

Klasterisasi Prioritas Kerusakan Pipa PDAM Blora Cabang Cepu dengan Metode K-Means

Anang Triyulianto Nugraha Putra ^{1*}, Adhika Pramita Widayarsi ²

¹ Sekolah Tinggi Teknologi Ronggolawe ; anangtrijulianto@gmail.com

² Sekolah Tinggi Teknologi Ronggolawe ; dikasari9@gmail.com

Abstrak: Penelitian ini bertujuan mengelompokkan prioritas kerusakan pipa di PDAM Blora Cabang Cepu menggunakan algoritma K-Means berdasarkan tiga variabel: jarak lokasi (km), ukuran pipa (inci), dan durasi perbaikan (jam). Kebocoran pipa pada jaringan distribusi air bersih sering menyebabkan kerugian finansial dan menurunnya kualitas pelayanan, sehingga diperlukan solusi berbasis data untuk mendukung pengambilan keputusan. Data diperoleh dari 228 laporan pengaduan periode 2024–2025 melalui website Billing PDAM Blora. Tahapan penelitian meliputi preprocessing, normalisasi Min-Max, serta clustering K-Means yang dievaluasi dengan Elbow Method dan Silhouette Score. Proses clustering dilakukan menggunakan Python di Google Colab. Hasil evaluasi menunjukkan jumlah cluster optimal adalah tiga (K=3) dengan Silhouette Score tertinggi 0,660, mengungguli K=2 (0,645) dan K=4 (0,650). Ketiga cluster yang terbentuk adalah: Prioritas Tinggi (20 kejadian) dengan rata-rata jarak 2,50 km, ukuran pipa 5,35 inci, dan durasi 9,20 jam; Prioritas Sedang (117 kejadian) dengan jarak 3,72 km, ukuran 1,98 inci, durasi 1,58 jam; serta Prioritas Rendah (91 kejadian) dengan jarak 1,00 km, ukuran 2,11 inci, durasi 1,84 jam. Penelitian ini membuktikan bahwa K-Means efektif untuk menentukan prioritas penanganan kerusakan pipa secara objektif. Hasil clustering dapat menjadi dasar pengambilan keputusan operasional PDAM Cabang Cepu dalam menjadwalkan perbaikan dan alokasi sumber daya secara lebih efisien.

Kata Kunci: K-Means, Clustering, Kerusakan Pipa, PDAM, Prioritas Penanganan

DOI: <https://doi.org/10.47134/jacis.v6i2.183>

*Correspondensi: Anang Triyulianto Nugraha Putra

Email: anangtrijulianto@gmail.com

Receive: 29 April 2026

Accepted: 25 Mei 2026

Published: 1 Juni 2026



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Abstract: This study aims to classify pipe damage priorities at PDAM Blora Cepu Branch using the K-Means algorithm based on three variables: distance (km), pipe size (inches), and repair duration (hours). Pipe leaks in clean water distribution networks often cause financial losses and decline in service quality, therefore a data-driven solution is needed to support decision-making. Data were obtained from 228 complaint reports during the 2024–2025 period through the PDAM Blora Billing website. The research stages included preprocessing, Min-Max normalization, and K-Means clustering evaluated using the Elbow Method and Silhouette Score. The clustering process was conducted using Python on Google Colab. The evaluation results showed that the optimal number of clusters was three (K=3) with the highest Silhouette Score of 0.660, outperforming K=2 (0.645) and K=4 (0.650). The three clusters formed were: High Priority (20 incidents) with an average distance of 2.50 km, pipe size of 5.35 inches, and duration of 9.20 hours; Medium Priority (117 incidents) with a distance of 3.72 km, pipe size of 1.98 inches, duration of 1.58 hours; and Low Priority (91 incidents) with a distance of 1.00 km, pipe size of 2.11 inches, duration of 1.84 hours. This study proves that

K-Means is effective for objectively determining pipe damage repair priorities. The clustering results can serve as a basis for operational decision-making at PDAM Cepu Branch in scheduling repairs and allocating resources more efficiently.

Keywords: K-Means, Clustering, Pipe Damage, PDAM, Handling Priority

PENDAHULUAN

Sistem distribusi air bersih merupakan infrastruktur penting dalam memenuhi kebutuhan air bagi masyarakat. Perusahaan Daerah Air Minum (PDAM) memiliki peran penting dalam menyediakan layanan air bersih yang berkualitas melalui jaringan distribusi pipa yang luas dan kompleks. Namun dalam pelaksanaannya, sistem distribusi air sering menghadapi berbagai permasalahan teknis, salah satunya adalah kebocoran pada jaringan pipa yang dapat menyebabkan kehilangan air serta menurunkan efisiensi sistem distribusi [1].

Kebocoran pada jaringan pipa distribusi air merupakan permasalahan umum yang terjadi pada banyak sistem penyediaan air di berbagai daerah. Faktor seperti usia pipa, tekanan air yang tidak stabil, kondisi tanah, serta kualitas material pipa dapat menjadi penyebab utama terjadinya kerusakan pada jaringan distribusi air [2]. Kebocoran pipa yang tidak segera terdeteksi dapat mengakibatkan kerugian finansial bagi perusahaan pengelola air serta menurunkan kualitas pelayanan kepada masyarakat [3].

Perkembangan teknologi informasi dan analisis data memberikan peluang untuk mengatasi permasalahan tersebut melalui pemanfaatan teknik analisis data dalam pengelolaan sistem distribusi air. Pendekatan berbasis data memungkinkan pengelola jaringan untuk mengidentifikasi pola kerusakan atau kebocoran pipa secara lebih efektif sehingga proses pengambilan keputusan dapat dilakukan secara lebih cepat dan akurat [4]. Salah satu pendekatan yang banyak digunakan dalam analisis data adalah teknik data mining. Data mining merupakan proses pengolahan data dalam jumlah besar untuk menemukan pola, hubungan, maupun informasi yang berguna dalam pengambilan keputusan [5]. Teknik data mining telah banyak diterapkan dalam berbagai bidang, termasuk dalam analisis sistem distribusi air untuk mendeteksi serta mengidentifikasi pola kebocoran pada jaringan pipa [6].

Dalam data mining terdapat beberapa teknik analisis data, salah satunya adalah *clustering*. *Clustering* merupakan teknik pengelompokan data berdasarkan kesamaan karakteristik sehingga objek yang berada dalam satu kelompok memiliki tingkat kemiripan yang tinggi dibandingkan dengan objek pada kelompok lainnya [7]. Teknik ini banyak digunakan dalam berbagai penelitian untuk mengidentifikasi pola data yang tersembunyi tanpa memerlukan label data sebelumnya [8]. Metode *clustering* dapat membantu dalam mengelompokkan data operasional jaringan pipa seperti tekanan air, lokasi pipa, frekuensi kerusakan, serta jenis kerusakan sehingga dapat ditemukan pola kebocoran yang terjadi pada sistem distribusi air [9]. Dengan adanya pengelompokan data tersebut, pihak pengelola jaringan dapat menentukan prioritas perbaikan dan pemeliharaan jaringan secara lebih efektif [10].

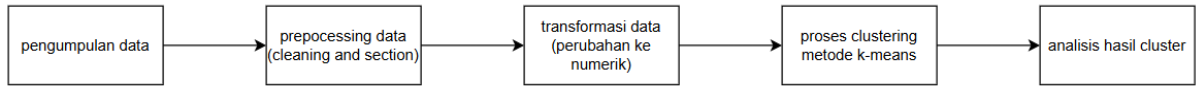
Salah satu algoritma *clustering* yang paling populer dalam data mining adalah K-Means. Algoritma K-Means bekerja dengan cara membagi data ke dalam sejumlah *cluster* berdasarkan jarak antara data dengan pusat *cluster* (*centroid*) sehingga data yang memiliki karakteristik serupa akan berada dalam kelompok yang sama [11]. Metode ini dikenal memiliki proses komputasi yang sederhana serta mampu mengolah data dalam jumlah besar secara efisien [12]. Selain itu, metode K-Means juga banyak digunakan dalam berbagai penelitian karena mampu menghasilkan pengelompokan data yang jelas serta mudah diinterpretasikan [13]. Keunggulan tersebut menjadikan metode K-Means sebagai salah satu metode *clustering* yang paling sering digunakan dalam analisis data [14].

Meskipun demikian, terdapat beberapa metode *clustering* lain yang juga sering digunakan dalam penelitian, di antaranya Fuzzy C-Means, DBSCAN, dan K-Medoids. Metode Fuzzy C-Means [15] memungkinkan suatu data memiliki tingkat keanggotaan pada lebih dari satu *cluster* sehingga lebih fleksibel dalam menangani data dengan batas *cluster* yang tidak jelas [16]. Namun metode ini memiliki kompleksitas perhitungan yang lebih tinggi sehingga membutuhkan waktu komputasi yang lebih lama dibandingkan metode K-Means [17]. Metode DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) merupakan metode *clustering* berbasis kepadatan yang mampu mengidentifikasi *cluster* dengan bentuk yang tidak beraturan serta dapat mendeteksi data pencilan atau noise [18]. Akan tetapi metode ini sangat bergantung pada pemilihan parameter kepadatan sehingga kurang optimal apabila digunakan pada dataset dengan distribusi data yang relatif seragam. Sementara itu, metode K-Medoids merupakan pengembangan dari metode K-Means yang menggunakan objek data sebagai pusat *cluster* atau medoid. Metode ini lebih tahan terhadap pengaruh data pencilan, namun memiliki proses komputasi yang lebih kompleks sehingga kurang efisien ketika digunakan pada dataset yang berukuran besar [19].

Berdasarkan perbandingan beberapa metode *clustering* tersebut, metode K-Means dipilih dalam penelitian ini karena memiliki proses perhitungan yang sederhana, efisien dalam mengolah data berukuran besar, serta mampu menghasilkan pengelompokan data yang jelas dan mudah diinterpretasikan. Keunggulan tersebut menjadikan metode K-Means sangat sesuai untuk digunakan dalam analisis data operasional jaringan pipa PDAM [20]. Penelitian ini bertujuan untuk mengelompokkan tingkat kebocoran pipa ke dalam kategori kecil, sedang, dan besar menggunakan metode *clustering* K-Means berdasarkan karakteristik teknis dan data operasional jaringan pipa PDAM Blora. Dengan adanya pengelompokan tersebut diharapkan dapat membantu dalam mengidentifikasi pola kebocoran serta mendukung pengambilan keputusan dalam proses pemeliharaan jaringan distribusi air secara lebih efektif dan efisien.

METODE

Penelitian ini menggunakan K-Means untuk mengelompokkan tingkat prioritas kerusakan pipa di PDAM Blora Cabang Cepu. Data kerusakan seperti lokasi, frekuensi, dan lama perbaikan dikumpulkan lalu dinormalisasi. Selanjutnya, data diproses dengan K-Means [21] menggunakan jarak *Euclidean* hingga terbentuk tiga *cluster* prioritas: tinggi, sedang, dan rendah. Hasil *cluster* kemudian dicek kesesuaiannya dengan kondisi di lapangan. Tahapan alur penelitian yang dilakukan, mulai dari pengumpulan data hingga analisis hasil *clustering*, digambarkan secara sistematis pada Gambar 1.



Gambar 1. Tahapan Penelitian

Berikut penjelasan setiap tahapan-tahapan yang akan dilakukan pada penelitian:

1. Pengumpulan Data

Data penelitian ini dikumpulkan dari sistem pelaporan kerusakan pipa PDAM Cabang Cepu yang mencakup 228 laporan pengaduan. Data yang diambil meliputi tiga variabel utama yaitu kode wilayah kejadian, ukuran pipa yang rusak (dalam satuan inci), dan durasi penanganan (dalam satuan jam) yang dihitung dari waktu respons hingga waktu selesai perbaikan.

a) Sumber Data

Dataset yang digunakan berasal dari website resmi Billing PDAM Blora yang meliputi semua pengaduan di cabang PDAM Blora. Dataset tersebut mencatat jumlah pengaduan berdasarkan berbagai jenis semua pengaduan kebocoran PDAM Kabupaten Blora selama Periode 2024-2025, sehingga memungkinkan untuk melakukan analisis guna mengelompokkan pengaduan kebocoran pipa. Dataset yang memiliki catatan lengkap mengenai pengaduan kebocoran pipa dari semua cabang PDAM Kabupaten Blora yaitu PDAM Kabupten Blora cabang Cepu. Pengaduan kebocoran pipa di PDAM Kabupaten Blora cabang Cepu berjumlah 228 data dengan 10 atribut, yaitu nomor urutan pengaduan, tanggal pengaduan, cabang, nomor pelanggan, nama pengaduan, alamat pengaduan, keterangan pengaduan, tanggal resepon pengaduan, tanggal tindakan pengaduan, dan tanggal hasil pengaduan. Gambar 2 menunjukkan dataset awal pengaduan PDAM Blora Cabang Cepu dalam bentuk excel.

DAFTAR PENGADUAN

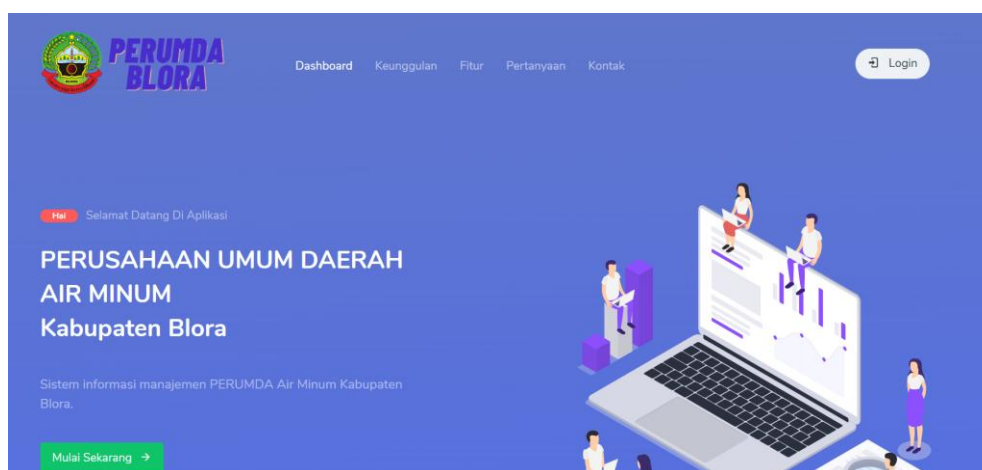
TANGGAL : 01/01/2024 - 31/12/2025
 CABANG : CEPU
 JENIS PENGADUAN: BOCOR PIPA TRANS DIST

NO	TGL	CABANG	NO PELANGGAN	NAMA PENGADUAN	ALAMAT	KETERANGAN	TGL RESPON	TGL TINDAKAN	TGL HASIL
1	01/02/2025	2	204051797	AGUS RIFAI	TAMBAKWATU	PIPA BOCOR 2"	01/02/2025	01/02/2025	01/02/2025
2	01/04/2024	2	204040525	SUYONO	JATIREJO	PIPA BOCOR 2"	01/04/2024	01/04/2024	01/04/2024
3	01/07/2025	2	204050637	KUSNIO	CEPU ASRI	PIPA BOCOR DI JALAN	01/07/2025	01/07/2025	01/07/2025
4	02/01/2024	2	204051118	BUDI	CEPU ASRI	PIPA BOCOR 6"	02/01/2024	02/01/2024	02/01/2024
5	02/01/2025	2	204050807	ROCHIM	NGLEBOK	PIPA BOCOR DI JALAN	02/01/2025	02/01/2025	02/01/2025
6	02/01/2025	2	203072795	SUYONO	PERUM MEGAL PERMAY	PIPA BOCOR DI JALAN	02/01/2025	02/01/2025	02/01/2025
7	02/02/2024	2	204050792	BUDI-HARIYANTO	SUMBER AGUNG	PIPA BOCOR DI JALAN	02/02/2024	02/02/2024	02/02/2024
8	02/02/2025	2	204051523	SUYONO	MENTUL	PIPA BOCOR 2"	02/02/2025	02/02/2025	02/02/2025
.....
221	30/01/2024	2	204050803	AGUS RIFAI	JATIREJO	PIPA BOCOR 2"	30/01/2024	30/01/2024	30/01/2024
222	30/01/2025	2	204051171	AGUS KRISWANTO-SUYONO	JL GAJAH MADA MENTUL TURIBANG	PIPA BOCOR DI SEMAK SEMAK	30/01/2025	30/01/2025	30/01/2025
223	30/03/2024	2	201010194	AGUS RIFAI	BALUN LR IV	PIPA BOCOR DI JALAN	30/03/2024	30/03/2024	30/03/2024
224	30/05/2024	2	204051665	BUDI	MUTIARA BLOK D7	PIPA BOCOR	30/05/2024	30/05/2024	30/05/2024
225	30/06/2025	2	203071469	RIYANTO-AHAW	GRAHA SENTOSA	PIPA BOCOR DI JALAN	30/06/2025	30/06/2025	30/06/2025
226	30/08/2024	2	205010160	AGUS RIFAI	RUSUNAWA	PIPA BOCOR DI SAWAH	30/08/2024	30/08/2024	30/08/2024
227	30/09/2025	2		BUDI	JL BY PASS	PIPA BOCOR DI JALAN	30/09/2025	30/09/2025	30/09/2025
228	31/05/2025	2	204051797	BACHTIAR ARIF	TAMBAKWATU KARANGBOYO	PIPA BOCOR DI JALAN	31/05/2025	31/05/2025	31/05/2025

Gambar 2. Dataset Awal Pengaduan PDAM Blora Cabang Cepu

b) Teknik Pengumpulan Data

Teknik pengumpulan data yang dilakukan dengan mengunduh dataset dari Billing PDAM Blora Dataset yang digunakan berasal dari website resmi *Billing* PDAM Blora, yang merupakan sumber yang dapat dipercaya. Selain itu, dilakukan proses verifikasi dan validasi data untuk memastikan akurasi. Langkah ini penting untuk menjamin bahwa data yang digunakan dalam analisis bebas dari kesalahan teknis maupun konseptual, serta relevan dengan tujuan penelitian. Gambar 3 adalah tampilan website Billing PDAM Blora.



Gambar 3. Website Billing PDAM Blora

2. *Preprocessing* Data

Sebelum data bisa dianalisis, harus dilakukan *preprocessing* terlebih dahulu. Ini adalah tahap membersihkan dan menyiapkan data mentah (data awal). Karena data dari lapangan atau sistem itu jarang yang sempurna. Seringkali ditemukan data kosong, format yang tidak seragam, nilai yang tidak wajar, atau skala antar variabel yang berbeda. *Preprocessing* hadir untuk memperbaiki semua kekurangan tersebut.

Tahapan ini dimulai dengan pemilihan data yang relevan dari dataset yang berisi informasi pengaduan kebocoran pipa di PDAM Blora cabang Cepu berdasarkan wilayah pengaduan, ukuran pipa, dan durasi pengerjaan pengaduan. Dalam proses ini, hanya atribut yang dianggap relevan untuk analisis yang diambil. Pemilihan ini bertujuan untuk memfokuskan analisis hanya pada pengaduan kebocoran pipa. Gambar 4 merupakan hasil cleaning dan section.

DAFTAR PENGADUAN

TANGGAL : 01/01/2024 - 31/12/2025
 CABANG : CEPU
 JENIS PENGADUAN: BOCOR PIPA TRANS DIST

NO	ALAMAT	JARAK (KM)	KETERANGAN	UKURAN PIPA	TGL RESPON	TGL TINDAKAN	TGL HASIL	DURASI (JAM)
1	TAMBAKWATU	4,3	PIPA BOCOR 2"	2	01/02/2025	01/02/2025	01/02/2025	1
2	JATIREJO	4,3	PIPA BOCOR 2"	2	01/04/2024	01/04/2024	01/04/2024	1
3	CEPU ASRI	4,3	PIPA BOCOR DI JALAN	2	01/07/2025	01/07/2025	01/07/2025	1
4	CEPU ASRI	4,3	PIPA BOCOR 6"	6	02/01/2024	02/01/2024	02/01/2024	7
5	NGLEBOK	3,4	PIPA BOCOR DI JALAN	2	02/01/2025	02/01/2025	02/01/2025	1
6	PERUM MEGAL PERMAY	1	PIPA BOCOR DI JALAN	2	02/01/2025	02/01/2025	02/01/2025	2
7	SUMBER AGUNG	4,3	PIPA BOCOR DI JALAN	2	02/02/2024	02/02/2024	02/02/2024	2
8	MENTUL	4,3	PIPA BOCOR 2"	2	02/02/2025	02/02/2025	02/02/2025	3
9	PASAR BALUN	1	PIPA BOCOR DI DEKAT GAPURA	2	02/03/2025	02/03/2025	02/03/2025	1
10	BALUN LR 4	1	PIPA BOCOR 4"	4	02/03/2025	02/03/2025	02/03/2025	5
11	JL JATIREJO LR 5	4,3	PIPA BOCOR 1"	1	02/04/2024	03/04/2024	03/04/2024	1
12	JL BYPASS	1	PIPA BOCOR 2"	2	02/04/2024	03/04/2024	03/04/2024	2
13	CEPU GANG 8	3	PIPA BOCOR	2	02/07/2024	02/07/2024	02/07/2024	1
14	JL RONGGOLAWE	1	PIPA BOCOR DI JALAN	2	02/09/2024	02/09/2024	02/09/2024	2
15	JL BRI	1	PIPA BOCOR DI JEMBATAN	2	02/09/2024	02/09/2024	02/09/2024	2
...
220	TAMBAKWATU DEPAN PONDOK AL ISTIGHFAR	4,3	PIPA BOCOR PINGGIR JALAN	2	29/11/2025	29/11/2025	29/11/2025	1
221	JATIREJO	4,3	PIPA BOCOR 2"	2	30/01/2024	30/01/2024	30/01/2024	1
222	JL GAJAH MADA MENTUL TURIBANG	4,3	PIPA BOCOR DI SEMAK SEMAK	2	30/01/2025	30/01/2025	30/01/2025	2
223	BALUN LR IV	1	PIPA BOCOR DI JALAN	2	30/03/2024	30/03/2024	30/03/2024	2
224	MUTIARA BLOK D7	1	PIPA BOCOR	2	30/05/2024	30/05/2024	30/05/2024	1
225	GRAHA SENTOSA	4,3	PIPA BOCOR DI JALAN	2	30/06/2025	30/06/2025	30/06/2025	2
226	RUSUNAWA	1	PIPA BOCOR DI SAWAH	2	30/08/2024	30/08/2024	30/08/2024	1
227	JL BY PASS	1	PIPA BOCOR DI JALAN	2	30/09/2025	30/09/2025	30/09/2025	2
228	TAMBAKWATU KARANGBOYO	4,3	PIPA BOCOR DI JALAN	2	31/05/2025	31/05/2025	31/05/2025	1

Gambar 4. Data Hasil Cleaning dan Section

3. Transformasi Data

Mengubah data yang telah melalui *preprocessing* agar dapat digunakan secara optimal oleh algoritma yang akan diterapkan. Proses transformasi data dalam penelitian ini mencakup inisialisasi dataset, penyesuaian format data, serta penggantian nama variabel agar lebih sesuai untuk pemrosesan lebih lanjut di Google Colab. Tahapan transformasi data dilakukan untuk memastikan dataset siap diolah secara optimal oleh algoritma, yang rincian prosesnya disajikan pada Tabel 1.

Tabel 1. Trasnformasi Data

Jarak (Kilometer)	Ukuran Pipa (Inch)	Durasi (Jam)
4,3	2	1
4,3	2	1
4,3	2	1
4,3	6	7
3,4	2	1
1	2	2
4,3	2	2
4,3	2	3
1	2	1
1	4	5
4,3	1	1
1	2	2
3	2	1
1	2	2
1	2	2
1	2	2
1	1	1

4. Proses *Clustering*

Data yang sudah bersih dan seragam kemudian dikelompokkan menggunakan algoritma K-Means dengan 3 *cluster*. Proses *clustering* K-means dilakukan dengan python di Google Colab. Algoritma bekerja dengan menghitung jarak setiap data ke titik pusat *cluster*, memasukkannya ke *cluster* terdekat, lalu mengulang proses hingga titik pusat stabil. Adapun tahapan K-means sebagai berikut:

a) Normalisasi Data

Normalisasi data dilakukan untuk menyamakan skala nilai pada dataset agar berada dalam rentang yang seragam, yaitu antara 0 hingga 1, guna menghindari dominasi variabel dengan rentang nilai besar terhadap hasil perhitungan algoritma. Dalam penelitian ini, proses tersebut dihitung menggunakan metode *Min-Max Scaling* melalui persamaan (1).

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

Keterangan:

X_{norm} : nilai hasil normalisasi (rentang 0 sampai 1)

X : nilai data asli

X_{min} : nilai minimum dari seluruh data pada variabel tersebut

X_{max} : nilai maksimum dari seluruh data pada variabel tersebut

b) Menghitung Penentuan Centroid Awal

Tahapan penentuan centroid awal merupakan langkah krusial dalam algoritma K-Means, yang kemudian akan diperbarui secara iteratif melalui penghitungan rata-rata posisi seluruh anggota dalam satu kelompok. Untuk menentukan koordinat pusat atau centroid baru, digunakan persamaan (2)

$$v = \frac{\sum_{i=1}^n x_i}{n} \quad (2)$$

Dengan $i = 1, 2, 3, \dots, n$

Atau dalam bentuk yang lebih sederhana:

$$v = \frac{(x_1 + x_2 + x_3 + \dots + x_n)}{n} \quad (3)$$

Keterangan:

v : centroid baru (nilai rata-rata)

x_i : nilai data ke- i dalam suatu *cluster*

n : jumlah data dalam cluster tersebut

Σ : penjumlahan seluruh data dari $i=1$ sampai n

c) Menentukan Jarak Pusat Cluster

Jarak antara setiap data dengan titik pusat cluster dihitung menggunakan rumus Euclidian Distance. Setelah nilainya dibandingkan, data akan memilih pusat *cluster* yang paling dekat dan bergabung ke dalam kelompoknya. Persamaan (4) adalah rumus *Euclidian Distance* yang dimaksud.

$$d(x,y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

Dengan $i = 1, 2, 3, \dots, n$

Atau dalam bentuk yang lebih sederhana:

$$d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 + \dots + (x_n - y_n)^2} \quad (5)$$

Keterangan:

d : jarak Euclidean antara dua titik

x : titik data (nilai normalisasi dari variabel)

y : titik centroid (pusat cluster)

n : jumlah variabel yang digunakan (dalam kasus ini: wilayah, ukuran pipa, dan durasi)

d) Pengelompokan data

Ketika sekumpulan data ditemukan berada paling dekat dengan salah satu *centroid*, maka data tersebut secara otomatis akan dimasukkan ke dalam kelompok yang sesuai dengan *centroid* terdekat itu.

e) Perhitungan Ulang

Langkah selanjutnya adalah mengulangi proses dan menggunakan rumus yang sama untuk menentukan letak *centroid* yang baru.

f) Kondisi Pengulangan

Proses akan kembali ke langkah 3 (pengelompokan ulang) jika masih terjadi perpindahan data antar kelompok, jika perubahan nilai *centroid* melebihi batas toleransi yang telah ditentukan, atau jika nilai fungsi tujuan yang digunakan masih berubah.

5. Analisis Hasil *Cluster*

Setelah *clustering* selesai, setiap cluster dianalisis karakteristiknya. Dihitung rata-rata wilayah, ukuran pipa, dan durasi dari masing-masing cluster. Hasilnya kemudian diinterpretasikan untuk mengetahui cluster mana yang termasuk ringan, sedang, atau berat.

HASIL DAN PEMBAHASAN

Hasil Normalisasi Dataset

Proses normalisasi menggunakan metode Min-Max telah berhasil mengonversi variabel Jarak Kantor Ke Titik Perbaikan dalam Kilometer, Ukuran Pipa, Durasi Jam. Langkah ini memastikan seluruh atribut memiliki skala yang seragam sehingga siap untuk diolah oleh algoritma clustering. Cuplikan hasil transformasi dataset tersebut ditunjukkan pada Gambar 5.

JARAK_KM	UKURAN_PIPA_INCH	DURASI_JAM
43.000000	43.000000	43.000000
2.704651	3.232558	4.779070
1.407286	1.743521	3.798578
1.000000	1.000000	1.000000
1.000000	2.000000	2.000000
3.000000	3.000000	3.000000
4.300000	4.000000	7.500000
4.300000	7.000000	15.750000

*228 rows x 4 columns

Gambar 5. Hasil Normalisasi Dataset

Hasil Clustering Prioritas Ringan, Sedang, Berat dengan Google Colab

Proses clustering menggunakan Google Colab telah berhasil membagi data ke dalam tiga tingkat prioritas, yaitu Tinggi, Sedang, dan Rendah. Hasil pengelompokan ini ditentukan berdasarkan kemiripan karakteristik data yang telah diolah. Detail pembagian kategori tersebut dapat dilihat pada Gambar 6.

Berdasarkan gambar dari proses *clusterisasi* dengan algoritma *K-Means* pada 228 data kerusakan pipa di PDAM Blora Cabang Cepu, teridentifikasi tiga tingkatan prioritas yang berbeda. *Cluster 0* (Prioritas Tinggi) mencakup 20 kasus dengan ciri rata-rata jarak tempuh 2,50 km, diameter pipa 5,35 inci, serta waktu perbaikan rata-rata 9,20 jam (kisaran 2 hingga 24 jam), sehingga wajib ditangani secepatnya karena memakan waktu lama dan memerlukan sumber daya besar.

```

Cluster 0: PRIORITAS TINGGI
  Jumlah kejadian   : 20
  Rata-rata Jarak   : 2.50 km
  Rata-rata Ukuran  : 5.35 inch
  Rata-rata Durasi  : 9.20 jam
  Rentang Durasi    : 2.0 - 24.0 jam

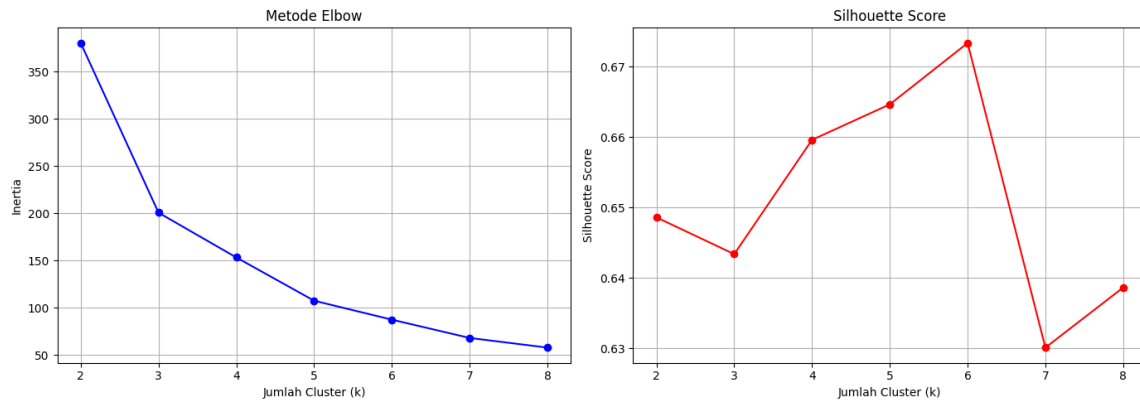
Cluster 1: PRIORITAS SEDANG
  Jumlah kejadian   : 117
  Rata-rata Jarak   : 3.72 km
  Rata-rata Ukuran  : 1.98 inch
  Rata-rata Durasi  : 1.58 jam
  Rentang Durasi    : 1.0 - 5.0 jam

Cluster 2: PRIORITAS RENDAH
  Jumlah kejadian   : 91
  Rata-rata Jarak   : 1.00 km
  Rata-rata Ukuran  : 2.11 inch
  Rata-rata Durasi  : 1.84 jam
  Rentang Durasi    : 1.0 - 5.0 jam

```

Gambar 6. Hasil *Clustering* K=3

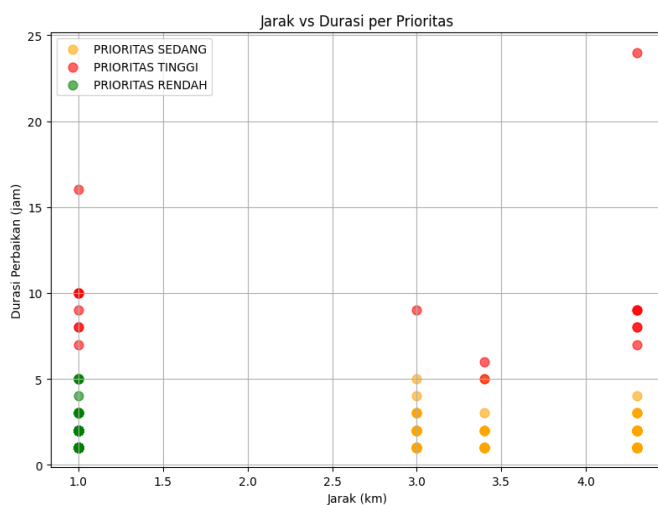
Cluster 1 (Prioritas Sedang) menjadi kelompok paling dominan dengan 117 kejadian, memiliki rata-rata jarak 3,72 km, ukuran pipa 1,98 inci, dan durasi perbaikan 1,58 jam (kisaran 1-5 jam), di mana lokasinya paling jauh namun waktu perbaikannya relatif singkat sehingga masih bisa dijadwalkan dalam rentang 4-12 jam. *Cluster 2* (Prioritas Rendah) terdiri dari 91 kejadian dengan rata-rata jarak 1,00 km, ukuran pipa 2,11 inci, dan durasi 1,84 jam (kisaran 1-5 jam), yang karena lokasinya sangat dekat dan waktu. Gambar 7 adalah visualisasi Metode *Elbow* dan *Silhouette Score* pada *K-Means Clustering*.



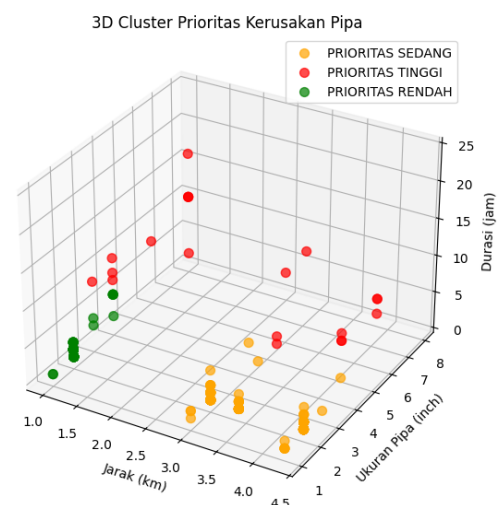
Gambar 7. visualisasi Metode *Elbow* dan *Silhouette Score*

Berdasarkan data yang tersaji dalam gambar 7, dapat disimpulkan bahwa jumlah *cluster* terbaik adalah 3, karena didukung secara konsisten oleh dua metode evaluasi. Pada *Metode Elbow*, penurunan nilai inersia terbesar terjadi dari $k=2$ ke $k=3$ (dari 2375 menjadi 2000), kemudian penurunannya mulai berkurang drastis pada $k=4$ dan seterusnya, menandakan titik siku (*elbow*) berada di $k=3$. Sementara itu, pada *Metode Silhouette Score*, nilai tertinggi juga diraih oleh $k=3$, yaitu 0,660, yang berarti *cluster* pada $k=3$ memiliki keseimbangan terbaik antara kekompakan internal dan pemisahan antar *cluster* dibandingkan k lainnya (misalnya $k=2$: 0,645; $k=4$: 0,650; $k=5$: 0,645). Dengan demikian, keselarasan hasil dari kedua metode ini memperkuat bukti bahwa pengelompokan data ke dalam tiga *cluster* merupakan pilihan yang paling optimal.

Gambar 8 menunjukkan jarak dengan durasi per prioritas dan Gambar 9 menunjukkan *cluster* berdasarkan wilayah dengan durasi.



Gambar 8. *Cluster* Berdasarkan Jarak Vs Durasi per Prioritas



Gambar 9. Visual 3D Prioritas Kerusakan Pipa

Dari gambar 8. dan gambar 9, terlihat jelas bahwa tiga cluster ini punya karakteristik yang benar-benar beda. Prioritas Tinggi (merah) adalah yang paling darurat karena durasi perbaikinya paling lama (9,20 jam) dan ukuran pipa paling besar (5,35 inci), meskipun jaraknya lumayan, artinya kerusakannya rumit dan butuh penanganan segera. Prioritas Sedang (kuning) punya jarak paling jauh (3,72 km) tapi durasi dan ukuran pipanya kecil, jadi masih bisa menyusul. Sementara Prioritas Rendah (biru) paling ringan karena jaraknya paling dekat (1 km), durasi singkat, dan ukuran pipa kecil, sehingga bisa dikerjakan belakangan. Jadi, pengelompokan jadi tiga prioritas ini udah pas dan masuk akal buat kondisi di lapangan.

SIMPULAN

Penelitian ini berhasil menerapkan algoritma K-Means untuk mengelompokkan prioritas kerusakan pipa PDAM Cabang Cepu berdasarkan tiga variabel yaitu wilayah, ukuran pipa, dan durasi penanganan. Dari 228 data pengaduan, dihasilkan tiga *cluster* dengan karakteristik berbeda. *Cluster 0* (81 kasus) dan *Cluster 2* (124 kasus) dikategorikan sebagai kerusakan ringan dengan ukuran pipa rata-rata 2 inci dan durasi penanganan kurang dari 2 hari, sehingga menjadi prioritas rendah. Sebaliknya, *Cluster 1* (23 kasus) merupakan kerusakan berat dengan ukuran pipa rata-rata 5,17 inci dan durasi penanganan mencapai 8,57 hari, sehingga harus menjadi prioritas tinggi. Ukuran pipa terbukti sebagai indikator yang lebih kuat dalam memprediksi durasi perbaikan dibandingkan wilayah. Metode K-Means efektif digunakan untuk menentukan prioritas penanganan kerusakan pipa secara objektif dan dapat dijadikan dasar pengambilan keputusan operasional PDAM Cabang Cepu.

Penelitian selanjutnya disarankan untuk menambahkan variabel lain seperti jenis kerusakan, lokasi kerusakan, dan ketersediaan suku cadang guna memperkaya analisis. Perbandingan dengan metode *clustering* lain seperti K-Medoids atau DBSCAN juga perlu dilakukan untuk menentukan algoritma terbaik. Hasil clustering sebaiknya diintegrasikan ke dalam sistem informasi PDAM secara real-time agar setiap pengaduan baru dapat langsung dikelompokkan dan ditentukan prioritas penanganannya secara otomatis. Selain itu, model perlu dievaluasi secara berkala setiap 6-12 bulan serta memperluas cakupan data ke seluruh cabang PDAM Blora agar diperoleh gambaran yang lebih komprehensif.

DAFTAR PUSTAKA

- [1] J. Lo Presti, C. Giudicianni, C. Toffanin, E. Creaco, L. Magni, and G. Galuppini, "Combining clustering and regularised neural network for burst detection and localization and flow/pressure sensor placement in water distribution networks," *J. Water Process Eng.*, vol. 63, no. April, p. 105473, 2024, doi: 10.1016/j.jwpe.2024.105473.
- [2] W. Huang, X. Huang, Z. Chen, J. Zhan, H. Yang, and X. Li, "Parameter Analysis and Optimization of a Leakage Localization Method Based on Spatial Clustering," *Water (Switzerland)*, vol. 17, no. 1, 2025, doi: 10.3390/w17010106.
- [3] G. M. Komba, T. E. Mathonsi, and P. A. Owolawi, "Water Pipeline Leak Detection and Localisation in Water Distribution Networks," *2023 Int. Conf. Emerg. Trends Networks Comput. Commun. ETNCC 2023 - Proc.*, pp. 140–145, 2023, doi: 10.1109/ETNCC59188.2023.10284934.

- [4] Lucas Roberto Tomazini, Rodrigo Pita Rolle, Eduardo Paciência Godoy, Esther Luna Colombini, and Alexandre da Silva Simões, "GraphLeak: A Realistic Dataset for Analyzing Leaks in Water Distribution Systems," *Proc. do Congr. Bras. Automática*, 2024, doi: 10.20906/cba2024/4655.
- [5] X. Li, S. Chu, T. Zhang, T. Yu, and Y. Shao, "Leakage localization using pressure sensors and spatial clustering in water distribution systems," *Water Supply*, vol. 22, no. 1, pp. 1020–1034, 2022, doi: 10.2166/ws.2021.219.
- [6] A. Fatkhudin, A. Khambali, F. A. Artanto, M. Mundriyah, and N. A. P. Zade, "Implementasi Algoritma Clustering K-Means Dalam Pengelompokan Mahasiswa," *J. Minfo Polgan*, vol. 12, no. 1, pp. 777–783, 2023, doi: 10.33395/jmp.v12i1.12494.
- [7] J. Faran and A. Triayudi, "Penerapan Algoritma K-Means Data Mining untuk Clustering KinerjaKaryawan Koperasi," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 4, pp. 2096–2108, 2024, doi: 10.30865/klik.v4i4.1728.
- [8] A. Mulyana, Y. Hermawan, and N. J. Saputri, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Rekomendasi Pilihan Program Studi Pada Mahasiswa Baru (Studi Kasus di Institut Bisnis dan Informatika Kesatuan)," *KERNEL J. Ris. Inov. Bid. Inform. dan Pendidik. Inform.*, vol. 5, no. 1, pp. 60–72, 2024, doi: 10.31284/j.kernel.2024.v5i1.7624.
- [9] H. Saputra *et al.*, "Sistem Informasi Geografis Kebocoran Pipa Air Berbasis Web," vol. 12, no. 2, 2022.
- [10] K. Saputra, M. U. Harun, A. Rasyid, and M. L. Perceptron, "Deteksi Kebocoran Pipa Air Menggunakan Machine," pp. 110–121, 2022.
- [11] P. Air, M. Tirta, and S. Asahan, "Pengelompokan Wilayah Berdasarkan Kubik Air menggunakan Algoritma K-Means pada Perumda Air Minum Tirta Silaupiasa Kabupaten Asahan Regional Grouping Based on Cubic Water Using K - Means Algorithm at," vol. 13, pp. 789–801, 2025.
- [12] E. Kurnia, A. P. Juledi, I. R. Munthe, and M. Nasution, "Penerapan Data Mining Menggunakan Algoritma," vol. 9, no. 2, pp. 3551–3559, 2025, [Online]. Available: www.penerbitlitnus.co.id
- [13] J. Ilmiah, D. Karya, P. Teknik, I. Fakultas, and U. M. Bengkulu, "Pengelompokan Pelanggan Pdam Berdasarkan Penggunaan Air Oleh Konsumen Menggunakan Algoritma K-Medoids Perusahaan Air Minum Daerah (PERUMDA) Tirta Hidayah Kota Bengkulu " Pengelompokan Data Penjualan Mie Berdasarkan Bulan Dengan Menggunakan Algoritma," vol. 1, no. 5, 2023.
- [14] K. Annisa, B. S. Ginting, and M. A. Syari, "Penerapan Data Mining Pengelompokan Data Pengguna Air Bersih Berdasarkan Keluhannya Menggunakan Metode Clustering Pada PDAM Langkat," vol. 6341, no. April, 2022.
- [15] P. Pane, "Optimasi Pendistribusian Air Minum Kemasan Menggunakan Metode Improved Zero Poin," vol. 3, no. 3, pp. 127–134, 2024, doi: 10.47065/jogtc.v3i3.6263.
- [16] F. S. Tambunan, "Perencanaan Jaringan Distribusi Sistem Penyediaan Air Bersih di Kecamatan Kutalimbaru Kabupaten Deli Serdang Dengan Sumber Air Baku Sungai Tuntungan," vol. 9, pp. 12–21, 2026.
- [17] R. Liu *et al.*, "Leak detection in water distribution networks using micro-electromechanical systems-based accelerometers: a machine learning approach," *Water Pract. Technol.*, vol. 20, no. 9, pp. 1900–1920, 2025, doi: 10.2166/wpt.2025.109.
- [18] A. T. Zy and W. Hadikristanto, "E ISSN : 2809-4069 Klasifikasi Kualitas Air dengan K-Means dan Decision Tree," vol. 5, no. 2, pp. 166–176, 2025.

-
- [19] R. Z. Singal and N. A. Jamal, "Perencanaan Sistem Jaringan Distribusi Air Bersih (Studi Kasus Desa Panca Agung Kabupaten Bulungan)," vol. 8, no. 2, pp. 108–119, 2022.
- [20] D. C. Rada, U. Maritim, and R. Ali, "Analisis Kebutuhan Air Bersih dan Kapasitas Jaringan Distribusi Wilayah Pengaliran Bambang Utoyo Dwi Cahya Rada Corresponding Author : dwicahyarada@umrah.ac.id penyediaan air bersih , terutama karena keterbatasan sumber daya air dan tidak meratanya".
- [21] K. Gustipartsani *et al.*, "Data Mining Clustering Menggunakan Algoritma K-Means Pada," vol. 7, no. 6, pp. 3595–3601, 2023.