

PENGELOMPOKAN WILAYAH PENGEMBANGAN SAMBUNGAN BARU PDAM TIRTA SILAUIPIASA BERBASIS METODE CLUSTERING K-MEDOIDS

Syahputri Amelia¹, Dicky Apdilah²

Universitas Asahan, Asahan

e-mail: ¹syahputriamelia@gmail.com, ²dickyapdi1404@gmail.com

Abstract : *The development of new pipe connections is one of the important aspects in improving clean water services provided by the Regional Water Company (PDAM). However, the determination of priority areas for pipe network development at PDAM Tirta Silaupiasa Asahan Regency is still carried out subjectively, so it is less objective and data-based. Therefore, this research aims to apply the K-Medoids clustering algorithm to group development areas of new pipe connections based on their characteristics. The data used consisted of 22 sub-district areas with three input variables, namely the number of prospective new customers, distance to the existing pipe network, and population density. The data was normalized using the Min-Max method, then processed using the K-Medoids algorithm with the number of clusters $k = 3$ and Euclidean Distance as the distance measure. The clustering results were evaluated using the Silhouette Score method. Based on the test results, the clustering process converged after three iterations with the total cost decreasing from 6.3947 to 5.3436. The clustering produced three groups: Cluster 1 (High Priority) with 1 area, Cluster 2 (Medium Priority) with 16 areas, and Cluster 3 (Low Priority) with 5 areas. The average Silhouette Score obtained was 0.3981, which is included in the reasonable structure category. The system was implemented as a web-based application using PHP and MySQL, so it can help PDAM in determining priority areas for pipe network development objectively and based on data.*

Keywords: *K-Medoids, Clustering, Water Pipe Network, Silhouette Score, Data Mining*

Abstrak : Pengembangan pasang baru jalur pipa merupakan salah satu aspek penting dalam meningkatkan pelayanan air bersih yang disediakan oleh Perusahaan Daerah Air Minum (PDAM). Namun, penentuan wilayah prioritas pengembangan jaringan pipa pada PDAM Tirta Silaupiasa Kabupaten Asahan masih dilakukan secara subjektif, sehingga kurang objektif dan berbasis data. Oleh karena itu, penelitian ini bertujuan untuk menerapkan algoritma clustering K-Medoids dalam mengelompokkan wilayah pengembangan pasang baru jalur pipa berdasarkan karakteristiknya. Data yang digunakan terdiri dari 22 wilayah kelurahan dengan tiga variabel input, yaitu jumlah calon pelanggan baru, jarak ke jaringan pipa eksisting, dan kepadatan penduduk. Data dinormalisasi menggunakan metode Min-Max, kemudian diproses menggunakan algoritma K-Medoids dengan jumlah cluster $k = 3$ dan Euclidean Distance sebagai ukuran jarak. Hasil clustering dievaluasi menggunakan metode Silhouette Score. Berdasarkan hasil pengujian, proses clustering konvergen setelah tiga iterasi dengan total cost menurun dari 6,3947 menjadi 5,3436. Pengelompokan menghasilkan tiga kelompok, yaitu Cluster 1 (Prioritas Tinggi) dengan 1 wilayah, Cluster 2 (Prioritas Sedang) dengan 16 wilayah, dan Cluster 3 (Prioritas Rendah) dengan 5 wilayah. Nilai rata-rata Silhouette Score yang diperoleh sebesar 0,3981 yang termasuk dalam kategori reasonable structure. Sistem diimplementasikan sebagai aplikasi berbasis web menggunakan PHP dan MySQL, sehingga dapat membantu PDAM dalam menentukan wilayah prioritas pengembangan jaringan pipa secara objektif dan berbasis data.

Kata kunci: K-Medoids, Clustering, Jaringan Pipa Air, Silhouette Score, Data Mining

PENDAHULUAN

Air bersih merupakan kebutuhan dasar masyarakat yang harus tersedia secara memadai, baik dari sisi kuantitas, kualitas, maupun kontinuitas pelayanan. Perusahaan Daerah Air Minum (PDAM) sebagai penyedia layanan publik yang vital memegang peranan penting dalam penyaluran air bersih dari instalasi produksi menuju pelanggan melalui jaringan distribusi perpipaan. Peningkatan jumlah penduduk, pertumbuhan permukiman, dan kebutuhan pelayanan dalam jangka panjang menuntut pengembangan jaringan perpipaan yang lebih terarah dan efisien (Hasnawiyah et al., 2022).

PDAM sebagai penyedia layanan air minum sering menghadapi persoalan dalam menentukan wilayah prioritas pengembangan pasang baru jalur pipa. Tidak semua wilayah memiliki karakteristik kebutuhan dan potensi pelanggan yang sama. Sebagian wilayah memiliki kepadatan penduduk tinggi, potensi pelanggan besar, dan akses jaringan yang memungkinkan, sementara wilayah lain memiliki potensi lebih rendah atau memerlukan biaya pengembangan lebih besar. Apabila penentuan wilayah pengembangan dilakukan secara manual atau hanya berdasarkan pertimbangan subjektif, maka terdapat risiko ketidaktepatan dalam penetapan prioritas, penggunaan anggaran yang kurang efisien, dan hasil pelayanan yang tidak optimal.

Permasalahan tersebut menuntut adanya pendekatan yang lebih objektif, sistematis, dan berbasis data. Salah satu pendekatan yang dapat digunakan adalah data mining khususnya teknik clustering. Clustering bertujuan mengelompokkan data ke dalam kelompok-kelompok yang memiliki karakteristik serupa, sehingga dapat membantu pengambilan keputusan berdasarkan pola yang muncul dari data. Salah satu algoritma clustering yang relevan digunakan adalah K-Medoids. K-Medoids merupakan metode partisi yang menggunakan objek data nyata sebagai pusat cluster (medoid), sehingga lebih

stabil terhadap data ekstrem dibanding metode yang berbasis rata-rata seperti K-Means (Yafi et al., 2023).

Penelitian terdahulu oleh Annisa et al. (2022) menerapkan data mining untuk pengelompokan data pengguna air bersih berdasarkan keluhanannya menggunakan metode clustering pada PDAM Langkat. Penelitian lain oleh Effendie & Kariyam (2023) mengembangkan indeks rasio deviasi berbasis medoid untuk menentukan jumlah klaster dalam suatu dataset. Berdasarkan hal tersebut, penerapan algoritma K-Medoids untuk clustering wilayah pengembangan pasang baru jalur pipa PDAM Tirta Silaupiasa menjadi penting untuk dilakukan. Penelitian ini diharapkan menghasilkan pengelompokan wilayah secara objektif berdasarkan karakteristik data yang dimiliki, sehingga PDAM memperoleh dasar analitis dalam menentukan prioritas pengembangan jaringan pasang baru.

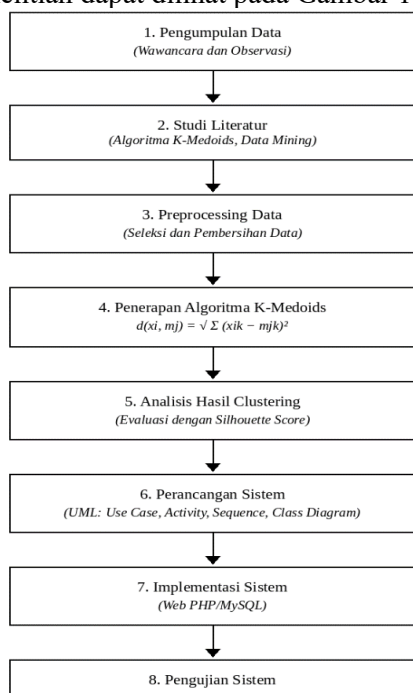
Data mining merupakan proses penggalian informasi, pola, dan pengetahuan yang tersembunyi dari sekumpulan data dalam jumlah besar dengan menggunakan teknik statistik, matematika, dan kecerdasan buatan (Maulia et al., 2021). Salah satu teknik dalam data mining adalah clustering, yaitu metode pengelompokan data yang bertujuan memisahkan data ke dalam beberapa kelompok berdasarkan kesamaan karakteristik yang dimiliki (Ina et al., 2022). Data yang memiliki karakteristik serupa akan dikelompokkan dalam satu cluster, sedangkan data yang berbeda akan berada pada cluster yang berbeda.

Algoritma K-Medoids termasuk ke dalam teknik partitioning clustering. Metode ini mengelompokkan data ke dalam beberapa cluster berdasarkan kedekatan karakteristik data, dengan pusat setiap cluster disebut medoid, yaitu objek yang memiliki total jarak minimum terhadap data lain dalam cluster yang sama. Penggunaan medoid menjadikan metode ini lebih tahan terhadap pengaruh data ekstrem dibandingkan metode clustering lain seperti K-Means (Yafi et al., 2023). Untuk mengukur kedekatan

antar data, digunakan Euclidean Distance sebagai ukuran jarak. Adapun langkah-langkah algoritma K-Medoids adalah menentukan jumlah cluster (k), memilih k objek sebagai medoid awal, menghitung jarak antara setiap data dengan medoid, mengelompokkan data ke cluster terdekat, memilih medoid baru dengan meminimalkan total jarak, serta mengulangi proses hingga tidak ada perubahan cluster.

METODE

Metode penelitian yang digunakan dalam penelitian ini adalah metode data mining dengan pendekatan clustering menggunakan algoritma K-Medoids. Metode ini dipilih karena kemampuannya dalam mengelompokkan wilayah berdasarkan kemiripan karakteristik data serta ketahanannya terhadap data yang bersifat outlier. Tahapan penelitian dilakukan secara sistematis mulai dari pengumpulan data, studi literatur, preprocessing data, penerapan algoritma K-Medoids, analisis hasil clustering, perancangan sistem, implementasi, hingga pengujian sistem. Kerangka kerja penelitian dapat dilihat pada Gambar 1.



Gambar 1 Kerangka Kerja Penelitian

Adapun uraian tahapan kerangka kerja penelitian adalah sebagai berikut. (1) Pengumpulan data dilakukan melalui wawancara dan observasi langsung ke PDAM. (2) Studi literatur dilakukan dengan mempelajari teori algoritma K-Medoids, data mining, teknik clustering, serta studi terdahulu yang relevan. (3) Preprocessing data mencakup seleksi data, pembersihan data dari nilai kosong atau tidak konsisten, serta normalisasi. (4) Penerapan algoritma K-Medoids untuk mengelompokkan wilayah berdasarkan kemiripan karakteristik. (5) Analisis hasil clustering dengan Silhouette Score. (6) Perancangan sistem menggunakan pendekatan UML. (7) Implementasi sistem sebagai aplikasi berbasis web. (8) Pengujian sistem sebagai dasar evaluasi sebelum sistem diimplementasikan.

Pengumpulan Data

Pengumpulan data dilakukan melalui wawancara dan observasi langsung ke PDAM Tirta Silaupiasa Kisaran. Data yang digunakan bersumber dari tiga sumber utama, yaitu Laporan Sambungan Baru PDAM Tirta Silaupiasa Tahun Anggaran 2024–2025 untuk variabel jumlah calon pelanggan baru (X_1), estimasi visual berdasarkan Peta Jaringan Perpipaan untuk variabel jarak ke jaringan pipa eksisting (X_2), dan publikasi BPS Kabupaten Asahan untuk variabel kepadatan penduduk (X_3).

Penerapan Algoritma K-Medoids

Algoritma K-Medoids diterapkan untuk mengelompokkan wilayah berdasarkan tingkat kemiripan karakteristik data. Tahapan penerapannya adalah: menentukan jumlah cluster (k), memilih k objek data secara acak sebagai medoid awal, menghitung jarak setiap data terhadap seluruh medoid menggunakan Euclidean Distance, mengelompokkan setiap data ke cluster dengan medoid terdekat, menghitung total cost, serta melakukan pertukaran medoid dan non-medoid hingga tidak terjadi perubahan cluster (konvergen). Kualitas hasil clustering dievaluasi menggunakan Silhouette Score, dan sistem

diimplementasikan sebagai aplikasi berbasis web menggunakan PHP dan MySQL.

HASIL DAN PEMBAHASAN

Analisa Data

Data penelitian diperoleh dari tiga sumber utama dan menghasilkan 22 data wilayah kelurahan beserta nilai ketiga variabel input yang digunakan sebagai masukan proses clustering. Data penelitian selengkapnya disajikan pada Tabel 1.

Tabel 1 Data Penelitian

No	Kelurahan	X1	X2	X3
1	Kisaran Baru	77	0.3	7.943
2	Mutiara	42	0.4	5.562
3	Siumbut Umbut	29	0.8	2.889
4	Siumbut Baru	28	1.2	2.886
5	Kisaran Timur	24	0.5	6.098
6	Sidodadi	22	1.5	4.060
7	Selawan	22	0.6	2.761
8	Gambir Baru	21	0.9	995
9	Sidomukti	21	1.3	755
10	Lestari	19	1.0	1.402
11	Kedai Ledang	16	1.4	1.753
12	Sei Renggas	15	1.8	698
13	Kisaran Naga	13	0.7	2.958
14	Sendang Sari	12	1.6	15.071
15	Bunut Barat	12	0.8	3.850
16	Mekar Baru	12	1.1	5.348
17	Sentang	10	2.5	2.503
18	Kisaran Barat	8	0.4	8.157
19	Teladan	7	0.5	9.281
20	Karang Anyar	5	2.0	747
21	Sidomulyo	3	1.7	755
22	Subur	2	3.0	644

Normalisasi Data

Sebelum diterapkan algoritma K-Medoids, seluruh data dinormalisasi menggunakan metode Min-Max agar setiap variabel berada pada rentang nilai yang sama, yaitu antara 0 hingga 1. Normalisasi diperlukan karena ketiga variabel memiliki satuan dan skala yang berbeda. Rumus normalisasi Min-Max adalah sebagai berikut.

$$x_{norm} = (x - x_{min}) / (x_{max} - x_{min}) \quad (1)$$

Berdasarkan data penelitian, diperoleh nilai minimum dan maksimum setiap variabel sebagai berikut: $X1_{min} = 2$, $X1_{max} = 77$; $X2_{min} = 0,3$, $X2_{max} = 3$; $X3_{min} = 644$, $X3_{max} = 15.071$. Sebagai contoh, hasil normalisasi untuk wilayah Kisaran Baru adalah $X1 = (77-2)/(77-2) = 1$; $X2 = (0,3-0,3)/(3-0,3) = 0$; $X3 = (7943-644)/(15071-644) = 0,5059$. Hasil normalisasi seluruh wilayah disajikan pada Tabel 2.

Tabel 2 Hasil Normalisasi Data

No	Kelurahan	X1	X2	X3
1	Kisaran Baru	1,0000	0,0000	0,5059
2	Mutiara	0,5333	0,0370	0,3409
3	Siumbut Umbut	0,3600	0,1852	0,1556
4	Siumbut Baru	0,3467	0,3333	0,1554
5	Kisaran Timur	0,2933	0,0741	0,3780
6	Sidodadi	0,2667	0,4444	0,2368
7	Selawan	0,2667	0,1111	0,1467
8	Gambir Baru	0,2533	0,2222	0,0243
9	Sidomukti	0,2533	0,3704	0,0077
10	Lestari	0,2267	0,2593	0,0525
11	Kedai Ledang	0,1867	0,4074	0,0769
12	Sei Renggas	0,1733	0,5556	0,0037
13	Kisaran Naga	0,1467	0,1481	0,1604
14	Sendang Sari	0,1333	0,4815	1,0000

No	Kelurahan	X1	X2	X3
15	Bunut Barat	0,1333	0,1852	0,2222
16	Mekar Baru	0,1333	0,2963	0,3261
17	Sentang	0,1067	0,8148	0,1289
18	Kisaran Barat	0,0800	0,0370	0,5208
19	Teladan	0,0667	0,0741	0,5987
20	Karang Anyar	0,0400	0,6296	0,0071
21	Sidomulyo	0,0133	0,5185	0,0077
22	Subur	0,0000	1,0000	0,0000

Nilai hasil normalisasi tersebut kemudian digunakan sebagai masukan proses iterasi K-Medoids

Penerapan Algoritma K-Medoids

Setelah data dinormalisasi, algoritma K-Medoids diterapkan dengan jumlah cluster $k = 3$ yang merepresentasikan tiga kategori prioritas pengembangan wilayah, yaitu Prioritas Tinggi, Prioritas Sedang, dan Prioritas Rendah. Berikut ini disajikan perhitungan manual proses clustering pada iterasi pertama.

Langkah 1. Penentuan Medoid Awal

Tabel 3 Medoid Awal

Medoid	Nama Wilayah	X1	X2	X3
M1	Kisaran Baru	1,0000	0,0000	0,5059
M2	Gambir Baru	0,2533	0,2222	0,0243
M3	Subur	0,0000	1,0000	0,0000

Langkah 2. Perhitungan Jarak Euclidean

Jarak setiap wilayah terhadap ketiga medoid dihitung menggunakan rumus Euclidean Distance berikut.

$$d(W, M) = \sqrt{[(X_{1W} - X_{1M})^2 + (X_{2W} - X_{2M})^2 + (X_{3W} - X_{3M})^2]} \quad (3)$$

Sebagai contoh, perhitungan untuk wilayah Kisaran Baru ($X_1 = 1$; $X_2 = 0$;

$X_3 = 0,5059$) terhadap ketiga medoid adalah sebagai berikut.

$$\begin{aligned} d(Kisaran Baru, M1) &= \sqrt{[(1 - 1)^2 + (0 - 0)^2 + (0,5059 - 0,5059)^2]} = \sqrt{0} = 0 \\ d(Kisaran Baru, M2) &= \sqrt{[(1 - 0,2533)^2 + (0 - 0,2222)^2 + (0,5059 - 0,0243)^2]} \\ &= \sqrt{0,8389} = 0,9159 \\ d(Kisaran Baru, M3) &= \sqrt{[(1 - 0)^2 + (0 - 1)^2 + (0,5059 - 0)^2]} = \sqrt{2,2559} \\ &= 1,5020 \end{aligned}$$

Jarak terkecil adalah $d(M1) = 0$, sehingga wilayah Kisaran Baru dikelompokkan ke dalam Cluster C1.

Contoh berikutnya untuk wilayah Mutiara ($X_1 = 0,5333$; $X_2 = 0,037$; $X_3 = 0,3409$):

$$\begin{aligned} d(Mutiara, M1) &= \sqrt{[(0,5333 - 1)^2 + (0,037 - 0)^2 + (0,3409 - 0,5059)^2]} \\ &= \sqrt{0,2464} = 0,4964 \\ d(Mutiara, M2) &= \sqrt{[(0,5333 - 0,2533)^2 + (0,037 - 0,2222)^2 + (0,3409 - 0,0243)^2]} \\ &= \sqrt{0,2129} = 0,4614 \\ d(Mutiara, M3) &= \sqrt{[(0,5333 - 0)^2 + (0,037 - 1)^2 + (0,3409 - 0)^2]} \\ &= \sqrt{1,3280} = 1,1524 \end{aligned}$$

Jarak terkecil adalah $d(M2) = 0,4614$, sehingga wilayah Mutiara dikelompokkan ke dalam Cluster C2.

Contoh untuk wilayah Siumbut Umbut ($X_1 = 0,36$; $X_2 = 0,1852$; $X_3 = 0,1556$):

$$\begin{aligned} d(Siumbut Umbut, M1) &= \sqrt{[(0,36 - 1)^2 + (0,1852 - 0)^2 + (0,1556 - 0,5059)^2]} \\ &= \sqrt{0,5666} = 0,7527 \\ d(Siumbut Umbut, M2) &= \sqrt{[(0,36 - 0,2533)^2 + (0,1852 - 0,2222)^2 + (0,1556 - 0,0243)^2]} \\ &= \sqrt{0,0300} = 0,1732 \\ d(Siumbut Umbut, M3) &= \sqrt{[(0,36 - 0)^2 + (0,1852 - 1)^2 + (0,1556 - 0)^2]} \\ &= \sqrt{0,8178} = 0,9043 \end{aligned}$$

Jarak terkecil adalah $d(M2) = 0,1732$, sehingga wilayah Siumbut Umbut dikelompokkan ke dalam Cluster C2.

Contoh untuk wilayah Sentang ($X_1 = 0,1067$; $X_2 = 0,8148$; $X_3 = 0,1289$) yang masuk cluster berbeda:

$$\begin{aligned} d(Sentang, M1) &= \sqrt{[(0,1067 - 1)^2 + (0,8148 - 0)^2 + (0,1289 - 0,5059)^2]} \\ &= \sqrt{1,6040} = 1,2665 \\ d(Sentang, M2) &= \sqrt{[(0,1067 - 0,2533)^2 + (0,8148 - 0,2222)^2 + (0,1289 - 0,0243)^2]} \\ &= \sqrt{0,3837} = 0,6194 \\ d(Sentang, M3) &= \sqrt{[(0,1067 - 0)^2 + (0,8148 - 1)^2 + (0,1289 - 0)^2]} \\ &= \sqrt{0,0623} = 0,2496 \end{aligned}$$

Jarak terkecil adalah $d(M3) = 0,2496$, sehingga wilayah Sentang dikelompokkan ke dalam Cluster C3

Langkah 3. Pengelompokan Iterasi 1

Perhitungan jarak yang sama dilakukan untuk seluruh 22 wilayah, kemudian setiap wilayah dikelompokkan

ke cluster dengan jarak terkecil (jarak minimum). Hasil selengkapnya iterasi 1 disajikan pada Tabel 4.

Tabel 4 Hasil Iterasi 1

No	Kelurahan	d(M1)	d(M2)	d(M3)	Min	Cl
1	Kisaran Baru	0	0,9159	1,5020	0	C1
2	Mutiara	0,4964	0,4614	1,1524	0,4614	C2
3	Siumbut Umbut	0,7527	0,1732	0,9043	0,1732	C2
4	Siumbut Baru	0,8129	0,1956	0,7674	0,1956	C2
5	Kisaran Timur	0,7220	0,3855	1,0422	0,3855	C2
6	Sidodadi	0,8987	0,3077	0,6602	0,3077	C2
7	Selawan	0,8241	0,1658	0,9396	0,1658	C2
8	Gambir Baru	0,9159	0	0,8184	0	C2
9	Sidomukti	0,9711	0,1491	0,6787	0,1491	C2
10	Lestari	0,9332	0,0537	0,7764	0,0537	C2
11	Kedai Ledang	1,0057	0,2037	0,6261	0,2037	C2
12	Sei Renggas	1,1155	0,3435	0,4770	0,3435	C2
13	Kisaran Naga	0,9324	0,1881	0,8792	0,1881	C2
14	Sendang Sari	1,1078	1,0167	1,1343	1,0167	C2
15	Bunut Barat	0,9306	0,2344	0,8550	0,2344	C2
16	Mekar Baru	0,9334	0,3331	0,7870	0,3331	C2
17	Sentang	1,2665	0,6194	0,2496	0,2496	C3
18	Kisaran Barat	0,9209	0,5575	1,0977	0,5575	C2
19	Teladan	0,9408	0,6218	1,1046	0,6218	C2
20	Karang Anyar	1,2517	0,4602	0,3726	0,3726	C3
21	Sidomulyo	1,2209	0,3817	0,4817	0,3817	C2
22	Subur	1,5020	0,8184	0	0	C3

Langkah 4. Pengelompokan Iterasi 2

Dengan medoid baru tersebut, perhitungan jarak diulang untuk seluruh

wilayah. Hasil iterasi 2 disajikan pada Tabel 5, dengan medoid M2 = Bunut Barat dan M3 = Sentang.

Tabel 5 Hasil Iterasi 2

No	Kelurahan	d(M1)	d(M2)	d(M3)	Min	Cl
1	Kisaran Baru	0	0,9306	1,2665	0	C1
2	Mutiara	0,4964	0,4428	0,9121	0,4428	C2
3	Siumbut Umbut	0,7527	0,2363	0,6792	0,2363	C2
4	Siumbut Baru	0,8129	0,2682	0,5387	0,2682	C2
5	Kisaran Timur	0,7220	0,2494	0,8034	0,2494	C2
6	Sidodadi	0,8987	0,2919	0,4177	0,2919	C2
7	Selawan	0,8241	0,1703	0,7219	0,1703	C2
8	Gambir Baru	0,9159	0,2344	0,6194	0,2344	C2
9	Sidomukti	0,9711	0,3077	0,4834	0,3077	C2
10	Lestari	0,9332	0,2074	0,5734	0,2074	C2
11	Kedai Ledang	1,0057	0,2708	0,4184	0,2708	C2

No	Kelurahan	d(M1)	d(M2)	d(M3)	Min	CI
12	Sei Renggas	1,1155	0,4319	0,2955	0,2955	C3
13	Kisaran Naga	0,9324	0,0733	0,6686	0,0733	C2
14	Sendang Sari	1,1078	0,8323	0,9331	0,8323	C2
15	Bunut Barat	0,9306	0	0,6370	0	C2
16	Mekar Baru	0,9334	0,1521	0,5554	0,1521	C2
17	Sentang	1,2665	0,6370	0	0	C3
18	Kisaran Barat	0,9209	0,3376	0,8714	0,3376	C2
19	Teladan	0,9408	0,3982	0,8780	0,3982	C2
20	Karang Anyar	1,2517	0,5025	0,2315	0,2315	C3
21	Sidomulyo	1,2209	0,4141	0,3335	0,3335	C3
22	Subur	1,5020	0,8550	0,2496	0,2496	C3

Langkah 5 Pengelompokan Iterasi 3**Tabel 6 Hasil Iterasi 3**

No	Kelurahan	d(M1)	d(M2)	d(M3)	Min	CI
1	Kisaran Baru	0	0,9306	1,2517	0	C1
2	Mutiara	0,4964	0,4428	0,8402	0,4428	C2
3	Siumbut Umbut	0,7527	0,2363	0,5674	0,2363	C2
4	Siumbut Baru	0,8129	0,2682	0,4515	0,2682	C2
5	Kisaran Timur	0,7220	0,2494	0,7144	0,2494	C2
6	Sidodadi	0,8987	0,2919	0,3721	0,2919	C2
7	Selawan	0,8241	0,1703	0,5829	0,1703	C2
8	Gambir Baru	0,9159	0,2344	0,4602	0,2344	C2
9	Sidomukti	0,9711	0,3077	0,3357	0,3077	C2
10	Lestari	0,9332	0,2074	0,4172	0,2074	C2
11	Kedai Ledang	1,0057	0,2708	0,2753	0,2708	C2
12	Sei Renggas	1,1155	0,4319	0,1525	0,1525	C3
13	Kisaran Naga	0,9324	0,0733	0,5165	0,0733	C2
14	Sendang Sari	1,1078	0,8323	1,0082	0,8323	C2
15	Bunut Barat	0,9306	0	0,5025	0	C2
16	Mekar Baru	0,9334	0,1521	0,4707	0,1521	C2
17	Sentang	1,2665	0,6370	0,2315	0,2315	C3
18	Kisaran Barat	0,9209	0,3376	0,7853	0,3376	C2
19	Teladan	0,9408	0,3982	0,8120	0,3982	C2
20	Karang Anyar	1,2517	0,5025	0	0	C3
21	Sidomulyo	1,2209	0,4141	0,1143	0,1143	C3
22	Subur	1,5020	0,8550	0,3726	0,3726	C3

Tabel 7 Perbandingan Total Cost Setiap Iterasi

Iterasi	Medoid C1	Medoid C2	Medoid C3	Total Cost
1	Kisaran Baru	Gambir Baru	Subur	6,3947
2	Kisaran Baru	Bunut Barat	Sentang	5,5828
3	Kisaran Baru	Bunut Barat	Karang Anyar	5,3436

Evaluasi dengan Silhouette Score

Kualitas hasil pengelompokan dievaluasi menggunakan metode Silhouette Score. Nilai Silhouette Score berkisar antara -1 hingga 1, di mana nilai yang semakin mendekati 1 menunjukkan bahwa setiap data memiliki kemiripan yang lebih tinggi dengan anggota clusternya sendiri dibandingkan cluster lain. Rumus Silhouette Score adalah sebagai berikut.

$$s(i) = (b(i) - a(i)) / \max(a(i), b(i)) \quad (2)$$

dengan $a(i)$ adalah rata-rata jarak data ke- i terhadap seluruh anggota cluster yang sama, dan $b(i)$ adalah rata-rata jarak minimum data ke- i terhadap anggota cluster lain yang terdekat. Berikut

disajikan beberapa contoh perhitungan Silhouette Score.

Wilayah Kisaran Baru (Cluster C1): $a(i) = 0$; $b(i) = \min(0,8812 ; 1,2713) = 0,8812$; sehingga $s = (0,8812 - 0) / 0,8812 = 1,0000$ (Baik).

Wilayah Selawan (Cluster C2): $a(i) = 0,3118$; $b(i) = \min(0,8241 ; 0,6440) = 0,6440$; sehingga $s = (0,6440 - 0,3118) / 0,6440 = 0,5158$ (Baik).

Wilayah Karang Anyar (Cluster C3): $a(i) = 0,2177$; $b(i) = \min(1,2517 ; 0,5695) = 0,5695$; sehingga $s = (0,5695 - 0,2177) / 0,5695 = 0,6177$ (Baik).

Perhitungan yang sama dilakukan untuk seluruh 22 wilayah. Hasil akhir pengelompokan dan nilai Silhouette Score seluruh wilayah disajikan pada Tabel 8.

Tabel 8 Hasil Akhir Clustering dan Silhouette Score

No	Kelurahan	Cluster	Prioritas	s(i)
1	Kisaran Baru	C1	Tinggi	1,0000
2	Mutiara	C2	Sedang	0,0461
3	Siumbut Umbut	C2	Sedang	0,4914
4	Siumbut Baru	C2	Sedang	0,3494
5	Kisaran Timur	C2	Sedang	0,5086
6	Sidodadi	C2	Sedang	0,1348
7	Selawan	C2	Sedang	0,5158
8	Gambir Baru	C2	Sedang	0,3586
9	Sidomukti	C2	Sedang	0,0500
10	Lestari	C2	Sedang	0,3391
11	Kedai Ledang	C2	Sedang	-0,0622
12	Sei Renggas	C3	Rendah	0,4314
13	Kisaran Naga	C2	Sedang	0,4733
14	Sendang Sari	C2	Sedang	0,1616
15	Bunut Barat	C2	Sedang	0,4751
16	Mekar Baru	C2	Sedang	0,3858
17	Sentang	C3	Rendah	0,5854
18	Kisaran Barat	C2	Sedang	0,4572
19	Teladan	C2	Sedang	0,4246
20	Karang Anyar	C3	Rendah	0,6177
21	Sidomulyo	C3	Rendah	0,4596
22	Subur	C3	Rendah	0,5555
	Rata-rata Silhouette Score			0,3981

Nilai rata-rata Silhouette Score yang diperoleh adalah 0,3981. Nilai

tersebut menunjukkan bahwa pengelompokan yang dihasilkan oleh

algoritma K-Medoids termasuk dalam kategori cukup baik (reasonable structure), yang berarti setiap wilayah memiliki kemiripan yang memadai dengan anggota clusternya sendiri dan perbedaan yang cukup signifikan terhadap cluster lain.

Berdasarkan hasil pengelompokan, Cluster 1 (Prioritas Tinggi) terdiri dari 1 wilayah dengan medoid Kisaran Baru; Cluster 2 (Prioritas Sedang) terdiri dari 16 wilayah dengan medoid Bunut Barat; serta Cluster 3 (Prioritas Rendah) terdiri dari 5 wilayah yaitu Sei Renggas, Sentang, Karang Anyar, Sidomulyo, dan Subur dengan medoid Karang Anyar. Distribusi jumlah wilayah setiap cluster disajikan pada Tabel 9.

Tabel 9 Distribusi dan Karakteristik Cluster

Cluster	Prioritas	Jumlah	Medoid	Karakteristik
C1	Tinggi	1	Kisaran Baru	Calon pelanggan tinggi, jarak dekat, kepadatan tinggi
C2	Sedang	16	Bunut Barat	Karakteristik menengah pada ketiga variabel
C3	Rendah	5	Karang Anyar	Calon pelanggan sedikit, jarak jauh, kepadatan rendah

Wilayah yang masuk ke dalam Cluster 1 merupakan wilayah yang paling diprioritaskan untuk mendapatkan pengembangan jaringan pipa baru, karena memiliki jumlah calon pelanggan yang tinggi, jarak yang dekat ke jaringan pipa eksisting, dan kepadatan penduduk yang

tinggi. Wilayah pada Cluster 2 memiliki potensi menengah sehingga dapat dijadwalkan pada tahap pengembangan berikutnya, sedangkan Cluster 3 memiliki potensi paling rendah sehingga pengembangannya dapat dipertimbangkan setelah kedua cluster lainnya.

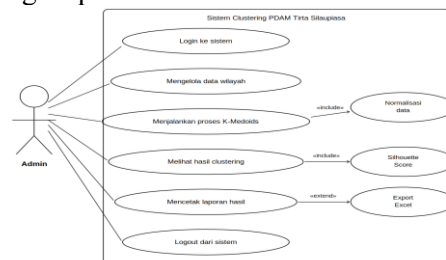
Hasil ini sejalan dengan penelitian terdahulu yang menunjukkan bahwa algoritma K-Medoids mampu menghasilkan pengelompokan yang representatif dan tahan terhadap data ekstrem. Adanya wilayah Sendang Sari dengan nilai kepadatan penduduk yang sangat tinggi (nilai normalisasi $X_3 = 1$) tidak menggeser pusat cluster secara signifikan karena medoid menggunakan objek data nyata, bukan nilai rata-rata. Hal ini membuktikan keunggulan K-Medoids dalam menangani outlier dibandingkan metode berbasis rata-rata seperti K-Means.

Pembahasan

Analisis Sistem

Analisis sistem dilakukan untuk memodelkan alur kerja sistem informasi clustering menggunakan pendekatan Unified Modeling Language (UML), yang meliputi Use Case Diagram, Activity Diagram, Sequence Diagram, dan Class Diagram.

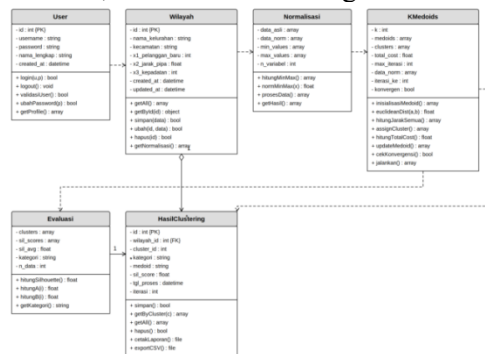
Use Case Diagram menggambarkan interaksi antara aktor Admin dengan sistem. Terdapat enam use case utama, di mana use case Menjalankan Proses K-Medoids memiliki relasi «include» dengan Normalisasi Data dan Silhouette Score karena kedua fungsi tersebut wajib dieksekusi secara otomatis, serta use case Mencetak Laporan memiliki relasi «extend» dengan Export Excel sebagai fungsi opsional.



Gambar 2 Use Case Diagram Sistem

Activity Diagram menggambarkan alur aktivitas untuk setiap use case, meliputi proses login, pengelolaan data wilayah, proses K-Medoids, tampilan hasil clustering, pencetakan laporan, dan logout. Pada proses K-Medoids, sistem mengambil data wilayah, melakukan normalisasi Min-Max, menginisialisasi medoid awal, menjalankan iterasi perhitungan jarak, penetapan cluster, pembaruan medoid, dan penghitungan total cost hingga konvergen, kemudian menghitung Silhouette Score.

Sequence	Diagram
menggambarkan urutan interaksi antar objek berdasarkan waktu, melibatkan empat objek utama yaitu Admin, Controllor, Model, dan View. Class Diagram menggambarkan struktur kelas sistem yang terdiri dari kelas User, Wilayah, Normalisasi, KMedoids, Evaluasi, dan Hasil Clustering.	



Gambar 3 Class Diagram Sistem

Implementasi Sistem

Sistem clustering wilayah pengembangan pasang baru jalur pipa PDAM Tirta Silaupiasa diimplementasikan sebagai aplikasi berbasis web menggunakan bahasa pemrograman PHP dan basis data MySQL, dengan antarmuka menggunakan framework Bootstrap 5. Sistem terdiri dari enam halaman utama. Berikut langkah-langkah menjalankan sistem yang dibangun.

Halaman login merupakan antarmuka pertama yang ditampilkan saat pengguna mengakses sistem. Pengguna diwajibkan memasukkan username dan password yang telah terdaftar.



Sistem Clustering Wilayah

PDAM Tirta Silaupiasa Kabupaten Asahan

 **Username**

 **Password**



PDAM Tirta Silaupiasa © 2026

Gambar 4 Halaman Login

Halaman dashboard merupakan halaman utama yang ditampilkan setelah Admin berhasil login, menampilkan kartu ringkas total wilayah, jumlah cluster, nilai Silhouette Score, dan tabel ringkasan hasil clustering.

Dashboard

General status, Administration, Backend, Integration, system, clustering.

22 Hosts Up

3 Cluster Down

3.981 Data

17 Jan 2024 Proses Transfer

Elegance Host Clustering

Cluster	Kategori	Status	Jumlah Hosts	Klik Detail
C1	Available Hosts	Active Data	3 - updata	
C2	Available Hosts	Active Data	16 - updata	
C3	Available Hosts	Active Data	3 - updata	
C4	Available Hosts	Active Data	3 - updata	

Informasi Proses Transfer

Jumlah Cluster 30	3
Jumlah Transfer	3
Total Cluster 30	3.9847
Total Data 30	5.6326
Gagal Transfer	3.9847
Status	Available
Target Proses	17 Jan 2024 10:39

Akses Log

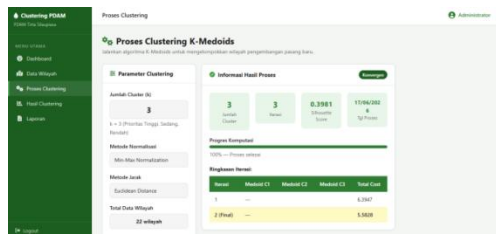
Gambar 5 Halaman Dashboard

Halaman data wilayah berfungsi untuk pengelolaan data wilayah, dilengkapi fitur pencarian, template CSV, import CSV secara massal, serta tombol tambah, edit, dan hapus data.

[illegible]

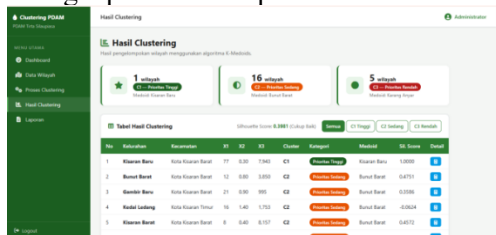
Gambar 6 Halaman Data Wilayah

Halaman proses clustering memungkinkan Admin menjalankan algoritma K-Medoids dengan menentukan jumlah cluster dan metode normalisasi, kemudian menampilkan jumlah iterasi, total cost, nilai Silhouette Score, dan status konvergensi.



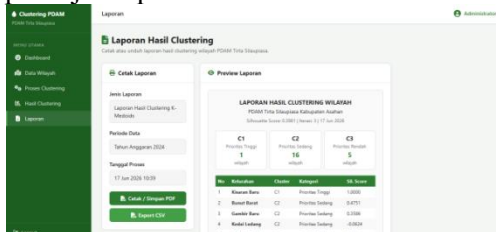
Gambar 7 Halaman Proses Clustering

Halaman hasil clustering menampilkan hasil pengelompokan seluruh wilayah dalam bentuk tabel lengkap beserta kolom cluster, kategori prioritas, dan nilai Silhouette Score, dilengkapi fitur filter per cluster.



Gambar 8 Halaman Hasil Clustering

Halaman laporan menyediakan fitur pencetakan hasil clustering dalam format PDF dan ekspor CSV, dilengkapi pratinjau laporan sebelum diunduh.



Gambar 9 Halaman Laporan

SIMPULAN

Berdasarkan hasil penelitian dan pembahasan mengenai clustering wilayah pengembangan pasang baru jalur pipa PDAM Tirta Silaupiasa Kabupaten Asahan menggunakan algoritma K-Medoids, dapat diambil kesimpulan sebagai berikut.

1. Algoritma K-Medoids berhasil diterapkan untuk mengelompokkan 22 wilayah kelurahan ke dalam tiga cluster berdasarkan tiga variabel input, yaitu jumlah calon pelanggan baru, jarak ke jaringan pipa eksisting,

dan kepadatan penduduk. Proses clustering dilakukan melalui normalisasi Min-Max, penentuan medoid awal, serta iterasi pengelompokan hingga konvergen setelah tiga iterasi dengan total cost menurun dari 6,3947 menjadi 5,3436.

2. Hasil pengelompokan dengan $k = 3$ menghasilkan Cluster 1 (Prioritas Tinggi) dengan 1 wilayah (Kisaran Baru), Cluster 2 (Prioritas Sedang) dengan 16 wilayah (medoid Bunut Barat), dan Cluster 3 (Prioritas Rendah) dengan 5 wilayah (medoid Karang Anyar).
3. Kualitas pengelompokan dievaluasi menggunakan Silhouette Score dengan nilai rata-rata 0,3981 yang termasuk kategori cukup baik (reasonable structure), sehingga hasil pengelompokan dapat diandalkan sebagai dasar pengambilan keputusan dalam perencanaan pengembangan jaringan pipa baru PDAM Tirta Silaupiasa.

DAFTAR PUSTAKA

- Afrilia, D., dkk. (2021). Perancangan Sistem Informasi Berbasis Unified Modeling Language (UML). Jurnal Teknik Informatika.
- Annisa, dkk. (2022). Penerapan Data Mining Pengelompokan Data Pengguna Air Bersih Berdasarkan Keluhannya Menggunakan Metode Clustering pada PDAM Langkat. Jurnal Sistem Informasi dan Ilmu Komputer.
- Effendie, & Kariyam. (2023). Indeks Rasio Deviasi Berbasis Medoid untuk Menentukan Jumlah Klaster dalam Suatu Dataset. Jurnal Statistika dan Aplikasinya.
- Elgamar. (2020). Konsep Dasar Pemrograman Website dengan PHP. Malang: Ahlimedia Press.
- Fasaee, M. A. K., dkk. (2021). Analysis of Water Distribution Network Systems. Water Resources Management.

-
- Fathoroni, dkk. (2020). Buku Saku Pemrograman Menggunakan Sublime Text. Bandung: Kreatif Industri Nusantara.
- Gusman, A. P. (2019). Pemrograman Web Dinamis dengan PHP dan MySQL. Jurnal Teknologi Informasi.
- Habibi, R., & Aprilian, R. (2019). Aplikasi Sistem Informasi Berbasis UML. Bandung: Kreatif Industri Nusantara.
- Hasnawiyah, dkk. (2022). Pengembangan Jaringan Perpipaan Air Bersih Berbasis Analisis Kebutuhan Wilayah. Jurnal Teknik Sipil.
- Ina, dkk. (2022). Penerapan Metode Clustering dalam Pengelompokan Data. Jurnal Sistem Informasi.
- Kadir, A. (2019). Pemrograman Database MySQL untuk Pemula. Yogyakarta: MediaKom.
- Kaban, R., & Sembiring, F. (2021). Pemanfaatan Web Browser dalam Pencarian Informasi. Jurnal Teknologi dan Sistem Informasi.
- Kaufman, L., & Rousseeuw, P. J. (2023). Finding Groups in Data: An Introduction to Cluster Analysis. Hoboken: Wiley.
- Maulia, dkk. (2021). Implementasi Data Mining dengan Teknik Clustering. Jurnal Informatika.
- Nata, dkk. (2021). Perancangan Website Sebagai Media Informasi. Jurnal Teknologi Informasi.
- Nugroho, dkk. (2021). Konsep Web Server dan Aplikasinya. Jurnal Rekayasa Perangkat Lunak.
- Purwaningsih, dkk. (2022). Konsep Dasar Sistem Informasi. Jurnal Teknologi Informasi.
- Sa'ad, M. I. (2020). Otodidak Web Programming: Membuat Website Edutainment. Jakarta: Elex Media Komputindo.
- Sureja, dkk. (2022). Systematic Data-Based Planning for Clean Water Supply Systems. Journal of Water Engineering.
- Susilowati, S. (2019). Perancangan Website Sebagai Media Promosi dan Informasi. Jurnal Sistem Informasi.
- Yafi, dkk. (2023). Penerapan Algoritma K-Medoids untuk Pengelompokan Data Wilayah. Jurnal Data Mining dan Sistem Informasi.