dalam gambar 4.





Gambar 4. Uji coba validasi menggunakan *Rapidminer*

Pada uji validasi, data yang di ambil dari tabel 2. Setelah memasukkan data yang dijadikan acuan untuk uji validasi, data tersebut akan melewati uji *K-means Clustering*, dengan jumlah kluster sebanyak tiga, lima, dan tujuh. Setelah mendapatkan hasil uji *k-means clustering* untuk tiap kluster, tiap kluster yang diuji akan diambil hasilnya untuk melakukan penghitungan nilai *davies boulder index* dan pada uji validasi ini, nilai *davies boulder index* untuk jumlah kluster tiga adalah sebesar 0.027, kluster dengan jumlah lima sebesar 0.073 dan kluster dengan jumlah tujuh sebesar 0.055. setelah mendapatkan nilai *davies boulder index*, model kluster dengan nilai yang paling kecil akan diambil sebagai model yang optimal. Pada penelitian ini, model dengan jumlah kluster tiga dianggap sebagai yang paling optimal dikarenakan nilai *davies boulder index* yang diperoleh lebih kecil diantara yang lainnya.

#### 4.6 Pembahasan

Dari penelitian ini, berdasarkan hasil *k-means clustering*, daerah–daerah pada kecamatan Bawen dibagi menjadi tiga kluster berdasarkan tingkat kerawanan terhadap kecelakaan lalu lintas. Pada kluster pertama atau kluster sangat rawan kecelakaan beranggotakan kelurahan bawen, pada kluster kedua atau kluster rawan kecelakaan beranggotakan desa Harjosari dan pada kluster ketiga atau kluster agak rawan kecelakaan beranggotakan desa Asinan, Dopleng, Kandangan, Lemahireng, Polosiri, Poncoruso dan Samban.

Dengan adanya hasil uji *k-means clustering* dan hasil pemetaan pada gambar 2, dapat membantu dalam menggambarkan bagaimana pola persebaran kecelakaan lalu lintas didalam kecamatan Bawen dan pada kasus ini, pada gambar 2 dapat dilihat bagaimana persebaran kejadian kecelakaan pada kecamatan Bawen dari tahun 2021 hingga tahun 2023, dimana daerah dengan kasus kecelakaan lalu lintas paling banyak terjadi pada kelurahan Bawen dan diikuti oleh Desa Harjosari. Berdasarkan hasil yang ada, hal ini sesuai dengan penelitian milik Sachin Kumar dan Durga Toshniwal yang berjudul *a data mining approach to characterize road accident locations* dimana dalam penelitiannya, penulis menggunakan *K-means Clustering* untuk membagi area berdasarkan intensitas kecelakaan dan terciptalah 3 kluster yang mewakili area frekuensi kecelakaan tinggi, kecelakaan sedang dan rendah kecelakaan. Dari pembagian ketiga area inilah yang dijadikan parameter dalam menentukan pola persebaran kecelakaan lalu lintas[17] dan berdasarkan penelitian milik B.Topcuoglu, dkk berjudul *speed-related traffic accident analysis using GIS-based DBSCAN and NNH clustering* bahwa dengan adanya pemetaan menggunakan sistem kluster dapat berperan dalam menentukan pola dimana kejadian kecelakaan lalu lintas sering terjadi dan dapat dijadikan dasar dalam menentukan aksi untuk mengurangi jumlah kecelakaan lalu lintas[18]

Hasil pengujian membuktikan nilai *davies-boulder index* pada *k-means clustering* dengan jumlah kluster 3 lebih optimal dibandingkan model *k-means clustering* dengan jumlah kluster yang lima dan tujuh. Hal ini sejalan dengan penelitian milik Devanta Abraham Tarigan, dimana dia membahas semakin kecil nilai *davies-boulder index* yang dimiliki suatu model *K-means Clustering* maka semakin optimal model *k-means clustering* yang digunakan[16].

#### 5. Simpulan

Hasil akhir dari penelitian ini membagi desa atau kelurahan yang menjadi bagian Kecamatan Bawen kedalam tiga klaster, klaster pertama merupakan daerah yang sangat rawan kecelakaan dengan satu anggota yaitu kelurahan Bawen. Klaster kedua merupakan daerah yang rawan kecelakaan dengan satu anggota yaitu Harjosari dan klaster ketiga merupakan daerah yang agak rawan kecelakaan dengan tujuh anggota yang terdiri dari Asinan, Dopleng, Kandangan, Lemahireng, Polosiri, Poncoruso dan Samban.

Dalam memastikan bahwa pemakaian *K-means Clustering* dengan jumlah klaster sebanyak tiga merupakan model yang optimal, maka dilakukan uji validitas menggunakan model *Davies Boulder Index*. Pada tahap ini, model dengan jumlah tiga klaster dibandingkan dengan model dengan jumlah lima klaster dan tujuh klaster. Dari hasil perbandingan tersebut, klaster dengan jumlah tiga menghasilkan nilai sebesar 0.027, klaster dengan jumlah lima menghasilkan nilai sebesar 0.073 dan klaster dengan jumlah tujuh menghasilkan nilai sebesar 0.055. dan hasil akhir menyatakan bahwa penggunaan model dengan klaster berjumlah tiga lebih optimum dibandingkan penggunaan model dengan lima dan tujuh klaster.

Dengan adanya penelitian ini, dapat digunakan bagi pemerintah dan kepolisian setempat dalam membuat kebijakan dalam keamanan berkendara di kecamatan bawen berdasarkan pemetaan yang telah dilakukan. Selain dapat berperan penting bagi pemerintah dan kepolisian setempat, penelitian ini juga dapat berperan penting bagi Masyarakat umum. Dengan adanya penelitian ini, Masyarakat bisa dapat lebih tau mengenai kerawanan dalam kecelakaan lalu lintas pada daerah–daerah pada kecamatan Bawen.

Saran untuk peneliti yang ingin mengembangkan topik ini, disarankan untuk menggunakan model *clustering* yang lain agar hasil yang didapatkan lebih efektif dan optimal.

#### Daftar Referensi

- [1] H. Wahyono, H. Setiaji, T. Hartati, and N. Wiliyani, "K-Means Clustering for Identifying Traffic Accident Hotspots in Depok City," vol. 2, no. 1, pp. 159–170, 2024, doi: 10.61098/jarcis.v2i1.182.
- [2] D. Puspitasari, M. Wahyudi, M. Rizaldi, A. Nurhadi, K. Ramanda, and Sumanto, "K-Means Algorithm for Clustering the Location of Accident-Prone on the Highway," *J. Phys. Conf. Ser.*, vol. 1641, no. 1, pp. 1–6, 2020, doi: 10.1088/1742-6596/1641/1/012086.
- [3] C. Sinclair and S. Das, "Traffic Accidents Analytics in UK Urban Areas using k-means Clustering for Geospatial Mapping," *2021 Int. Conf. Sustain. Energy Futur. Electr. Transp.*, pp. 1–7, 2021, doi: 10.1109/SeFet48154.2021.9375817.
- [4] B. A. Fauzan, M. Jamaris, and J. Junadhi, "Implementation of K-Means Clustering Algorithm for Grouping Traffic Violation Levels in Siak," *J. Teknol. Dan Open Source*, vol. 5, no. 1, pp. 81–88, 2022, doi: 10.36378/jtos.v5i1.2427.
- [5] H. Kamh, S. H. Alyami, A. Khattak, M. Alyami, and H. Almujiabah, "Exploring Road Traffic Accidents Hotspots Using Clustering Algorithms and GIS-based Spatial Analysis," *IEEE Access*, vol. PP, p. 1, 2024, doi: 10.1109/ACCESS.2024.3443245.
- [6] A. Pratama, M. D. Irawan, and S. D. Andriana, "Implementation of K-Means Clustering in Recognizing Crime Hotspots and Traffic Issues Through GIS," *J. Comput. Networks, Archit. High Perform. Comput.*, vol. 6, no. 2, pp. 771–782, 2024, doi: 10.47709/cnahpc.v6i2.3771.
- [7] K. P. Siregar *et al.*, "Implementation Of Purity K-Means Algorithm In Accident Data Clustering In North Padang Lawas," *2nd Int. Conf. Multidiscip. Eng.*, vol. 2, pp. 1–11, 2024.
- [8] W. F. W. Yaacob, S. B. Ibrahim, A. S. Afizan, N. A. F. Azran, S. A. Md Nasir, and N. C. Harun, "Spatio-Temporal Clustering of Road Accidents in Kelantan, Malaysia," *Int. J. Acad. Res. Bus. Soc. Sci.*, vol. 11, no. 9, pp. 522–534, 2021, doi: 10.6007/ijarbss/v11-i9/11036.
- [9] L. L. Latifah, S. A. Hudjimartsu, and I. Yanuarsyah, "Analisis Daerah Rawan Kecelakaan Lalu Lintas Menggunakan Cluster Analysis Di Kota Bogor Berbasis Webgis," *J. Ilm. Teknol. Infomasi Terap.*, vol. 8, no. 2, pp. 235–244, 2022, doi: 10.33197/jitter.vol8.iss2.2022.760.
- [10] N. Rahmadani, E. Rahayu, and A. Lestari, "K-Means Clustering Areas Prone To Traffic Accidents in Asahan Regency," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 6, no. 2, pp. 181–186, 2021, doi: 10.33480/jitk.v6i2.1519.K-MEANS.
- [11] A. A. Vernanda, A. Faisol, and N. Vendyansyah, "Penerapan Metode K-Means

- Clustering Untuk Pemetaan Daerah Rawan Kecelakaan Lalu Lintas Di Kota Malang Berbasis Website,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 5, no. 2, pp. 836–844, 2021, doi: 10.36040/jati.v5i2.3791.
- [12] V. Irawan, A. Rizal, and I. Purnamasari, “Penerapan Algoritma K-Mean Clustering Pada Kecelakaan Lalu Lintas Berdasarkan Kabupaten/Kota Di Provinsi Jawa Tengah Tahun 2020,” *J. Ilm. Wahana Pendidik.*, vol. 8, no. July 2020, pp. 293–300, 2022, doi: <https://doi.org/10.5281/zenodo.6820090>.
- [13] A. H. A. N. Karsa and A. R. Hidayat, “Metode Algoritma K-Means Untuk Clustering Data Produk Paling Laku Pada Toko Tono Grosir Plumbon Cirebon,” *Syntax Lit. ; J. Ilm. Indones.*, vol. 7, no. 9, pp. 15984–15996, 2024, doi: 10.36418/syntax-literate.v7i9.15144.
- [14] D. Remawati, D. J. Aji Putra, and T. Irawati, “Metode K-Means Untuk Pemetaan Persebaran Usaha Mikro Kecil Dan Menengah,” *J. Teknol. Inf. dan Komun.*, vol. 9, no. 2, pp. 39–46, 2021, doi: 10.30646/tikomsin.v9i2.574.
- [15] Y. Sopyan, A. D. Lesmana, and C. Juliane, “Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1464–1470, 2022, doi: 10.47065/bits.v4i3.2697.
- [16] D. A. Tarigan, “Optimization of the K-Means Clustering Algorithm Using Davies Bouldin Index in Iris Data Classification,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 545–552, 2023, doi: 10.30865/klik.v4i1.964.
- [17] S. Kumar and D. Toshniwal, “A data mining approach to characterize road accident locations,” *J. Mod. Transp.*, vol. 24, no. 1, pp. 62–72, 2016, doi: 10.1007/s40534-016-0095-5.
- [18] B. Topcuoglu, T. Memisoglu Baykal, and H. Tuydes Yaman, “Speed-Related Traffic Accident Analysis Using Gis-Based DbSCAN and Nnh Clustering,” *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, vol. 48, no. 4/W1-2022, pp. 487–494, 2022, doi: 10.5194/isprs-archives-XLVIII-4-W1-2022-487-2022.