E-ISSN : 3025-1311
https://ejournal.warunayama.org/kohesi

Kohesi: Jurnal Multidisiplin Saintek
Volume 7 No 2 Tahun 2025

PENGGUNAAN K-MEANS UNTUK IDENTIFIKASI DAERAH RAWAN KECELAKAAN LALU LINTAS DI KOTA YOGYAKARTA

Khairil Hakim[1], Yoga Pramudya[2], Fajar Hidayat[3], Tedy Setiadi[4]

[1,2, 3, 4] Informatika, Universitas Ahmad Dahlan
Jl. Ringroad Selatan, Kragilan, Tamanan, Kec. Banguntapan, Kabupaten Bantul, Daerah Istimewa Yogyakarta 55191

E-mail : 2200018263@webmail.uad.ac.id[1], 2200018265@webmail.uad.ac.id[2], 2200018275@webmail.uad.ac.id[3], tedy.setiadi@tif.uad.ac.id[4]

Abstrak

Penelitian ini bertujuan untuk mengidentifikasi daerah rawan kecelakaan lalu lintas di kota Yogyakarta dengan menggunakan algoritma *K-Means*. Algoritma ini merupakan teknik pengelompokan non-hierarki yang mengelompokkan data berdasarkan karakteristik serupa. Data kecelakaan lalu lintas tahun 2023 diperoleh dari situs *Open Data Kota Yogyakarta*. Proses analisis meliputi identifikasi masalah, pengumpulan data, pengolahan data, klasterisasi, dan analisis hasil. Dengan membagi informasi menjadi tiga kelompok, yakni rendah, sedang, dan tinggi, penelitian ini berhasil mengidentifikasi pola kecelakaan dan daerah rawan kecelakaan. Temuan menunjukkan bahwa metode *K-Means* berguna dalam meneliti data tentang kecelakaan jalan dan memberikan wawasan yang penting untuk keputusan kebijakan. Penelitian ini diharapkan dapat berkontribusi dalam meningkatkan keselamatan lalu lintas di Kota Yogyakarta.

Kata kunci : kecelakaan lalu lintas, *K-Means*, *clustering*, Yogyakarta.

Abtract

The K-Means clustering algorithm, a non-hierarchical method that groups data according to shared criteria, is being used in this project to find Yogyakarta regions that are prone to traffic accidents. Yogyakarta's Open Data portal provided the 2023 traffic accident data. Problem identification, data collection, preprocessing, clustering, and result interpretation are all included in the analysis. This study found high-risk areas and accident patterns by grouping the data into low, medium, and high clusters. The results demonstrate how useful K-Means is for examining data on traffic accidents and offer insightful information for policymakers. This study helps to increase Yogyakarta's traffic safety.

Keywords : *traffic accident, K-Means, clustering, Yogyakarta.*

Article History
Received: Januari 2025
Reviewed: Januari 2025
Published: Januari 2025

Plagirism Checker No 234
Prefix DOI : Prefix DOI : 10.8734/Kohesi.v1i2.365

Publish by : Kohesi

1. Pendahuluan

Meningkatnya jumlah penduduk telah menyebabkan permintaan yang lebih besar untuk transportasi bermotor, khususnya kendaraan pribadi. Orang cenderung lebih menyukai kendaraan pribadi karena dianggap lebih mudah beradaptasi dan hemat biaya [1]. Kecelakaan lalu lintas menjadi salah satu tantangan utama dalam sektor transportasi yang memerlukan penanganan serius, selain isu kemacetan [2]. Tingginya aktivitas lalu lintas di Kota Yogyakarta menyebabkan angka kecelakaan lalu lintas relatif besar. Kecelakaan lalu lintas merupakan

kejadian yang tidak terduga di jalan, yang melibatkan kendaraan bermotor dengan atau tanpa pengguna jalan lain dan dapat menimbulkan kerugian harta benda, cedera, atau bahkan korban jiwa [3].

Jumlah insiden lalu lintas di Yogyakarta pada tahun 2023 tercatat sebanyak 830 kejadian [4]. Masalah ini memberikan dampak besar terhadap masyarakat, baik dari segi keselamatan, ekonomi, maupun sosial, sehingga upaya mitigasi sangat diperlukan untuk menciptakan lingkungan lalu lintas yang lebih aman. Salah satu cara untuk melakukannya adalah dengan mengenali area-area yang berisiko kecelakaan, yang dapat menjadi dasar untuk merencanakan langkah-langkah pencegahan yang lebih efektif.

Dalam hal ini, teknologi analitik berperan penting, terutama dengan memanfaatkan algoritma *K-Means*, yang merupakan metode untuk mengatur data berdasar pada lokasi. *K-Means* adalah metode untuk mengatur data yang tidak mengikuti struktur *hierarki*, di mana data dikelompokkan ke dalam beberapa kategori berdasarkan kemiripan atribut tertentu. Dengan menggunakan *K-Means*, Data dengan sifat yang serupa akan dikelompokkan ke dalam satu kategori, sedangkan data yang memiliki sifat yang berbeda akan dikelompokkan ke dalam kategori yang lain [5]. *K-Means* adalah salah satu metode pengelompokan yang menggunakan pendekatan pemisahan berdasarkan titik pusat. [6]. Tujuan dari *K-Means* merupakan pengelompokan informasi dengan tujuan untuk meningkatkan kesamaan data di dalam satu kelompok [7]. Dengan penerapan *K-Means*, daerah rawan kecelakaan dapat teridentifikasi dengan lebih tepat untuk mendukung pengambilan keputusan yang lebih baik.

Algoritma *K-Means* digunakan untuk proses *clustering* dimana atribut yang berisi data kontinu menjadi kategorikal dalam bentuk *cluster* [8]. Dengan penerapan *K-Means*, daerah rawan kecelakaan dapat teridentifikasi dengan lebih tepat untuk mendukung pengambilan keputusan yang lebih baik. Algoritma *K-Means* mengelompokkan data menjadi beberapa *cluster*, di mana informasi yang memiliki ciri-ciri serupa dikelompokkan menjadi satu *cluster*, sedangkan informasi dengan ciri-ciri yang tidak sama dikelompokkan ke dalam *cluster* yang berbeda [9].

Penelitian ini bertujuan untuk mengidentifikasi area berisiko tinggi terhadap kecelakaan lalu lintas di Kota Yogyakarta dengan memanfaatkan algoritma *K-Means*. Algoritma ini dipilih karena kemudahan implementasi, efisiensi, dan kemampuannya dalam mengelompokkan wilayah rawan kecelakaan secara akurat. Diharapkan temuan dari studi ini dapat memberikan sumbangan yang signifikan untuk meningkatkan keamanan di jalan raya serta berfungsi sebagai acuan dalam pengambilan keputusan yang terkait.

2. Metodologi

Studi ini melibatkan tinjauan literatur yang ada, dimulai dengan identifikasi masalah yang berkaitan dengan subjek penelitian [10]. Penelitian ini menggunakan teknik penambangan data, khususnya memanfaatkan Algoritma *K-Means*. Secara umum, *clustering* merupakan teknik untuk membedakan kumpulan data menjadi beberapa kelompok berdasarkan kecocokan tertentu [11]. Algoritma *K-Means* merupakan salah satu metode yang paling banyak digunakan untuk melakukan pengelompokan. Metode ini berfungsi untuk mengatur informasi ke dalam berbagai kelompok berdasarkan seberapa dekat atau mirip fitur tertentu [12]. Proses ini dilakukan dengan mengelompokkan objek ke dalam klaster berdasarkan kedekatan terhadap pusat klaster (*centroid*), yang diperbarui secara iteratif hingga hasil optimal tercapai.

Metode ini terdiri dari beberapa tahapan utama yaitu Identifikasi masalah, pengumpulan data, pengolahan data, proses klasterisasi, dan analisis hasil. Tahapan metodologi yang digunakan dalam penelitian ini dirinci di bawah ini:

Tahap Identifikasi Masalah

Penelitian ini berfokus pada analisis Kecelakan di jalan raya yang terjadi di daerah Yogyakarta dengan penerapan metode *K-Means* untuk mengelompokkan data menurut fitur tertentu guna mengidentifikasi pola atau wilayah rawan kecelakaan.

Tahap Pengumpulan Data

Informasi yang dipakai dalam penelitian ini diambil dari situs *Open Data Kota Yogyakarta*. (dataset.jogjakota.go.id). Dataset ini menyediakan informasi kecelakaan lalu lintas di wilayah Yogyakarta tahun 2023, mencakup data jumlah kejadian, lokasi, dan waktu. Data tersebut digunakan sebagai dasar untuk mendukung proses analisis dalam penelitian.

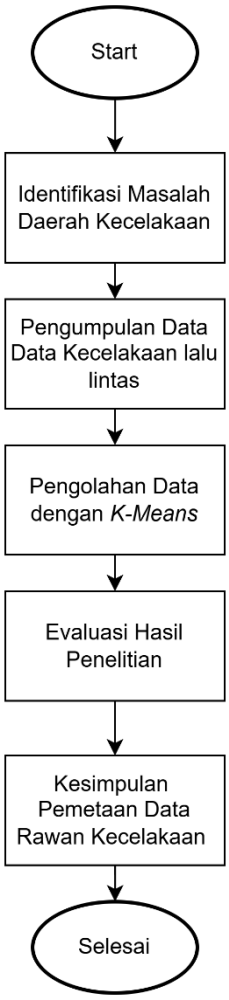
Tahap Pengolahan Data

Informasi yang telah terkumpul diproses untuk dikelompokkan atau dilaksanakan proses pengelompokan menggunakan Algoritma *K-Means*. Proses pengelompokan ini dilakukan dengan menentukan tiga kategori, yaitu rendah, sedang, dan tinggi, berdasarkan jumlah kejadian kecelakaan lalu lintas pada setiap bulan di wilayah tertentu. Data akan dipisahkan per bulan dan dianalisis untuk menentukan kategori:

1. Klaster Rendah: Bulan-bulan dengan jumlah kecelakaan relatif sedikit dibandingkan rata-rata total kejadian dalam setahun.
2. Klaster Sedang: Bulan-bulan dengan jumlah kecelakaan mendekati rata-rata total kejadian.
3. Klaster Tinggi: Bulan-bulan dengan jumlah kecelakaan jauh di atas rata-rata total kejadian.

Tahap Klasterisasi

Tahapan di bagian ini adalah ketika kerangka kerja digunakan untuk menggambarkan proses yang teratur dalam penelitian ini, mulai dari analisis kebutuhan data hingga penarikan kesimpulan hasil [13]. Alur proses Algoritma *K-Means* dalam penelitian ini dapat dijelaskan melalui konsep bagan berikut.



Gambar 1. *Flowchart* Alur kerja Penelitian

Tahap Analisis

Pada poin ini, dilakukan penilaian mengenai frekuensi kecelakaan jalan raya yang terjadi di Kota Yogyakarta dengan menggunakan informasi yang dihimpun dari kumpulan data yang ada. Evaluasi pengelompokan dilaksanakan untuk memahami seberapa tinggi mutu dari hasil pengelompokan [14]. Data ini mencatat total kecelakaan yang terjadi setiap bulan tanpa membedakan kategori korban. Data yang diperoleh kemudian dikelompokkan ke dalam tiga klaster: rendah, sedang, dan tinggi, berdasarkan frekuensi kejadian kecelakaan yang terjadi di masing-masing bulan. Analisis ini bertujuan untuk menemukan pola kecelakaan yang terjadi di Yogyakarta dan wilayah yang berisiko tinggi terhadap kecelakaan.

3. Hasil Dan Pembahasan

Dalam penelitian ini, terdapat dua belas kumpulan data yang masing-masing memiliki 2 variabel, yaitu informasi mengenai bulan di Provinsi Daerah Istimewa Yogyakarta pada tahun 2023. Data yang dikumpulkan mencakup total kecelakaan beserta jumlah korban setiap bulan selama periode satu tahun. Selanjutnya dikelompokkan 3 *cluster*:

Tabel 1. Data Kecelakaan Bulanan DIY Tahun 2023

_id	Bulan	Jumlah
1	Januari	64

2	Februari	50
3	Maret	64
4	April	55
5	Mei	78
6	Juni	59
7	Juli	59
8	Agustus	79
9	September	72
10	Oktober	81
11	Nopember	100
12	Desember	69

Tabel 2. Contoh Data Kecelakaan Terendah, Sedang, dan Tertinggi di DIY

_id	Bulan	Jumlah
2	Februari	50
3	Maret	69
11	Nopember	100

1. Hitung jarak antara data dan pusat *Euclidean* menggunakan rumus berikut:
d(x,y) = |x - y| = $\sqrt{\sum_{i=0}^{n}(xi - yi)^2}$ (1)

Dimana:
D(x,y)j = Jarak antara dua objek, yaitu x dan y
n = Ukuran dari data
Xi = Posisi dari objek xi
Xj = Posisi dari objek yi
Hasil ini didapatkan dari penghitungan pada iterasi yang pertama..

Tabel 3. Sebagai berikut adalah hasil perhitungan jarak.

_id	Bulan	Jumlah	*Distance from* m1	*Distance from* m2	*Distance from* m3	*Minimum Distance*	*Cluster Membership*	*(Minimum Distance)^2*
1	Januari	64	14	5	36	5	c2	25
2	Februari	50	0	19	50	0	c1	0
3	Maret	64	14	5	36	5	c2	25
4	April	55	5	14	45	5	c1	25
5	Mei	78	28	9	22	9	c2	81
6	Juni	59	9	10	41	9	c1	81
7	Juli	59	9	10	41	9	c1	81
8	Agustus	79	29	10	21	10	c2	100
9	September	72	22	3	28	3	c2	9

10	Oktober	81	31	12	19	12	c2	144
11	Nopember	100	50	31	0	0	c3	0
12	Desember	69	19	0	31	0	c2	0

2. Mengelompokkan data berdasarkan *cluster*. Mengelompokkan data ke dalam *cluster-cluster* yang berisi data dengan jarak terpendek atau terkecil. Jumlah jarak yang menunjukkan bahwa data paling dekat berada dalam satu kelompok dengan pusat *cluster* [15]. Seperti contoh Tabel **diatas (tulis urutan tabel)** terlihat jarak data ke centroid 2 lebih kecil dibandingkan *centroid* 1 dan 3. Lihat tabel di bawah untuk contoh pengelompokan data berdasarkan jarak terdekat:

Tabel 4. Pengelompokkan data *Membership*

Cluster Membership
c2
c1
c2
c1
c2
c1
c1
c2
c2
c2
c3
c2

3. Proses ulang ke langkah dua menggunakan *centroid* yang baru dari iterasi awal, yang dihitung berdasarkan nilai rata-rata dari masing-masing grup *cluster*. *Centroid* baru ditetapkan dengan membagi total semua nilai atribut pada *centroid* dengan total jumlah data dan diterapkan pada seluruh atribut *centroid*. Misalnya untuk atribut merek fokus pertama:

Tabel 5. *Centroid* Iterasi Pertama

	x
M1	55.75
M2	72.42857
M3	100

Tabel 6. *Centroid* Iterasi Kedua

	x
M1	58.5

M2	75.8
M3	100

Proses selanjutnya diulangi dengan fokus baru, sehingga menghasilkan nilai rata-rata yang tetap sama. Kondisi ini membuat proses perhitungan di penelitian ini dengan sudut pandang baru terhenti pada iterasi keempat. Pengamatan terakhir yang masih dapat ditemukan pada tabel berikut (terletak di bawah tabel):

Tabel 7. Pusat Akhir dari Iterasi Keempat

	x
M1	59.28571
M2	75.8
M3	100

Di bawah ini adalah hasil iterasi terakhir:

Tabel 8. Hasil Akhir Iterasi

_id	Bulan	Jumlah	*Distance from* m1	*Distance from* m2	*Distance from* m3	*Minimum Distance*	*Cluster Membership*	*(Minimum Distance)^2*
1	Januari	64	5.5	11.8	36	5.5	c1	30.25
2	Februari	50	8.5	25.8	50	8.5	c1	72.25
3	Maret	64	5.5	11.8	36	5.5	c1	30.25
4	April	55	3.5	20.8	45	3.5	c1	12.25
5	Mei	78	19.5	2.2	22	2.2	c2	4.84
6	Juni	59	0.5	16.8	41	0.5	c1	0.25
7	Juli	59	0.5	16.8	41	0.5	c1	0.25
8	Agustus	79	20.5	3.2	21	3.2	c2	10.24
9	September	72	13.5	3.8	28	3.8	c2	14.44
10	Oktober	81	22.5	5.2	19	5.2	c2	27.04
11	Nopember	100	41.5	24.2	0	0	c3	0
12	Desember	69	10.5	6.8	31	6.8	c2	46.24

4. *Davies-Bouldin Index*

Fase ini menggambarkan cara melakukan validasi internal pada *cluster* berdasarkan kepadatan data yang spesifik, perbedaan antara *cluster*, dan perbandingan antar cluster. Nilai dari ketiga elemen ini menjadi acuan untuk menetapkan indeks *Davies-Bouldin*. Proses yang dilalui meliputi perhitungan *Sum of Squares Within-cluster* (SSW), dilanjutkan dengan menghitung *Sum of Squares Between-cluster* (SSB), dan diakhiri dengan memperhitungkan Rasio dan DBI.

Langkah pertama untuk mendapatkan SSW menggunakan rumus yang dijelaskan sebagai berikut

$$SSW_i = \frac{1}{mi} \sum_{j=i}^{mi} d\,(x_j, c_i) \qquad (2)$$

Dimana:

Mi : total data pada kelompok yang ke i
Ci : pusat kelompok yang ke i
D(xj, ci) : jarak antara data yang ke j ke pusat kelompok yang ke i

Penghitungan SSW dilakukan dengan cara menentukan jarak antara setiap titik data dan *centroid*, kemudian menghitung rata-ratanya. Hasil lengkap dari perhitungan SSW dapat ditemukan dalam tabel di bawah ini:

Tabel 9. Nilai SSW untuk Setiap *Cluster*

SSW 1	3.928567
SSW 2	4.24
SSW 3	0

Setelah diperoleh nilai SSW, tentukan nilai SSB diperoleh dengan mengukur jarak antara *centroid* kluster menggunakan rumus berikut ini:
SSBij = d(ci,cj)
Hasil perhitungan SSB secara menyeluruh ditampilkan dalam tabel berikut:

Tabel 10. Jarak Antar *Centroid* (SSB)

	Centroid		
SSB	1	2	3
1	0	16.5	1657.7
2	16.5	0	585.6
3	1657.7	585.6	0

Setelah nilai SSB diperoleh, langkah berikutnya adalah mencari nilai rasio dan DBI. Hasil dari perhitungan rasio dan DBI disajikan dalam tabel berikut:

Tabel 11. Menunjukkan hasil Rasio dan DBI

R	1	2	3	R Max	DBI
1	0	0.494636	0.00237	0.49464	
2	0.494636	0	0.00724	0.49464	
3	0.00237	0.00724	0	0.00724	0.33217

Semakin rendah nilai DBI yang didapatkan (non-negatif >=0), maka kualitas *cluster* yang dihasilkan oleh algoritma *clustering* semakin baik. Berdasarkan perhitungan Rasio dan DBI, nilai DBI yang diperoleh adalah 0,440559459, yang menunjukkan bahwa hasil *cluster* tersebut sudah cukup memuaskan.

4. Kesimpulan

Penelitian ini berhasil mengidentifikasi area berisiko tinggi untuk Kecelakaan di jalan raya di Kota Yogyakarta menggunakan algoritma *K-Means*, berdasarkan data kecelakaan tahun 2023 yang dikelompokkan ke dalam tiga kelompok: rendah, sedang, dan tinggi. Hasil pengelompokan menunjukkan bahwa metode ini berhasil dalam mengatur data sesuai dengan pola kecelakaan, dengan iterasi yang menghasilkan *centroid* stabil. Studi ini menyajikan sumbangan yang signifikan untuk memahami pola kecelakaan di daerah Yogyakarta, sehingga

dapat dijadikan sebagai landasan untuk pengambilan keputusan yang lebih tepat dalam upaya meningkatkan keselamatan di jalan. Namun, penelitian ini memiliki keterbatasan pada penggunaan variabel data yang terbatas, sehingga hasil klasterisasi mungkin belum mencakup semua faktor yang memengaruhi kecelakaan lalu lintas. Penelitian selanjutnya diharapkan dapat mempertimbangkan variabel tambahan, seperti faktor cuaca, kondisi jalan, dan perilaku pengemudi, untuk analisis yang lebih komprehensif, serta mengeksplorasi penggabungan algoritma lain untuk meningkatkan akurasi. Dengan demikian, hasil penelitian ini diharapkan dapat mendukung upaya mitigasi kecelakaan di jalan raya dan membangun suasana berkendara yang lebih aman untuk masa depan.

Daftar Pustaka

[1] 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. 10, p. 295, 2020, [Online]. Available: https://doi.org/10.5281/zenodo.6820090

[2] G. S. B. M. Mina Yumei Santi, "Karakteristik Kecelakaan Lalu Lintas Dan Lokasi Black Spot Di Kab. Cilacap," *J. Tek. Sipil*, vol. 12, no. 4, pp. 259–266, 2016, doi: 10.24002/jts.v12i4.634.

[3] C. A. N. Sari and B. Afriandini, "Analysis of Traffic Accident Rates to Improve Road Safety in Yogyakarta City," *CIVeng J. Tek. Sipil dan Lingkung.*, vol. 2, no. 1, pp. 37–42, 2021, [Online]. Available: http://jurnalnasional.ump.ac.id?index.php/civeng

[4] K. R. K. Yogyakarta, "Jumlah Kecelakaan Lalu Lintas di Kota Yogyakarta Tahun 2023." https://dataset.jogjakota.go.id/dataset/kyda2023-113/resource/a2735ca4-d648-48d4-9fd2-20b61918e118 (accessed Dec. 11, 2024).

[5] A. Nur Khormarudin *et al.*, "Teknik Data Mining: Algoritma K-Means Clustering," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 1, no. 2, pp. 116–123, 2022, [Online]. Available: https://djournals.com/klik%0Ahttps://ilmukomputer.org/category/datamining/

[6] B. Harahap and A. Rambe, "Implementasi K-Means Clustering Terhadap Mahasiswa yang Menerima Beasiswa Yayasan Pendidikan Battuta di Universitas Battuta Tahun 2020/2021 Studi Kasus Prodi Informatika," *Informatika*, vol. 9, no. 3, pp. 90–97, 2021, doi: 10.36987/informatika.v9i3.2185.

[7] D. D. Darmansah and N. W. Wardani, "Analisis Pesebaran Penularan Virus Corona di Provinsi Jawa Tengah Menggunakan Metode K-Means Clustering," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 1, pp. 105–117, 2021, doi: 10.35957/jatisi.v8i1.590.

[8] F. Lende and L. L. Momo, "Penerapan K-Means Clustering Terhadap Kinerja Guru Menggunakan Metodologi Perencanaan Strategis Sistem Informasi (Studi kasus SD Negeri Lokokaki)," vol. 9, no. 4, pp. 61–68, 2023.

[9] E. Purwaningsih, "Analisis Kecelakaan Berlalu Lintas Di Kota Jakarta Dengan Menggunakan Metode K-Means," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 5, no. 1, pp. 139–144, 2019, doi: 10.33480/jitk.v5i1.712.

[10] N. Lucyana, R. Passarella, D. Kurniawan, P. Sari, and M. A. Buchari, "Analisis Penyebab Kecelakaan Pesawat Di Indonesia Menggunakan Metode K-Means," *J. Sist. Komput. Musirawas*, vol. 7, no. 2, pp. 89–106, 2022.

[11] E. A. Saputra and Y. Nataliani, "Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means," *J. Inf. Syst. Informatics*, vol. 3, no. 3, pp. 424–439, 2021, doi: 10.51519/journalisi.v3i3.164.

[12] M. Produk *et al.*, "Penerapan Metode K-Means Pada Data Penjualan Untuk," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 5, no. 1, pp. 228–236, 2023.

[13] A. Mubarok *et al.*, "Penerapan Algoritma K-Means Untuk Menentukan Kelas," vol. 23, no. November, 2021.
[14] E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *J. Sains dan Manaj.*, vol. 9, no. 1, p. 96, 2021, [Online]. Available: www.bps.go.id
[15] P. Alkhairi and A. P. Windarto, "Penerapan K-Means Cluster pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara," *Semin. Nas. Teknol. Komput. Sains*, pp. 762–767, 2019, [Online]. Available: https://prosiding.seminar-id.com/index.php/sainteks/article/download/228/223