



Penerapan Algoritma K-Mean Clustering Pada Kecelakaan Lalu Lintas Berdasarkan Kabupaten/Kota Di Provinsi Jawa Tengah Tahun 2020

Valdi Irawan¹, Adhi Rizal², Intan Purnamasari³

^{1,2,3}Dosen Universitas Singaperbangsa Karawang

Received: 16 Juni 2022

Revised: 24 Juni 2022

Accepted: 1 Juli 2022

Abstract

The increase in population has resulted in the need for motorized vehicles so that it triggers traffic accidents. The grouping of traffic accident data is important to do, because it will make it easier for the police, especially the traffic accident unit in each district/city police station in Central Java province in reducing the number of traffic accidents. In this study, using Data Mining with the K-Means Algorithm with the aim of obtaining information about traffic accident data groups based on the incidence rate so that it is known the level of traffic accidents that occur in each district / city in the province of Central Java. By using the K-means algorithm, 3 clusters with different accident rates are obtained, cluster 1 is the low accident rate, cluster 2 is moderate, and cluster 3 is the highest accident rate.

Keywords: *Data Mining, Clustering, K-Means, traffic accident*

(*) Corresponding Author:

How to Cite: Irawan, V., Rizal, A., & Purnamasari, I. (2022). Application of the K-Mean Clustering Algorithm in Traffic Accidents by Regency/City in Central Java Province in 2020. *Jurnal Ilmiah Wahana Pendidikan*, 8(10), 293-300. <https://doi.org/10.5281/zenodo.6820090>

PENDAHULUAN

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, bahwa setiap tahun di seluruh dunia ada lebih dari 1,25 juta korban meninggal akibat kecelakaan lalu lintas dan 50 juta orang luka berat.

Direktorat Lalu Lintas Polda Metro Jaya mencatat jumlah kecelakaan yang terjadi selama tahun 2020 mencapai 4559 kejadian kecelakaan lalu lintas yang diantaranya 346 meninggal dunia, 1 korban luka berat, dan 5715 korban luka ringan. Namun pada kenyataan dilapangan data tersebut bisa saja bertambah karena sering kali 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) di seluruh polres kabupaten/kota di Provinsi Jawa Tengah dalam upaya penanganan yang tepat untuk mengurangi angka kecelakaan lalu lintas.

Algoritma K-Means dapat melakukan clustering wilayah pada kecelakaan lalu lintas berdasarkan tinggi, sedang dan rendah tingkatan kecelakaan pada kabupaten/kota di provinsi Jawa Tengah. Cluster yang terbentuk menghasilkan pengetahuan terkait kasus perkara lalu lintas dan bisa digunakan sebagai kelas target pada Teknik klasifikasi (Delita & Fitri, 2014).

Berdasarkan permasalahan yang ada, maka pada kesempatan kali ini akan dilakukan penelitian terhadap data kecelakaan lalu lintas yang didapat dari Polres kabupaten/kota di provinsi Jawa Tengah. 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 tingkatan jumlah kecelakaan sehingga diketahui tingkatan kejadian kecelakaan lalu lintas di wilayah kabupaten/kota di Provinsi Jawa Tengah

METODE

Dalam penelitian ini metode yang digunakan adalah metode data mining dengan menggunakan Alogritma K-Means, terdapat beberapa tahapan diantaranya tahapan pengumpulan data, tahapan pengolahan data, tahapan clustering, dan tahap analisis. Tahapan dalam metode yang digunakan pada penelitian ini dijelaskan sebagai berikut :

Tahapan Pengumpulan Data

Sumber data pada penelitian ini diperoleh dari data yang dikumpulkan melalui situs Badan Pusat Statistika Provinsi Jawa Tengah (jateng.bps.go.id) yang memuat data kecelakaan lalu lintas di kabupaten/kota berdasarkan jumlah kejadian di Provinsi Jawa Tengah pada tahun 2020.

Tahap Pengolahan Data

Data yang sudah dikumpulkan diolah untuk dikelompokkan atau clustering menggunakan algoritma data mining K-Means. Data akan dikelompokkan menjadi 3 kluster yakni, rendah, sedang, dan tinggi. Pada penelitian ini, pengolahan data menggunakan aplikasi Microsoft Excel dan akan diproses pada tahap clustering.

Tahapan Clustering

Pada tahapan ini dilakukan pengelompokan atau clustering menggunakan teori data mining dengan algoritma K-Means. Berikut urutan alur proses algoritma K-Means ditunjukkan pada konsep bagan sebagai berikut



Tahap Analisis

Pada tahap analisis ini dilakukan analisis data banyaknya jumlah kejadian kecelakaan lalu lintas di Kabupaten/Kota di Provinsi Jawa Tengah. Data yang diperoleh dari hasil pengumpulan data terdiri dari akumulasi kecelakaan dengan korban jiwa atau meninggal, luka ringan, dan luka berat. Selanjutnya data dikelompokkan menjadi 3 cluster.

HASIL & PEMBAHASAN

Pada penelitian ini, data yang digunakan berjumlah 35 set data dengan 4 variable, yakni informasi nama kabupaten/kota dan jenis kecelakaan lalu lintas. Diperoleh tabel data seperti berikut :

No	Kabupaten/Kota	Meninggal	Luka Berat	Luka Ringan
1	Kabupaten Cilacap	172	1	1035
2	Kabupaten Banyumas	217	1	1765
3	Kabupaten Purbalingga	75	0	497
4	Kabupaten Banjarnegara	75	4	528
5	Kabupaten Kebumen	121	0	732

6	Kabupaten Puworejo	87	0	561
7	Kabupaten Wonosobo	53	2	319
8	Kabupaten Magelang	135	2	941
9	Kabupaten Boyolali	91	1	1087
10	Kabupaten Klaten	142	7	1712
11	Kabupaten Sukoharjo	80	1	1186
12	Kabupaten Wonogiri	65	0	754
13	Kabupaten Karanganyar	99	4	1208
14	Kabupaten Sragen	126	0	909
15	Kabupaten Grobogan	123	2	577
16	Kabupaten Blora	70	6	357
17	Kabupaten Rembang	93	0	397
18	Kabupaten Pati	153	1	1446
19	Kabupaten Kudus	109	0	890
20	Kabupaten Jepara	97	0	363
21	Kabupaten Demak	154	5	582
22	Kabupaten Semarang	121	1	539
23	Kabupaten Temanggung	59	0	350
24	Kabupaten Kendal	114	2	426
25	Kabupaten Batang	118	3	453
26	Kabupaten Pekalongan	53	1	254
27	Kabupaten Pemalang	100	1	507
28	Kabupaten Tegal	134	0	506
29	Kabupaten Brebes	171	1	972
30	Kota Magelang	25	0	227
31	Kota Surakarta	50	0	852
32	Kota Salatiga	33	0	264
33	Kota Semarang	121	1	539
34	Kota Pekalongan	25	2	117
35	Kota Tegal	17	0	254
Total		3508	48	24495

- a. Dalam mengelompokan data dengan metode clustering, hal pertama yang diperlukan adalah menentukan jumlah kluster. Pada penelitian ini, jumlah kluster yang ditentukan adalah 3 kluster. 3 kluster ini mengacu pada tingkatan kecelakaan lalu lintas yang terjadi pada kabupaten/kota di provinsi Jawa Tengah tersebut, yakni rendah, sedang dan tinggi.
- b. Setelah menentukan jumlah kluster yang ingin dikelompokan, tahap selanjutnya adalah menentukan centroid (titik pusat) awal. Jumlah centroid disamakan dengan jumlah kluster yang ditentukan. Data centroid diambil secara acak. Didapatkan tabel centroid awal berikut ini :

No	Kabupaten/Kota	Meninggal	Luka Berat	Luka Ringan
----	----------------	-----------	------------	-------------

1	Kabupaten Cilacap	25	2	177
2	Kabupaten Kudus	109	0	890
3	Kabupaten Banyumas	217	1	1765

- c. Setelah ditentukan titik pusat awal, hitung jarak setiap data ke masing – masing titik awal dengan menggunakan rumus:

$$d(x,y) = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Keterangan :

D(x,y)_j = Jarak objek antara objek x dan y n = Dimensi data

Xi = Koordinat dari obyek xi

Xj = Koordinat dari obyek yi

Diperoleh hasil dari perhitungan iterasi pertama. Tabel hasil perhitungan jarak diperoleh sebagai berikut

No	Kabupaten/Kota	C1	C2	C3	Hasil	Jarak Terpendek
1	Kabupaten Cilacap	870,5021539	158,0980708	731,3856712	2	158,0980708
2	Kabupaten Banyumas	1599,565253	881,6405163	0	3	0
3	Kabupaten Purbalingga	323,8888698	394,4679962	1275,926722	1	323,8888698
4	Kabupaten Banjarnegara	354,5490093	363,6151812	1245,127303	1	354,5490093
5	Kabupaten Kebumen	563,2450621	158,4550409	1037,451686	2	158,4550409
6	Kabupaten Puworejo	388,9781485	329,7347419	1210,998348	2	329,7347419
7	Kabupaten Wonosobo	144,7342392	573,7429738	1455,270765	1	144,7342392
8	Kabupaten Magelang	771,8782287	57,28001397	828,0706492	2	57,28001397
9	Kabupaten Boyolali	912,3908154	197,8231533	689,6085846	2	197,8231533
10	Kabupaten Klaten	1539,46062	822,6919229	92,03260292	3	92,03260292
11	Kabupaten Sukoharjo	1010,498392	297,4188965	594,9873948	2	297,4188965
12	Kabupaten Wonogiri	578,3882779	142,9405471	1022,362949	2	142,9405471
13	Kabupaten Karanganyar	1033,654197	318,1823377	569,3698271	2	318,1823377
14	Kabupaten Sragen	738,9377511	25,49509757	860,8240238	2	25,49509757
15	Kabupaten Grobogan	411,830062	313,3193259	1191,713472	2	313,3193259
16	Kabupaten Blora	185,5828656	534,4586046	1415,661683	1	185,5828656
17	Kabupaten Rembang	230,2780927	493,2595666	1373,608751	1	230,2780927
18	Kabupaten Pati	1275,439532	557,7391864	325,3567273	3	325,3567273
19	Kabupaten Kudus	717,933841	0	881,6405163	2	0
20	Kabupaten Jepara	199,459269	527,1366047	1407,126505	1	199,459269
21	Kabupaten Demak	425,0588195	311,3101348	1184,68308	2	311,3101348
22	Kabupaten Semarang	374,5143522	351,206492	1229,752821	2	351,206492
23	Kabupaten Temanggung	176,3207305	542,3098745	1423,794227	1	176,3207305
24	Kabupaten Kendal	264,4276839	464,0312489	1342,956068	1	264,4276839
25	Kabupaten Batang	291,2490343	437,1029627	1315,731356	1	291,2490343

26	KabupatenPekalongan	81,9390017	638,4614319	1519,874008	1	81,9390017
27	Kabupaten Pemalang	338,4169027	383,1070347	1263,429064	1	338,4169027
28	Kabupaten Tegal	346,5919791	384,8129416	1261,733332	1	346,5919791
29	Kabupaten Brebes	808,295738	102,8056419	794,3330536	2	102,8056419
30	Kota Magelang	50,03998401	668,3000823	1549,938386	1	50,03998401
31	Kota Surakarta	675,4657652	70,17834424	928,1481563	2	70,17834424
32	Kota Salatiga	87,38993077	630,596543	1512,236093	1	87,38993077
33	Kota Semarang	374,5143522	351,206492	1229,752821	2	351,206492
34	Kota Pekalongan	60	777,5532136	1659,14707	1	60
35	Kota Tegal	77,44029959	642,6196387	1524,179123	1	77,44029959

Dari perhitungan jarak terhadap centroid awal, didapatkan hasil 16 set data kluster 1, 16 set data kluster 2, dan 3 set data kluster 3.

- d. Selanjutnya, lakukan iterasi kedua dengan menentukan centroid baru dengan menghitung rata – rata pada set data anggota yang sama. Berikut hasil perhitungan rata – rata berdasarkan set data yang sama
- e. Pada iterasi kedua ini, lakukan hal yang sama seperti iterasi pertama, yaitu menghitung jarak ke centroid baru.
- f. Selanjutnya, jika hasil dari iterasi pertama dan kedua masih belum sama, lakukan iterasi ketiga dengan menentukan centroid baru dengan menghitung rata – rata pada set data anggota yang sama.
- g. Pada iterasi ketiga ini, lakukan hal yang sama seperti iterasi sebelumnya, yaitu menghitung jarak ke centroid baru
- h. Davies-Bouldin Index

Tahap ini menjelaskan proses validasi internal pada cluster berdasarkan tingkat densitas data dengan centroid, keterpisahan antar cluster dan rasio antar cluster. Hasil ketiga nilai tersebut dijadikan dasar untuk mendapatkan Davies-Bouldin Index. Tahapan yang dilakukan yaitu, hitung Sum of Square Within Cluster (SSW), selanjutnya hitung Sum of Square Between-cluster (SSB), dan terakhir hitung Rasio dan DBI.

- i. Langkah pertama memperoleh SSW dengan persamaan adalah sebagai berikut

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

Keterangan

Mi : Banyak data dalam cluster ke i

Ci : centroid kluster ke i

d(xj,ci) : Jarak antara data ke i ke cluster ke i

Perhitungan SSW didapatkan dengan menghitung jarak setiap data pada centroid dan dihitung rata – ratanya. Hasil perhitungan lengkap SSW ditunjukkan pada tabel sebagai berikut

SSW	120,693647
1	9
SSW	134,648413
2	4

SSW	134,956929
3	7

- ii. Setelah didapatkan nilai SSW, selanjutnya menentukan nilai SSB dengan menghitung jarak antar centroid suatu cluster dengan persamaan sebagai berikut :

$$SSB_{ij} = d(c_i, c_j)$$

Hasil Perhitungan lengkap SSB ditunjukkan pada tabel sebagai berikut

SS B	Centoi d		
	1	2	3
1	0	550,905562 1	1233,77266 9
2	550,905562 1	0	683,084622 5
3	1233,77266 9	683,084622 5	0

- iii. Setelah didapatkan nilai SSB, selanjutnya menentukan nilai rasio dan DBI. Hasil perhitungan Rasio dan DBI ditunjukkan pada tabel sebagai berikut

Rasio	1	2	3	Max	DBI
1	0	0,463495159 4	0,207210440 1	0,463495159 4	0,44055945 9
2	0,463495159 4	0	0,394688058 1	0,463495159 4	
3	0,207210440 1	0,394688058 1	0	0,394688058 1	

Semakin kecil nilai DBI yang diperoleh (non-negatif ≥ 0) maka semakin baik cluster yang diperoleh dari pengelompokan algoritma clustering. Dari hasil perhitungan Rasio dan DBI, diperoleh nilai DBI sebesar 0,440559459 sehingga dapat dikatakan hasil kluster cukup baik.

KESIMPULAN & SARAN

Kesimpulan

Dari hasil penelitian ini kabupaten/kota terbagi menjadi 3 cluster dengan tiap cluster memiliki arti yang berbeda – beda, cluster 1 menandakan bahwa wilayah tersebut memiliki tingkat kecelakaan lalu lintas yang rendah, cluster 2 menandakan bahwa wilayah tersebut memiliki tingkatan kecelakaan lalu lintas yang sedang, dan untuk cluster 3 menandakan bahwa wilayah tersebut memiliki tingkat kecelakaan yang tinggi. Kabupaten/kota yang masuk kedalam cluster 1 terdapat 21

kabupaten/kota, untuk cluster 2 terdapat 11 kabupaten/kota, dan cluster 3 terdapat 3 kabupaten/kota.

Saran

Dari hasil ini diharapkan dapat membantu Polres di provinsi Jawa Tengah agar dapat mengurangi tingkat kecelakaan lalu lintas di provinsi tersebut. Penelitian ini hanya menggunakan satu metode clustering, penulis berharap akan lebih baik apabila penelitian selanjutnya dapat menambahkan metode clustering lain sehingga dapat menjadi perbandingan metode mana yang lebih baik untuk digunakan dengan hasil maksimal.

DAFTAR PUSTAKA

- Allen, J., & Shockley, B. B. (1998). Potential engagements: Dialogue among school and university research communities. In B. S. Bisplinghoff & J. Allen (Eds.), *Engaging teachers: Creating teaching/researching relationships* (pp. 61-71). Portsmouth, NH: Heinemann.
- Ardiansyah, J. (2013). Peningkatan kompetensi guru bidang pendidikan di Kabupaten Tana Tidung. *EJournal Pemerintahan Integratif*, 1(1), 38-50.
- Anshori Iedam Fardian, Nuraini Yeni. 2020. *Pengelompokan Data Kecelakaan Lalu Lintas di Kota Tasikmalaya Menggunakan Algoritma K-Means*. Jurnal Responsif Universitas Adhirajasa Reswara Sanjaya
- Purwaningsih Esty. 2019. *Analisis Kecelakaan Berlalu Lintas di Kota Jakarta Dengan Menggunakan Metode K-Means*. Jurnal ilmu pengetahuan dan teknologi computer Universitas Bina Sarjana Informatika
- Aprianti Winda, Permadi Jaka. 2017. *K-Means Clustering Untuk Data Kecelakaan Lalu Lintas Jalan Raya di Kecamatan Pelaihari*. Jurnal Teknologi Informasi dan Ilmu Komputer Politeknik Negeri Tanah Laut
- Delita, Fitri. 2014. *Clustering Pelanggaran Berkendaraan Menggunakan Algoritma K-Means Pada Polres Banjai*. Jurnal STMIK Kaputama
- Rata – rata tiga orang meninggal setiap jam akibat kecelakaan jalan. Kominfo.go.id
- Jumlah Korban Kecelakaan Lalu Lintas di Wilayah Polda Jawa Tengah Tahun 2018-2020*. Jateng.bps.go.id