

Penerapan Algoritma K-Means Clustering untuk Segmentasi Kepadatan Penduduk Berbasis GIS

Rizki Amelia Putri*, Safwandi, Zahratul Fitri

Fakultas Teknik, Program Studi Teknik Informatika, Universitas Malikussaleh, Lhokseumawe, Indonesia

Email: ¹*rizki.200170034@mhs.unimal.ac.id, ²safwandi@unimal.ac.id, ³zahratul.fitri@unimal.ac.id

Email Penulis Korespondensi: rizki.200170034@mhs.unimal.ac.id

Submitted 26-05-2025; Accepted 28-06-2025; Published 30-06-2025

Abstrak

Penelitian ini bertujuan untuk mengembangkan sistem *clustering* menggunakan algoritma *K-means* untuk menganalisis data demografis kecamatan dari tahun 2020 hingga 2023. Sistem ini dirancang untuk mengelompokkan kecamatan berdasarkan variabel seperti jumlah penduduk, persentase penduduk, kepadatan penduduk, dan rasio jenis kelamin. Hasil *clustering* menunjukkan pola pengelompokan yang berbeda setiap tahun, mencerminkan dinamika data demografis yang terjadi. Evaluasi menggunakan *Davies-Bouldin Index* (DBI) menunjukkan bahwa hasil *clustering* memiliki kualitas yang cukup baik, dengan nilai DBI sebesar 1.1492 pada tahun 2020, 0.6859 pada tahun 2021, 1.2470 pada tahun 2022, dan 0.6805 pada tahun 2023. Nilai DBI terbaik tercatat pada tahun 2023 dengan 0.6805, menunjukkan bahwa hasil *clustering* pada tahun tersebut memiliki kualitas yang paling optimal dibandingkan tahun lainnya. Sistem ini juga mempermudah pengguna dengan visualisasi peta interaktif, sehingga mendukung analisis data dan pengambilan keputusan yang lebih baik. Hasil penelitian ini diharapkan dapat berkontribusi dalam pengelolaan data demografis dan mendukung pembuatan kebijakan berbasis data yang lebih akurat.

Kata Kunci: *Clustering*; Data Mining; Kepadatan Penduduk; Algoritma *K-means*; Sistem Informasi Geografis

Abstract

This study aims to develop a clustering system using the *K-means* algorithm to analyze demographic data of sub-districts from 2020 to 2023. The system is designed to cluster sub-districts based on variables such as population size, population percentage, population density, and gender ratio. The clustering results reveal different grouping patterns each year, reflecting the dynamics of demographic data over time. Evaluation using the Davies-Bouldin Index (DBI) indicates that the clustering results are of reasonably good quality, with DBI values of 1.1492 in 2020, 0.6859 in 2021, 1.2470 in 2022, and 0.6805 in 2023. The best DBI value was recorded in 2023 at 0.6805, demonstrating that the clustering results in that year were the most optimal compared to other years. The system also facilitates Users with interactive map visualizations, supporting better data analysis and decision-making processes. This research is expected to contribute to the management of demographic data and support more accurate data-driven policy-making.

Keywords: Clustering; Data Mining; Dense Population; *K-means* algorithm; Geographic Information System

1. PENDAHULUAN

Kepadatan penduduk adalah jumlah penduduk rata-rata per kilometer persegi di setiap wilayah. Jumlah penduduk di setiap wilayah tidak sama [1]. Tingginya tingkat kepadatan di suatu daerah sering kali merefleksikan intensitas aktivitas ekonomi, dinamika sosial, serta proses urbanisasi yang berkembang pesat. Namun demikian, apabila tidak dikelola secara optimal, kepadatan penduduk yang tinggi dapat menimbulkan berbagai permasalahan struktural, seperti meningkatnya tekanan terhadap infrastruktur, kemacetan lalu lintas, penurunan kualitas lingkungan, serta keterbatasan akses terhadap pelayanan publik yang memadai. Dua faktor yang mempengaruhi kepadatan penduduk: faktor internal daerah, seperti pertumbuhan penduduk yang relatif tinggi (kelahiran lebih tinggi daripada kematian), dan wilayah yang strategis; dan faktor eksternal, seperti banyaknya penduduk yang berpindah dan wilayah yang kurang produktif di sekitarnya [2]. Berdasarkan data yang dipublikasikan oleh Badan Pusat Statistik (BPS) Kota Lhokseumawe, laju pertumbuhan penduduk menunjukkan peningkatan dari tahun ke tahun, dengan persebaran yang terjadi di seluruh kecamatan di wilayah kota tersebut [3]. Sebagai salah satu kota utama di Provinsi Aceh, Kota Lhokseumawe mengalami pertumbuhan jumlah penduduk yang cukup signifikan dalam beberapa tahun terakhir. Kondisi ini mendorong perlunya kajian yang lebih mendalam terkait pola distribusi dan segmentasi kepadatan penduduk, khususnya pada tingkat kecamatan, sebagai dasar dalam merumuskan kebijakan pembangunan yang berbasis data dan berorientasi pada keberlanjutan.

Dengan kemajuan teknologi informasi, pendekatan tradisional dalam menganalisis data kependudukan kini dianggap kurang efektif. Penerapan metode data mining menjadi solusi yang relevan karena mampu mengolah dan menggali informasi dari data dalam jumlah besar dan kompleks. Data mining merupakan proses eksplorasi dan analisis data yang bertujuan untuk menemukan pola-pola yang bermanfaat, hubungan bersembunyi, dan informasi yang dapat diambil dari sekumpulan data besar [4]. Secara sederhana data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar [5]. Data mining juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data [6]. Salah satu teknik yang umum digunakan dalam data mining adalah *clustering*, atau pengelompokan, yang bertujuan untuk mengidentifikasi kelompok-kelompok dengan karakteristik serupa. Teknik ini mencari struktur dalam data tanpa memerlukan label kelas yang telah ditentukan sebelumnya [7]. Clustering tidak sama dengan klasifikasi; klasifikasi adalah penempatan data baru ke salah satu kategori (kelas) yang telah didefinisikan sebelumnya [8]. Analisis cluster digunakan untuk mengkategorikan objek ke dalam kelompok yang relatif serupa, di mana objek-objek dalam setiap kelompok cenderung mirip satu sama lain dan berbeda dari objek-objek dalam

kelompok lain. Dalam analisis cluster, objek-objek yang memiliki karakteristik yang mirip berada dalam satu kelompok, sedangkan karakteristik antar kelompok berbeda [9]. Konsep centroid dalam *clustering* sangat penting untuk menilai keanggotaan dan konektivitas lokal dalam kluster [10]. Salah satu teknik *clustering* yang paling sederhana dan umum digunakan adalah Algoritma *K-means*. *K-Means clustering* merupakan sebuah algoritma unsupervised learning yang digunakan dalam pengelompokan data dalam dataset yang tidak memiliki label kedalam sebuah cluster-cluster yang berbeda [11]. Algoritma ini mengelompokkan objek berdasarkan jarak, dan memiliki ketelitian yang cukup tinggi terhadap ukuran objek. Hal ini membuat Algoritma *K-means* relatif lebih terukur dan efisien dalam proses pengelompokannya [12]. Hasil *clustering* dari metode *K-means* sangat bergantung pada nilai awal yang dipilih untuk pusat-pusat kluster [13].

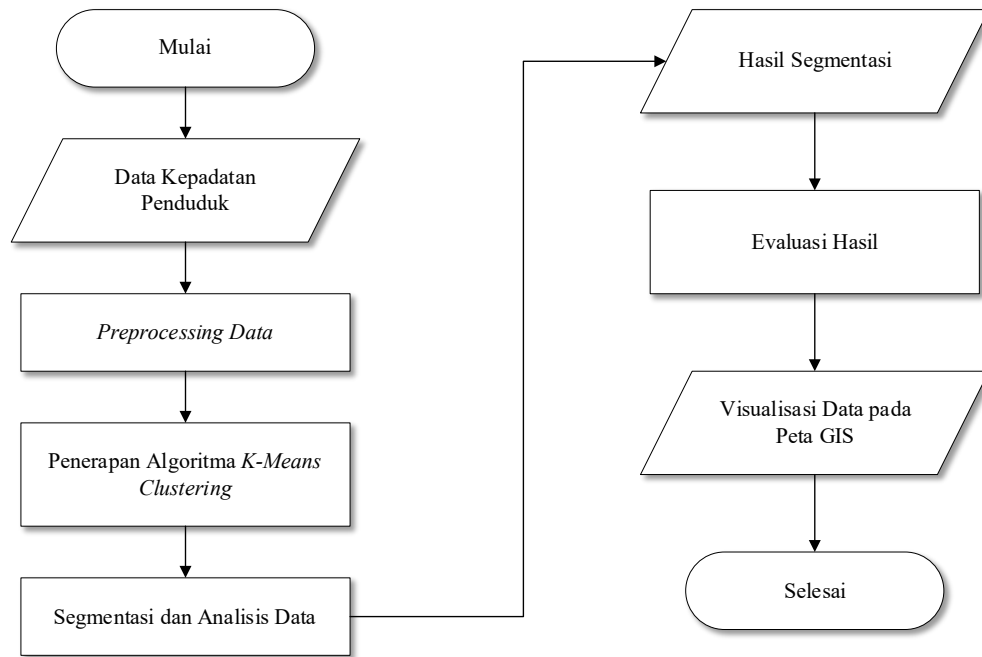
Penelitian terdahulu yang menjadi referensi dalam penelitian ini melibatkan lima studi terkait algoritma *K-means* dan memiliki hasil yang berbeda. Salah satu contohnya seperti penerapan *K-means* untuk pemetaan daerah rawan kecelakaan lalu lintas, di mana penelitian ini menghasilkan tiga kluster daerah rawan kecelakaan di Kota Malang, yaitu sangat rawan (SR), rawan (R), dan cukup rawan (CR) [14]. Penelitian kedua menerapkan *K-means clustering* pada pemetaan lahan kopi di Kabupaten Malang, menghasilkan kluster rendah (C1), sedang (C2), dan tinggi (C3). Data yang dianalisis dari tahun 2018 2020 menunjukkan bahwa mayoritas wilayah berada dalam kluster rendah dengan jumlah data 26 (2018), 24 (2019), dan 24 (2020). Kluster sedang memiliki 3 data (2018), 5 data (2019), dan 4 data (2020), sedangkan kluster tinggi masing-masing 4 data (2018 dan 2019) dan 5 data (2020) [15]. Penelitian ketiga yaitu penggunaan algoritma *K-means Clustering* untuk pemetaan kompetensi Teknologi Informasi dan Komunikasi (TIK) siswa SMKN 3 Lhokseumawe menghasilkan tiga kelompok, yaitu *cluster* "Sangat Kompeten" dengan 80 siswa, *cluster* "Kompeten" dengan 64 siswa, dan *cluster* "Kurang Kompeten" dengan 10 siswa [16]. Sementara itu, studi lain menerapkan pengklasteran kemiskinan kabupaten/kota di Provinsi Banten menggunakan metode *K-means* menghasilkan dua kluster dengan nilai struktur terbaik berdasarkan silhouette index. Kluster pertama terdiri dari Kota Tangerang, Kota Cilegon, dan Kota Tangerang Selatan, dengan sektor kemiskinan yang berpengaruh adalah persentase penduduk miskin usia 15 tahun ke atas yang bekerja di sektor informal (X6). Kluster kedua mencakup Kabupaten Pandeglang, Kabupaten Lebak, Kabupaten Tangerang, 3 Kabupaten Serang, dan Kota Serang, dengan faktor kemiskinan yang berpengaruh adalah persentase penduduk miskin usia 15 tahun ke atas dengan pendidikan di bawah SD (X1), tamat SD/SMP (X2), tamat SMA ke atas (X3), angka melek huruf usia 15-55 tahun (X4), persentase penduduk miskin usia 15 tahun ke atas yang bekerja di sektor formal (X7), dan persentase pengeluaran per kapita untuk makanan (X10). Pembentukan dua kluster ini menunjukkan akurasi terbaik dibandingkan dengan pembentukan tiga atau empat kluster [17]. Kemudian penelitian menggunakan metode *k means* pada kemiskinan di Jawa Timur tahun 2020. Dengan menggunakan tools Orange, hasil *clustering* menghasilkan tiga kluster, kluster 1 terdiri dari 6 kota/kabupaten dengan tingkat kemiskinan terendah sebesar 15%, kluster 2 mencakup 10 kota/kabupaten dengan tingkat kemiskinan menengah sebesar 26%, dan kluster 3 terdiri dari 23 kota/kabupaten dengan tingkat kemiskinan tertinggi sebesar 59% [18].

Penelitian ini berkontribusi melalui penerapan algoritma *K-Means* dalam mengelompokkan wilayah-wilayah di Kota Lhokseumawe berdasarkan tingkat kepadatan penduduk. Hal ini membantu dalam mengidentifikasi pola persebaran penduduk yang tidak mudah dikenali hanya dari data mentah. Dalam penelitian ini, algoritma *K-means Clustering* diterapkan untuk segmentasi kepadatan penduduk di Kota Lhokseumawe yang mencakup 4 Kecamatan diantaranya, Blang Mangat, Muara Dua, Muara Satu, dan Banda Sakti, dengan menggunakan data geospasial berbasis Geographic Information System (GIS). Sistem Informasi Geografis (SIG) merupakan sistem yang menyatukan komponen data tekstual dengan data peta sehingga dengan dipilihnya suatu lokasi dalam peta maka akan ditampilkan informasi yang berhubungan dengan lokasi yang diinginkan [19]. Penelitian ini memadukan algoritma *K-Means Clustering* dengan GIS untuk menghasilkan visualisasi tematik, yang membantu memperlihatkan distribusi spasial dari hasil segmentasi secara lebih jelas. Sebagian besar studi terdahulu yang menggunakan algoritma *K-Means* untuk segmentasi wilayah belum memanfaatkan integrasi dengan GIS, sehingga hasilnya terbatas pada representasi tabel dan grafik. Penelitian ini menjawab kekurangan tersebut dengan menggabungkan algoritma *K-Means* dan sistem visualisasi spasial berbasis GIS, memungkinkan penyajian hasil *clustering* dalam bentuk peta interaktif. Pemanfaatan GIS memungkinkan hasil segmentasi kepadatan penduduk dikaitkan langsung dengan posisi geografis suatu wilayah, sehingga pola persebaran dan konsentrasi penduduk dapat diidentifikasi dengan lebih jelas. Penggunaan GIS memberikan nilai tambah dalam visualisasi dan analisis data karena mampu memetakan dan menampilkan distribusi penduduk secara interaktif dan lebih informatif. Hasil dari segmentasi yang dilakukan dapat memberikan wawasan yang lebih mendalam mengenai karakteristik kepadatan penduduk di Kota Lhokseumawe dan diharapkan dapat menjadi dasar oleh pemerintah daerah Kota Lhokseumawe dalam pengambilan keputusan, seperti pemerataan pembangunan, perencanaan fasilitas umum, dan mitigasi risiko kepadatan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini memiliki tahapan-tahapan yang dapat dilihat pada Gambar 1 berikut.



Gambar 1. Tahapan Penelitian

Berikut adalah penjelasan rinci untuk setiap tahapan dalam penelitian ini:

1. **Mulai**
Penelitian dimulai dengan mengidentifikasi kebutuhan data dan tujuan akhir, yaitu segmentasi kepadatan penduduk di Kota Lhokseumawe menggunakan algoritma *K-means Clustering* berbasis GIS.
2. **Data Kepadatan Penduduk**
Pada tahap ini, data kependudukan yang mencakup informasi seperti jumlah penduduk, luas wilayah, dan parameter lain yang relevan dikumpulkan dari sumber resmi, seperti situs BPS Kota Lhokseumawe.
3. **Preprocessing Data**
Data yang telah dikumpulkan diproses untuk memastikan kualitasnya, termasuk penanganan data yang hilang, duplikat, atau tidak konsisten. Data kemudian dinormalisasi agar sesuai untuk proses *clustering*.
4. **Penerapan Algoritma *K-means Clustering***
Setelah *preprocessing* selesai, algoritma *K-means Clustering* diterapkan untuk mengelompokkan data ke dalam beberapa klaster berdasarkan kepadatan penduduk. Proses ini dimulai dengan inisialisasi *centroid* awal, diikuti dengan iterasi hingga hasil klasterisasi stabil.
5. **Segmentasi dan Analisis Data**
Hasil dari algoritma *K-means* digunakan untuk segmentasi wilayah Kota Lhokseumawe ke dalam beberapa kelompok berdasarkan kepadatan penduduk. Data yang telah dikelompokkan dianalisis untuk memahami pola distribusi penduduk.
6. **Hasil Segmentasi**
Output dari algoritma *K-means* adalah hasil segmentasi wilayah, yang menunjukkan klaster berdasarkan tingkat kepadatan penduduk (misalnya: rendah, sedang, tinggi).
7. **Evaluasi Hasil**
Hasil klasterisasi dievaluasi untuk memastikan keakuratannya menggunakan metode seperti uji *silhouette score* atau evaluasi lain yang sesuai. Evaluasi ini bertujuan memastikan bahwa hasil klasterisasi valid dan relevan.
8. **Visualisasi Data pada Peta GIS**
Data hasil segmentasi divisualisasikan dalam bentuk peta interaktif berbasis GIS. Peta ini memberikan gambaran yang mudah dipahami mengenai distribusi penduduk di berbagai wilayah di Kota Lhokseumawe.
9. **Selesai**
Hasil akhir berupa segmentasi wilayah dan visualisasi peta GIS dapat digunakan untuk mendukung pengambilan keputusan atau penelitian lebih lanjut.

2.2 *K-means*

K-means adalah salah satu teknik *clustering partitional* di mana setiap titik data harus dimasukkan ke dalam satu kluster tertentu dan dapat berpindah antar kluster pada setiap iterasi proses. Algoritma ini membagi data ke dalam k kluster terpisah, di mana k merupakan bilangan bulat positif yang telah ditentukan sebelumnya. *K-means* dikenal karena keunggulannya dalam mengelompokkan data besar dan mengatasi *outlier* dengan efisien dan cepat [20]. Langkah-langkah melakukan *clustering* dengan metode *K-means Clustering* adalah sebagai berikut [21]:

1. Menentukan nilai k untuk jumlah *cluster* yang ingin dibentuk.

2. Memilih nilai untuk pusat *cluster* awal (*centroid*) sebanyak k .
3. Menghitung jarak setiap data *input* terhadap masing-masing *centroid* menggunakan rumus jarak *Euclidean* (*Euclidean Distance*) hingga ditemukan jarak yang paling dekat dari setiap data dengan *centroid*. Berikut adalah persamaan *Euclidian Distance* :

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2} \quad (1)$$

Keterangan :

x_i : Data kriteria ke

μ_j : *Centroid cluster* ke- j

$\sum$: Penjumlahan

4. Mengklasifikasikan/mengelompokkan setiap data berdasarkan kedekatannya dengan *centroid* (jarak terkecil).
5. Memperbaharui nilai *centroid*. Nilai *centroid* baru di peroleh dari rata-rata *cluster* yang bersangkutan dengan menggunakan rumus :

$$\mu_j(t+1) = \frac{1}{N_{s_j}} \sum_{j \in s_j} x_j \quad (2)$$

Keterangan :

$\mu_j(t+1)$: *Centroid* baru pada iterasi ke-($t+1$).

N_{s_j} : Banyak data pada *cluster* S_j .

6. Melakukan perulangan dari langkah 3 hingga 5, sampai anggota tiap *cluster* tidak ada yang berubah. Jika langkah 6 telah terpenuhi, maka nilai pusat *cluster* (μ_j) pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Penelitian ini memanfaatkan data kependudukan Kota Lhokseumawe dari tahun 2020 hingga 2023. Data tersebut mencakup informasi jumlah penduduk, laju pertumbuhan penduduk, persentase penduduk, kepadatan penduduk per km², dan rasio jenis kelamin penduduk di empat kecamatan: Blang Mangat, Muara Dua, Muara Satu, dan Banda Sakti. Selanjutnya, variabel yang digunakan dikonversi dalam bentuk inisial untuk memudahkan proses pengolahan data nantinya, adapun konversi dari variabel tersebut adalah sebagai berikut.

a. Inisialisasi Variabel

Inisialisasi ini digunakan dalam proses segmentasi dengan algoritma *K-means Clustering* untuk mempermudah interpretasi hasil pengelompokan. Dengan variabel-variabel tersebut, penelitian dapat mengidentifikasi pola kepadatan penduduk di berbagai kecamatan secara lebih terstruktur dan sistematis.

Tabel 1. Inisialisasi Variabel

Nama Variabel	Inisial Variabel
Jumlah Penduduk (Ribu)	X1
Laju Pertumbuhan Penduduk per Tahun	X2
Persentase Penduduk	X3
Kepadatan Penduduk per km persegi (Km ²)	X4
Rasio Jenis Kelamin Penduduk	X5

b. Data Kependudukan Tahun 2020

Pada tahun 2020, Kecamatan Banda Sakti memiliki jumlah penduduk tertinggi dengan 77.802 jiwa dan kepadatan penduduk mencapai 6.922 penduduk per km², menjadikannya kecamatan dengan kepadatan tertinggi di Kota Lhokseumawe. Sebaliknya, Kecamatan Blang Mangat memiliki jumlah penduduk terendah, yaitu 26.162 jiwa, dengan kepadatan 466 penduduk per km².

Tabel 1. Data Kependudukan Tahun 2020

Kecamatan	X1	X2	X3	X4	X5
Blang Mangat	26162	1.57	13.86	466	99
Muara Dua	50929	1.12	26.99	881	97
Muara Satu	33820	0.37	17.92	605	100
Banda Sakti	77820	0.29	41.23	6922	98.9

c. Data Kependudukan Tahun 2021

Tahun 2021 menunjukkan adanya peningkatan jumlah penduduk di seluruh kecamatan. Banda Sakti tetap menjadi kecamatan dengan jumlah penduduk tertinggi (77.985 jiwa) dan kepadatan 6.938 penduduk per km². Blang Mangat mengalami peningkatan jumlah penduduk menjadi 26.559 jiwa dengan kepadatan 473 penduduk per km².

Tabel 3. Data Kependudukan Tahun 2021

Kecamatan	X1	X2	X3	X4	X5
Blang Mangat	26559	1.52	13.98	473	100
Muara Dua	51472	1.07	27.1	891	97
Muara Satu	33925	0.31	17.86	607	100
Banda Sakti	77985	0.23	41.06	6938	99

d. Data Kependudukan Tahun 2022

Pada tahun 2022, jumlah penduduk terus meningkat di semua kecamatan. Kecamatan Banda Sakti kembali menonjol dengan jumlah penduduk 78.256 jiwa dan kepadatan penduduk 6.962 penduduk per km². Kecamatan Blang Mangat tetap memiliki jumlah penduduk terendah dengan 26.992 jiwa, sementara kepadatan penduduknya meningkat menjadi 481 penduduk per km².

Tabel 4. Data Kependudukan Tahun 2022

Kecamatan	X1	X2	X3	X4	X5
Blang Mangat	26992	1.63	14.1	481	99.6
Muara Dua	52079	1.18	27.21	901	96.9
Muara Satu	34069	0.42	17.8	609	100.2
Banda Sakti	78256	0.35	40.89	6962	98.8

e. Data Kependudukan Tahun 2023

Tahun 2023 mencatatkan jumlah penduduk tertinggi selama periode penelitian, terutama di Kecamatan Banda Sakti dengan 79.850 jiwa dan kepadatan 7.100 penduduk per km². Blang Mangat juga mencatat peningkatan jumlah penduduk menjadi 28.146 jiwa, dengan kepadatan 500 penduduk per km². Laju pertumbuhan penduduk tertinggi dicapai oleh Blang Mangat dengan angka 2,29%.

Tabel 5. Data Kependudukan Tahun 2023

Kecamatan	X1	X2	X3	X4	X5
Blang Mangat	28146	2.29	14.36	0.5	99.6
Muara Dua	53068	2.08	27.07	0.92	97.92
Muara Satu	35003	1.74	17.85	0.63	99.75
Banda Sakti	79850	1.14	40.73	7.1	98.18

3.2 Normalisasi Data

Melakukan normalisasi data menggunakan metode Min-Max. Normalisasi diperlukan karena atribut-atribut seperti jumlah penduduk, laju pertumbuhan penduduk per tahun, persentase penduduk, kepadatan penduduk, dan rasio jenis kelamin memiliki rentang nilai yang berbeda, sehingga dapat memengaruhi hasil clustering jika tidak dinormalisasi.

Tabel 6. Hasil Normalisasi Data Tahun 2020

Kecamatan	X1	X2	X3	X4	X5
Blang Mangat	0	1	0	0	0.667
Muara Dua	0.480	0.648	0.480	0.064	0
Muara Satu	0.148	0.063	0.148	0.022	1
Banda Sakti	1	0	1	1	0.6

Tabel 7. Hasil Normalisasi Data Tahun 2021

Kecamatan	X1	X2	X3	X4	X5
Blang Mangat	0	1	0	0	1
Muara Dua	0.4844	0.6512	0.4845	0.0647	0
Muara Satu	0.1432	0.0620	0.1433	0.0207	1
Banda Sakti	1	0	1	1	0.7

Tabel 8. Hasil Normalisasi Data Tahun 2022

Kecamatan	X1	X2	X3	X4	X5
Blang Mangat	0	1	0	0	0.818
Muara Dua	0.489	0.648	0.489	0.065	0
Muara Satu	0.138	0.055	0.138	0.020	1

Banda Sakti	1	0	1	1	0.576
-------------	---	---	---	---	-------

Tabel 9. Hasil Normalisasi Data Tahun 2023

Kecamatan	X1	X2	X3	X4	X5
Blang Mangat	0	1	0	0	0.918
Muara Dua	0.4	0.817	0.482	0.064	0
Muara Satu	0.133	0.522	0.132	0.020	1
Banda Sakti	1	0	1	1	0.142

3.3 Pembentukan Jumlah Cluster

Jumlah cluster yang digunakan dalam algoritma K-means ditentukan berdasarkan jumlah data kecamatan di Kota Lhokseumawe, yang terdiri dari empat kecamatan. Dalam penelitian ini, diputuskan untuk menggunakan dua cluster, yaitu cluster Tinggi (C1) dan Rendah (C2). Keputusan untuk menetapkan dua cluster ini didasarkan pada pertimbangan bahwa jumlah data yang relatif sedikit akan lebih representatif jika dikelompokkan ke dalam dua kategori utama.

Tabel 10. Pembentukan Jumlah Cluster

Cluster	Label
C1	Rendah
C2	Tinggi

3.4 Proses Iterasi K-Means

Proses iterasi algoritma *K-means* yang digunakan untuk mengelompokkan data kecamatan di Kota Lhokseumawe ke dalam dua cluster, yaitu cluster Rendah (C1) dan Tinggi (C2). Proses iterasi ini bertujuan untuk meminimalkan total jarak antara titik data dengan *centroid cluster* yang telah ditentukan, sehingga setiap kecamatan dapat dikelompokkan ke dalam cluster yang paling sesuai berdasarkan nilai atribut yang dimiliki.

a. Menentukan Centroid Awal

Centroid awal (titik pusat) cluster ditentukan secara acak yang diperoleh dari data yang telah di normalisasi pada tabel 6 yaitu kecamatan Blang Mangat dan Muara Satu.

Tabel 11. Pembentukan *Centroid* Awal

C1	Blang Mangat	0	1	0	0	0.667
C2	Muara Satu	0.148	0.063	0.148	0.022	1

b. Menghitung jarak terdekat berdasarkan persamaan Euclidean Distance

Menghitung Distance dengan pusat Cluster pertama:

$$\sqrt{(0-0)^2 + (1-1)^2 + (0-0)^2 + (0-0)^2 + (0.667-0.667)^2} = 0$$

Menghitung Distance dengan pusat Cluster kedua:

$$d_x \sqrt{(0-0.148)^2 + (1-0.063)^2 + (0-0.148)^2 + (0-0.022)^2 + (0.667-1)^2} = 1.017$$

Perhitungan ini dilakukan pada seluruh data sampel, sehingga diperoleh hasil *clustering* iterasi ke-1 pada data sampel (data kependudukan 2020) sebagai berikut :

Tabel 12. Hasil Perhitungan Pada Iterasi Ke-1

No	C1	C2	Min	Hasil
1	0.000	1.017	0.000	C1
2	1.016	1.251	1.016	C1
3	1.017	0.000	0.000	C2
4	2.001	1.604	1.604	C2

c. Menentukan nilai *centroid* baru pada iterasi ke-2

Setelah diketahui hasil Cluster pada iterasi sebelumnya, langkah selanjutnya adalah menentukan nilai *Centeroid* baru pada iterasi ke-2.

Tabel 13. Nilai *Centroid* Baru (Iterasi Ke-2)

C1	Blang Mangat	0.240	0.824	0.240	0.032	0.333
C2	Muara Satu	0.574	0.031	0.574	0.511	0.8

- d. Menghitung kembali menggunakan *Centroid* baru pada iterasi ke-2
Menghitung Distance dengan pusat *Cluster* pertama:

$$d_x \sqrt{(0 - 0.240)^2 + (1 - 0.824)^2 + (0 - 0.240)^2 + (0 - 0.032)^2 + (0.667 - 0.333)^2} = 0.508$$

Menghitung Distance dengan pusat *Cluster* kedua:

$$d_x \sqrt{(0 - 0.574)^2 + (1 - 0.031)^2 + (0 - 0.574)^2 + (0 - 0.511)^2 + (0.667 - 0.8)^2} = 1.370$$

Perhitungan ini dilakukan pada seluruh data sampel, sehingga diperoleh hasil *clustering* iterasi ke-2 pada data sampel (data kependudukan 2020) sebagai berikut :

Tabel 14. Hasil Perhitungan Pada Iterasi Ke-2

No	C1	C2	Min	Hasil
1	0.508	1.370	0.508	C1
2	0.508	1.113	0.508	C1
3	1.021	0.802	0.802	C2
4	1.686	0.802	0.802	C2

Jika hasil perhitungannya masih ada data dengan *Cluster* yang tidak sama dengan *Cluster* sebelumnya, maka iterasi selanjutnya akan tetap dilakukan sampai tidak ada lagi *Cluster* yang berubah dan hasilnya sama semua.

- e. Menentukan nilai *centroid* baru pada iterasi ke-3

Setelah diketahui hasil *Cluster* pada iterasi sebelumnya, langkah selanjutnya adalah menentukan nilai *Centroid* baru pada iterasi ke-3.

Tabel 15. Nilai *Centroid* Baru (Iterasi Ke-3)

C1	Blang Mangat	0.240	0.824	0.240	0.032	0.333
C2	Muara Satu	0.574	0.031	0.574	0.511	0.8

- f. Menghitung kembali menggunakan *Centroid* baru pada iterasi ke-3
Menghitung Distance dengan pusat *Cluster* pertama:

$$d_x \sqrt{(0 - 0.240)^2 + (1 - 0.824)^2 + (0 - 0.240)^2 + (0 - 0.032)^2 + (0.667 - 0.333)^2} = 0.508$$

Menghitung Distance dengan pusat *Cluster* kedua:

$$d_x \sqrt{(0 - 0.574)^2 + (1 - 0.031)^2 + (0 - 0.574)^2 + (0 - 0.511)^2 + (0.667 - 0.8)^2} = 1.370$$

Perhitungan ini dilakukan pada seluruh data sampel, sehingga diperoleh hasil *clustering* iterasi ke-3 pada data sampel (data kependudukan 2020) sebagai berikut :

Tabel 16. Hasil Perhitungan Pada Iterasi Ke-3

No	C1	C2	Min	Hasil
1	0.508	1.370	0.508	C1
2	0.508	1.113	0.508	C1
3	1.021	0.802	0.802	C2
4	1.686	0.802	0.802	C2

3.5 Hasil Akhir Perhitungan *K-means*

Setelah dilakukan pengujian pada data kepadatan penduduk di Kota Lhokseumawe tahun 2020, proses perhitungan *K-means* berhasil diselesaikan pada iterasi ke-3. Hasil akhir dari perhitungan menunjukkan pembagian data kecamatan ke dalam dua cluster, yaitu cluster Rendah (C1) dan Tinggi (C2), sesuai dengan segmentasi yang telah ditentukan sebelumnya.

Tabel 17. Hasil Akhir Perhitungan *K-means* Data Sampel 2020

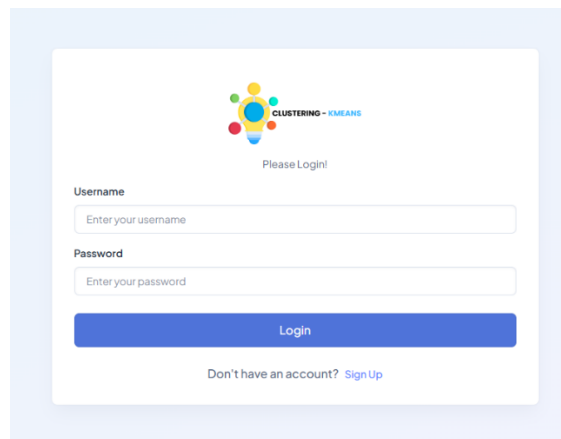
No	C1	C2	Min	Hasil
1	0.508	1.370	0.508	C1
2	0.508	1.113	0.508	C1
3	1.021	0.802	0.802	C2

4	1.686	0.802	0.802	C2
---	-------	-------	-------	----

Berdasarkan hasil akhir, kecamatan dengan tingkat kepadatan penduduk yang tinggi, seperti Blang Mangat dan Muara Dua, dikelompokkan ke dalam cluster Rendah (C1). Sementara itu, kecamatan dengan tingkat kepadatan penduduk yang lebih Tinggi, seperti Muara Satu dan Banda Sakti, dikelompokkan ke dalam cluster Tinggi (C2). Pembagian ini sesuai dengan karakteristik data yang telah dinormalisasi, di mana centroid dari masing-masing cluster telah stabil pada iterasi terakhir. Proses ini mengonfirmasi bahwa algoritma K-means mampu mengelompokkan data dengan baik, meskipun jumlah data kecamatan relatif sedikit. Hasil segmentasi ini memberikan gambaran yang jelas mengenai perbedaan tingkat kepadatan penduduk antar kecamatan, yang dapat digunakan sebagai dasar untuk analisis lebih lanjut dan perencanaan wilayah di Kota Lhokseumawe.

3.6 Implementasi Sistem

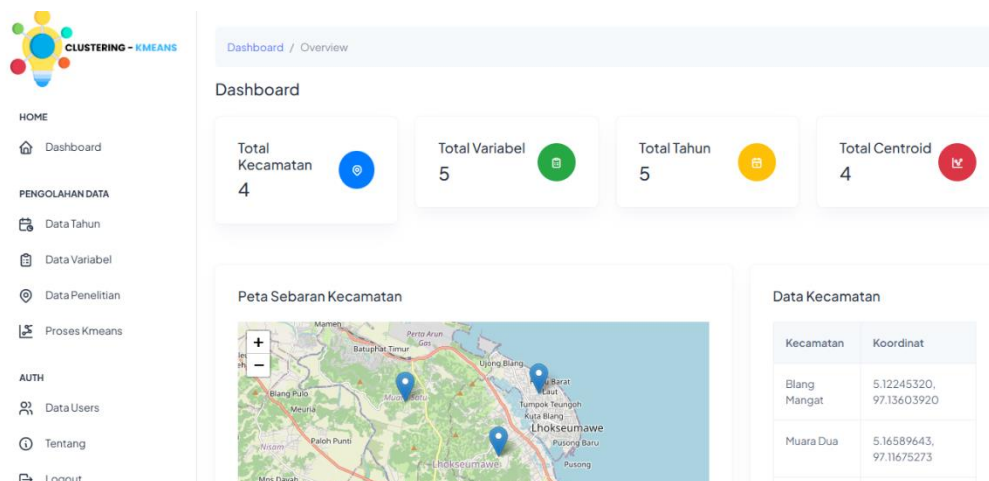
a. Halaman Login



Gambar 2. Halaman Login

Tampilan antarmuka Login ini dirancang untuk mempermudah pengguna mengakses sistem segmentasi kepadatan penduduk berbasis algoritma K-means. Pada bagian atas, terdapat logo dan judul aplikasi "Clustering - KMeans" yang memberikan identitas visual sistem.

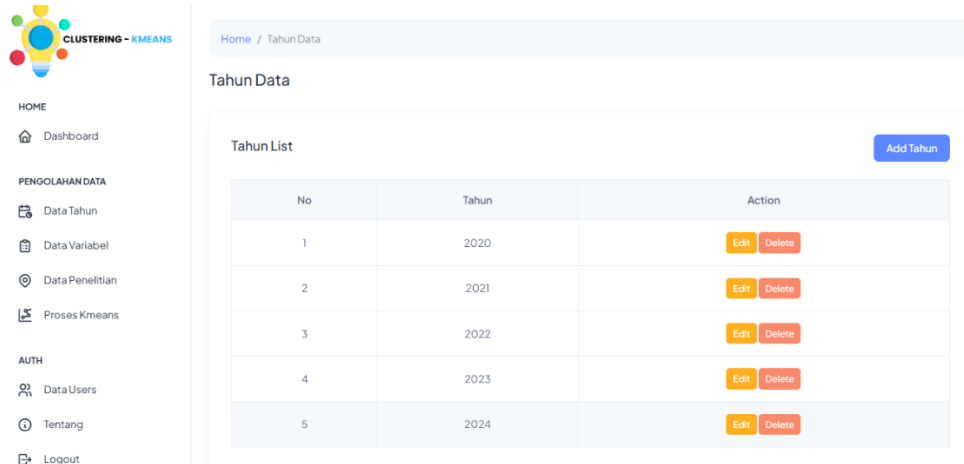
b. Halaman Dashboard



Gambar 3. Halaman Dashboard

Halaman Dashboard ini dirancang untuk memberikan gambaran umum sistem segmentasi kepadatan penduduk berbasis algoritma K-means. Pada bagian atas halaman, terdapat navigasi Dashboard / Overview yang menunjukkan lokasi pengguna dalam sistem.

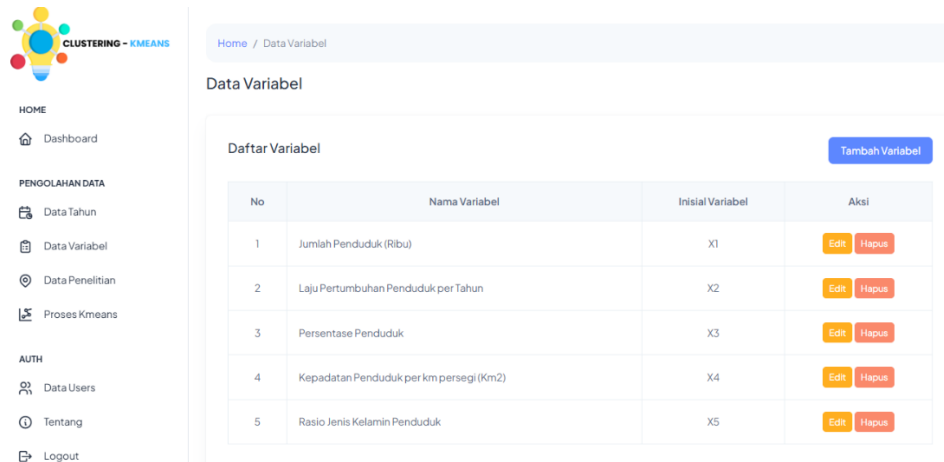
c. Halaman Data Tahun



Gambar 4. Halaman Data Tahun

Halaman Data Tahun ini dirancang untuk memudahkan pengguna dalam mengelola informasi tahun yang digunakan dalam sistem segmentasi kepadatan penduduk berbasis algoritma K-means.

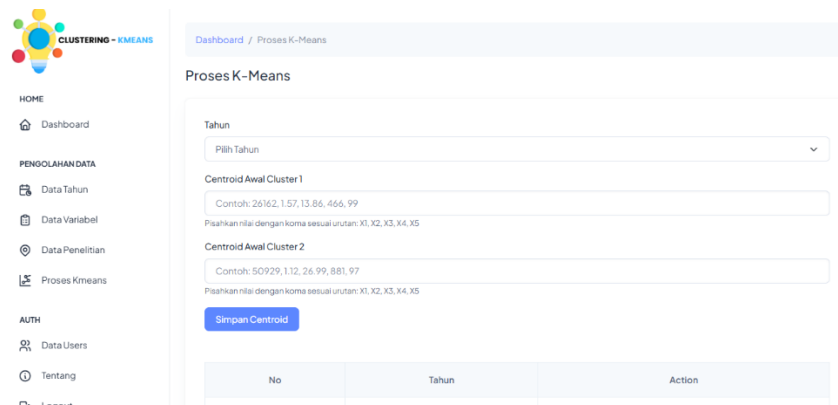
d. Halaman Data Variabel



Gambar 5. Halaman Data Variabel

Halaman Data Variabel dirancang untuk memudahkan pengguna dalam mengelola variabel yang digunakan dalam sistem segmentasi kepadatan penduduk berbasis algoritma K-means.

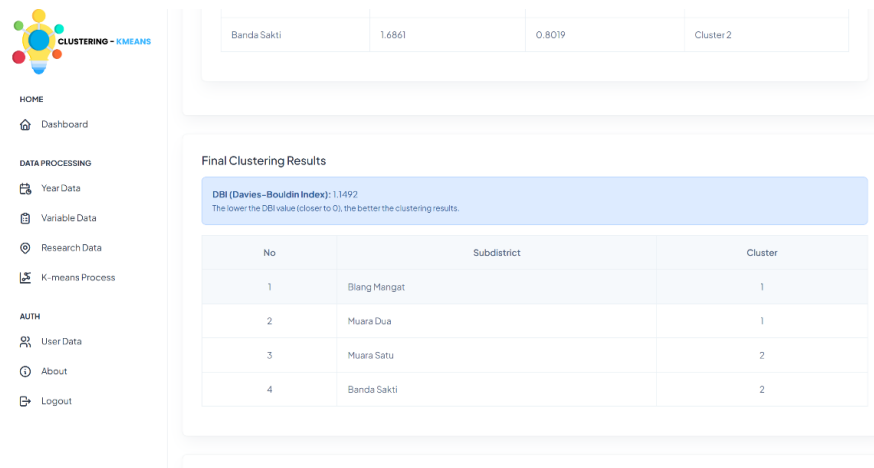
e. Halaman Proses K-Means



Gambar 6. Halaman Proses K-Means

Halaman Proses K-means dirancang untuk memfasilitasi pengguna dalam menjalankan proses segmentasi data kependudukan menggunakan algoritma K means. Pada bagian atas, pengguna dapat memilih tahun data yang ingin digunakan melalui menu dropdown Pilih Tahun, memastikan bahwa proses K-means dapat dijalankan berdasarkan data tahun tertentu.

f. Hasil Clustering



Gambar 7. Hasil *Clustering* Tahun 2020

Hasil akhir pengelompokan untuk data tahun 2020 menunjukkan pembagian kecamatan menjadi dua kluster berdasarkan analisis *K-means*. Kluster 1 meliputi Blang Mangat dan Muara Dua, sedangkan Kluster 2 terdiri dari Muara Satu dan Banda Sakti. Pembagian ini didasarkan pada variabel-variabel seperti jumlah penduduk, laju pertumbuhan, kepadatan penduduk, persentase, dan rasio jenis kelamin. Setiap kecamatan dimasukkan ke dalam kluster yang paling tepat berdasarkan jarak centroid. Indeks *Davies-Bouldin* (DBI) untuk pengelompokan ini adalah 1,1492, yang menunjukkan kualitas pengelompokan yang baik dengan pemisahan yang jelas antara kluster. Meskipun DBI berada di atas nilai optimal, namun DBI memberikan dasar yang kuat untuk analisis lebih lanjut. Pengelompokan ini mencerminkan pola demografi yang berbeda di seluruh wilayah, yang menawarkan wawasan berharga untuk perencanaan dan pembuatan kebijakan yang tepat sasaran berdasarkan karakteristik khusus setiap wilayah.

4. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma *K-means* dalam segmentasi kepadatan penduduk di Kota Lhokseumawe dengan memanfaatkan data kependudukan dari tahun 2020 hingga 2023. Sistem yang dibangun berbasis GIS mampu memvisualisasikan hasil *clustering* secara interaktif, memudahkan analisis dan pengambilan keputusan dalam perencanaan tata ruang dan kebijakan kependudukan. Hasil *clustering* menunjukkan dinamika perubahan pola pengelompokan tiap kecamatan dari tahun ke tahun. Misalnya, pada tahun 2020, kecamatan Blang Mangat dan Muara Dua terkelompok dalam *Cluster* 1, sementara Muara Satu dan Banda Sakti berada di *Cluster* 2. Pada tahun 2021, pola pengelompokan berubah dengan Blang Mangat tetap di *Cluster* 1, dan kecamatan lainnya di *Cluster* 2. Di tahun 2022, Blang Mangat dan Muara Dua kembali berada di *Cluster* 1, sedangkan Muara Satu dan Banda Sakti berada di *Cluster* 2. Pada tahun 2023, pola *clustering* lebih terstruktur, dengan Blang Mangat dan Muara Satu berada dalam *cluster* yang terpisah dari Muara Dua dan Banda Sakti, mencerminkan adanya perubahan signifikan dalam karakteristik demografis yang terdeteksi melalui algoritma *K-means*. Evaluasi kualitas *clustering* menggunakan indeks *Davies-Bouldin* (DBI) menunjukkan hasil yang memadai. Nilai DBI yang semakin rendah mengindikasikan hasil *clustering* yang lebih baik. Nilai DBI pada tahun 2020 adalah 1.1492, sementara pada tahun 2021 menurun menjadi 0.6859, menunjukkan peningkatan kualitas *clustering*. Pada tahun 2022, nilai DBI kembali meningkat menjadi 1.2470, namun pada tahun 2023, nilai DBI kembali menurun menjadi 0.6805, menunjukkan bahwa hasil *clustering* pada tahun tersebut memiliki kualitas terbaik dan terstruktur dengan baik. Dengan demikian, penelitian ini menunjukkan bahwa algoritma *K-means* efektif dalam mengidentifikasi pola kepadatan penduduk yang berubah-ubah, serta menyediakan wawasan yang berguna untuk perencanaan pengelolaan kependudukan berbasis data yang lebih akurat.

REFERENCES

- [1] Badan Pusat Statistik Kota Lhokseumawe Tahun 2024.
- [2] Abbas, R., Ayu Lizar, C., & Satriawan, H, "Analisis Tempat Penampungan Sementara (TPS) Menggunakan Sistem Informasi Geografis Untuk Pengelolaan Sampah Berkelanjutan Di Kota Lhokseumawe," *Jurnal Pengendalian Pencemaran Lingkungan (Jppl)*, 7(1), 2025.
- [3] Sani Putriana, Ernawati, "Clustering Data Titik Gempa Dengan Metode Fuzzy Possibilistic C-Means", 2021.
- [4] Agnes Vernanda, A., Faisol, A., & Vendyansyah, N, "Penerapan Metode K-Means Clustering Untuk Pemetaan Daerah Rawan Kecelakaan Lalu Lintas Di Kota Malang Berbasis Website," *In Jurnal Mahasiswa Teknik Informatika* (Vol. 5, Issue 2), 2021.
- [5] Alif, D., Fadhillah, F., Faisol, A., & Vendyansyah, N, "Penerapan Metode K-Means Clustering Pada Pemetaan Lahan Kopi Di Kabupaten Malang," *In Jurnal Mahasiswa Teknik Informatika* (Vol. 6, Issue 1), 2022.

- [6] Asyfani, Y., Manfaati Nur, I., Fathoni Amri, I., Yunanita, N., Hikmah Nur Rohim, F., Aura Hisani, Z., & Anggun Lestari, F, "Pengelompokan Kabupaten/Kota Di Jawa Tengah Berdasarkan Kepadatan Penduduk Menggunakan Metode Hierarchical Clustering" 2(1), 1–8, 2024.
- [7] J. Ilmu Matematika Dan Terapan And Desember, "Analisis Cluster Dengan Menggunakan Metode K-Means Untuk Pengelompokan Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014," 2017.
- [8] Butsianto, S., & Mayangwulan, N. T, "Penerapan Data Mining Untuk Prediksi Penjualan Mobil Menggunakan Metode K-Means Clustering," Jurnal Nasional Komputasi Dan Teknologi Informasi, 3(3), 2020.
- [9] Faisal, M., & Fitri, Z. (n.d.), "Information and Communication Technology Competencies Clustering for students for Vocational High School Students Using K-Means Clustering Algorithm" International Journal of Engineering, Science & Information Technology (IJESTY), 2022.
- [10] Gultom, D., Gunawan, I., Purnamasari, I., Andani, S. R., & Siregar, Z. A, "Penerapan Algoritma K-Means Dalam Pengelompokan Kepadatan Penduduk Menurut Kecamatan di Kabupaten Simalungun" TIN: Terapan Informatika Nusantara, 2(10), 622–628, 2022.
- [11] Indraputra, R. A., & Fitriana, R. (n.d.). K-Means Clustering Data COVID-19.
- [12] Nur Afifah, I. A., & Nurdianto, H, "Data Mining Clustering Dalam Pengelompokan Buku Perpustakaan Menggunakan Algoritma K-Means" Jipi (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika), 8(3), 802–814, 2023.
- [13] Peralatan, P., Pada, S., Wigglo, B., Menggunakan, D., Apriori, A., Syahril, M., Erwansyah, K., & Yetri, M. J-SISKO TECH Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD Penerapan Data Mining Untuk Menentukan Pola, 118(1), 118–136, 2020.
- [14] Tadi, M., & Arnawisuda Ningsi, B, "Analisis Klaster Kemiskinan Kabupaten Kota Di Provinsi Banten Menggunakan Metode K-Means" 4(1), 2023.
- [15] Veronica, M., Effendi, H., & Oktafian Saleh, Ari. (N.D.), "Clustering Tingkat Kedisiplinan Pegawai Pada Pengadilan Tinggi Palembang Menggunakan Algoritma K-Means".
- [16] Yulianeu, A., & Oktamala, R, "Sistem Informasi Geografis Trayek Angkutan Umum Di Kota Tasikmalaya Berbasis Web," Jutekin (Jurnal Teknik Informatika), 10(2), 2022.
- [17] Zheng, A., Cai, J., Yang, H., & Zhao, X. CPGAN: Curve Clustering Architecture Based On Projected Latent Vector Of Generative Adversarial Network. IEEE Access, 8, 86765–86776, 2020.
- [18] Zufria, I., & Damaiani Iskandar, I. "Clustering Pasien Rawat Inap Di RS USU Menggunakan Algoritma K-Means Clustering of Inpatients at USU Hospital Using the K-Means Algorithm," Journal Of Computer Science And Informatics Engineering (Cosie), 03(2), 54–63, 2024.
- [19] E. Muningsih, H. M. Nur, F. F. Dwi Imaniawan, Saifudin, V. R. Handayani, and F. Endiarto, "Comparative Analysis on Dimension Reduction Algorithm of Principal Component Analysis and Singular Value Decomposition for Clustering," J. Phys. Conf. Ser., vol. 1641, no. 1, 2020, doi: 10.1088/1742-6596/1641/1/012101.
- [20] Ferdiansyah, I., Huda, B., Hananto, A., & Tukino, T, "Analisis Clustering Menggunakan Metode K-Means Pada Kemiskinan Di Jawa Timur Tahun 2020," Innovative: Journal Of Social Science Research, 4(3), 858-869, 2024.
- [21] Handoko, S., Fauziah, F., & Handayani, E. T. E, "Implementasi data mining untuk menentukan tingkat penjualan paket data Telkomsel menggunakan metode K-Means clustering," Jurnal Ilmiah Teknologi dan Rekayasa, 25(1), 76-88, 2020.