

PERBANDINGAN ALGORITMA K-MEANS DAN K-MEDOIDS UNTUK CLUSTERING PENDERITA PNEUMONIA DI KABUPATEN KARAWANG

Muhammad Hardiansyah^{1*}, Betha Nurina Sari², Iqbal Maulana³

^{1,2,3}Universitas Singaperbangsa Karawang; Jl. HS. Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361

Keywords:

Pneumonia;
Clustering;
K-Means;
K-Medoids.

Correspondent Email:

mhardiansyah2810@gmail.com

Abstrak. *Pneumonia* merupakan penyakit serius yang memerlukan perhatian khusus, terutama dalam mengidentifikasi wilayah dengan jumlah kasus tinggi. Kabupaten Karawang dipilih sebagai lokasi penelitian karena memiliki jumlah kasus *pneumonia* yang cukup signifikan, diduga dipengaruhi oleh faktor lingkungan seperti keberadaan kawasan industri. Penelitian ini bertujuan untuk mengelompokkan wilayah berdasarkan jumlah penderita *pneumonia* menggunakan algoritma *K-means* dan *K-medoids*. Data yang digunakan berupa data sekunder dari Dinas Kesehatan Kabupaten Karawang periode 2019–2023. Proses analisis mengikuti tahapan *Knowledge Discovery in Databases* (KDD) meliputi *Data Selection*, *Pre-processing*, *Transformation*, *Data mining*, *Evaluasi* dan *Knowledge*. Tahapan normalisasi data menggunakan *StandardScaler* serta reduksi dimensi dengan *Principal Component Analysis* (PCA). Hasil evaluasi menunjukkan bahwa *K-means* menghasilkan performa *clustering* yang lebih baik, dengan nilai *Silhouette Score* sebesar 0.8066 dan *Davies-Bouldin Index* sebesar 0.1231. Sebaliknya, *K-medoids* menunjukkan hasil yang kurang optimal dengan *Silhouette Score* sebesar 0.5847 dan *Davies-Bouldin Index* sebesar 1.1531. Hasil *clustering* membentuk dua klaster wilayah berdasarkan tingkat sebaran kasus *pneumonia*, yaitu rendah dan tinggi yang berguna dalam menentukan wilayah prioritas penanganan di Kabupaten Karawang.



Copyright © [JITET](http://www.jitet.org) (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

Abstract. *Pneumonia* is a serious illness that requires special attention, particularly in identifying areas with a high number of cases. Karawang Regency was chosen as the research location because it has a significant number of *pneumonia* cases, suspected to be influenced by environmental factors such as the presence of industrial zones. This study aims to cluster regions based on the number of *pneumonia* sufferers using the *K-means* and *K-medoids* algorithms. The data used are secondary data from the Karawang District Health Office for the period 2019–2023. The analysis process follows the stages of *Knowledge Discovery in Databases* (KDD), including *Data Selection*, *Pre-processing*, *Transformation*, *Data Mining*, *Evaluation*, and *Knowledge*. The data normalization stage uses *StandardScaler*, and dimensionality reduction is performed using *Principal Component Analysis* (PCA). The evaluation results show that *K-means* produces better clustering performance, with a *Silhouette Score* of 0.8066 and a *Davies-Bouldin Index* of 0.1231. Conversely, *K-medoids* shows less optimal results, with a *Silhouette Score* of 0.5847 and a *Davies-Bouldin Index* of 1.1531. The clustering results form two regional clusters based on the level of *pneumonia* case distribution, namely low and high, which are useful in determining priority areas for intervention in Karawang Regency.

1. PENDAHULUAN

Pneumonia merupakan salah satu penyakit infeksi saluran pernapasan bawah yang berisiko tinggi menyebabkan kematian. Meskipun kesadaran akan pentingnya kesehatan terus meningkat, prevalensi *pneumonia* di Indonesia menunjukkan tren peningkatan signifikan, dengan 1.278 kasus dan 188 kematian tercatat pada tahun 2024 (CNN Indonesia, 2024). Kondisi ini menegaskan pentingnya identifikasi pola penyebaran penyakit, khususnya di wilayah dengan tingkat kasus tinggi seperti Kabupaten Karawang.

Karawang dipilih sebagai lokasi penelitian karena tingginya kasus *pneumonia* yang dipengaruhi oleh faktor lingkungan, seperti keberadaan kawasan industri besar, tingginya mobilitas penduduk, paparan polusi udara, serta kondisi iklim yang lembap dan kepadatan wilayah. Faktor-faktor tersebut meningkatkan risiko penyebaran penyakit pernapasan, termasuk *pneumonia*, yang menjadi tantangan serius bagi layanan kesehatan di daerah tersebut.

Untuk memahami pola sebaran penderita pneumonia secara lebih akurat, pendekatan *Data Mining* digunakan melalui metode *clustering* untuk mengelompokkan pasien atau wilayah berdasarkan karakteristik tertentu. Teknik ini mampu memproses data berskala besar dengan presisi tinggi dan tanpa memerlukan label awal, sehingga efektif dalam menemukan pola tersembunyi dalam data kesehatan. Salah satu manfaat penting dari *clustering* adalah kemampuannya dalam mendukung segmentasi pasien dan wilayah risiko untuk intervensi yang lebih efisien.

Penelitian ini membandingkan dua algoritma *clustering* populer, yaitu *K-means* dan *K-medoids*. *K-means* dikenal lebih cepat dan efisien untuk data dalam jumlah besar, namun sensitif terhadap *outlier* karena menggunakan nilai rata-rata (*centroid*) sebagai pusat klaster. Sebaliknya, *K-medoids* menggunakan medoid (data aktual dalam klaster) sebagai pusat, yang membuatnya lebih tahan terhadap *outlier* meskipun membutuhkan waktu komputasi yang lebih tinggi [1]. Oleh karena itu, membandingkan keduanya penting

untuk menentukan pendekatan paling tepat dalam pengelompokan data penderita *pneumonia*.

Selain membantu dalam pemetaan wilayah risiko, pendekatan *clustering* juga mendukung efisiensi pengelolaan sumber daya kesehatan. Dengan mengidentifikasi wilayah-wilayah yang termasuk dalam kategori risiko tinggi, pemerintah daerah dan tenaga medis dapat lebih tepat sasaran dalam melakukan intervensi, seperti peningkatan layanan puskesmas, penyediaan fasilitas pernapasan, atau kampanye kesehatan masyarakat yang lebih intensif. Hal ini penting mengingat keterbatasan sumber daya dan tingginya tuntutan layanan kesehatan di daerah padat penduduk seperti Karawang.

Lebih lanjut, penggunaan algoritma *data mining* seperti *K-means* dan *K-medoids* tidak hanya mendukung pengambilan keputusan berbasis data, tetapi juga dapat dikembangkan untuk sistem prediksi dan pemantauan penyakit di masa depan. Dengan pemanfaatan data sekunder dari Dinas Kesehatan Kabupaten Karawang periode 2019 – 2023, penelitian ini tidak hanya memberikan gambaran kondisi *pneumonia* saat ini, tetapi juga membuka peluang pengembangan sistem pendukung keputusan yang lebih adaptif terhadap perubahan pola penyakit akibat faktor lingkungan dan sosial.

Oleh karena itu, penelitian ini dilakukan untuk menganalisis dan membandingkan algoritma *K-means* dan *K-medoids* dalam mengelompokkan penderita *pneumonia* di Kabupaten Karawang, guna menghasilkan segmentasi data yang akurat sebagai dasar dalam menyusun strategi intervensi kesehatan yang lebih efektif, efisien, dan berbasis data.

2. TINJAUAN PUSTAKA

2.1. *Pneumonia*

Pneumonia adalah penyakit infeksi yang menyerang saluran pernapasan bagian bawah, seringkali menjadi penyebab kematian di negara-negara berkembang. Penyakit ini ditandai dengan gejala seperti batuk serius dan kesulitan bernapas. Infeksi ini terutama berbahaya bagi individu dengan sistem kekebalan tubuh yang lemah atau memiliki

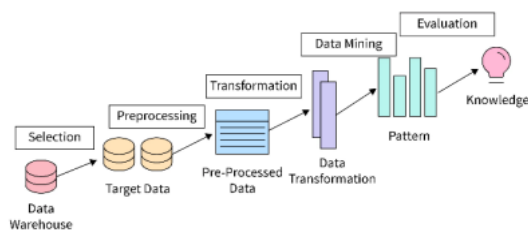
riwayat penyakit lain yang memperparah kondisi pernapasan [2].

2.2. Data Mining

Data mining merupakan alat yang memungkinkan pengguna untuk mengakses dan menganalisis data dalam jumlah besar secara cepat dan efisien. Secara lebih spesifik, *data mining* adalah teknik yang menggunakan analisis statistik untuk menggali informasi berharga dari kumpulan data yang luas. Proses ini bertujuan untuk mengekstraksi pola atau pengetahuan tersembunyi yang sebelumnya tidak diketahui, tetapi memiliki nilai guna dalam pengambilan keputusan [3].

2.3. Knowledge Discovery in Databases

Metode *Knowledge Discovery in Databases* (KDD) merupakan suatu proses yang terdiri dari serangkaian tahapan dalam mengekstraksi informasi yang bermakna dari kumpulan data yang besar. KDD mencakup berbagai langkah mulai dari data selection, preprocessing, transformation, data mining, hingga evaluasi hasil untuk memperoleh pola atau pengetahuan yang bermanfaat [4]. Ilustrasi tahapan KDD dapat dilihat pada **Gambar 2.1**



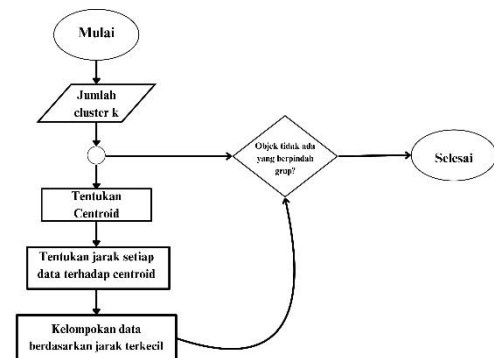
Gambar 2.1 Tahapan Knowledge Discovery in databases

2.4. Clustering

Clustering merupakan salah satu teknik dalam *data mining* yang digunakan untuk mengelompokkan sekumpulan objek berdasarkan tingkat kemiripan tertentu tanpa memerlukan label atau informasi kelas sebelumnya. Teknik ini memungkinkan objek-objek dengan karakteristik serupa dikelompokkan ke dalam satu grup, sementara objek yang memiliki perbedaan signifikan dimasukkan ke dalam kelompok lain. Hasil dari proses *clustering* menunjukkan bahwa objek dalam satu kelompok memiliki tingkat kesamaan yang lebih tinggi dibandingkan dengan objek di kelompok lainnya [5].

2.5. Algoritma K-Means

Algoritma *K-means* adalah salah satu metode *clustering* dalam *data mining* yang digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristiknya. Algoritma ini bekerja dengan cara membagi data ke dalam k kelompok atau kluster, di mana setiap data akan dimasukkan ke dalam kluster dengan pusat (*centroid*) terdekat [6].



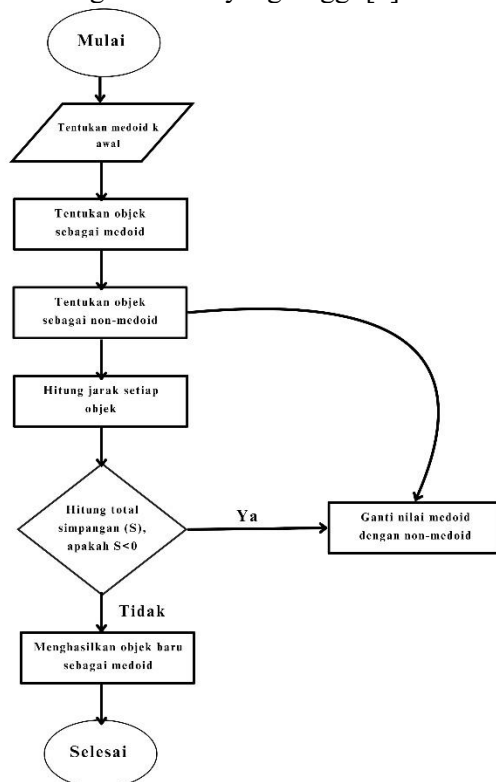
Gambar 2.2 Flowchart Algoritma K-Means

Gambar 2.2 menggambarkan alur algoritma *K-Means*, yang dimulai dengan menentukan jumlah kluster (k) dan menetapkan *centroid* awal secara acak. Selanjutnya, algoritma menghitung jarak setiap data ke masing-masing *centroid*, lalu mengelompokkan data ke kluster dengan jarak terdekat. Setelah itu, *centroid* diperbarui berdasarkan rata-rata posisi data dalam kluster. Proses ini diulang hingga tidak ada lagi data yang berpindah kluster, menandakan bahwa algoritma telah mencapai kondisi konvergen dan pengelompokan selesai.

2.6. Algoritma K-Medoids

K-medoids adalah metode pengelompokan berbasis partisi yang membagi sekumpulan objek ke dalam beberapa *cluster*. Dalam pendekatannya, *K-medoids* menggunakan medoid sebagai pusat *cluster*, yaitu objek yang paling merepresentasikan kelompok tersebut. Objek yang memiliki kedekatan dengan medoid akan dikelompokkan bersama dalam satu *cluster*. Keunggulan utama *K-medoids* adalah kemampuannya dalam mengatasi kelemahan *K-means* yang rentan terhadap *outlier*. Hasil pengelompokannya juga tetap stabil meskipun urutan data diacak [7]. Algoritma ini bekerja dengan prinsip meminimalkan ketidaksamaan antar objek, sehingga titik acuan yang

digunakan dalam metode ini lebih sesuai untuk data dengan variasi yang tinggi [8].



Gambar 2.3 Flowchart Algoritma K-Medoids

Gambar 2.3 menampilkan *flowchart* algoritma K-Medoids, yang diawali dengan menentukan medoid awal secara acak. Setelah itu, objek dipilih sebagai medoid dan non-medoid, lalu dihitung jarak antara masing-masing objek. Algoritma kemudian menghitung total simpangan (S) untuk mengevaluasi efisiensi medoid. Jika ditemukan total simpangan yang lebih kecil ($S < S_0$), maka medoid diganti dengan objek non-medoid yang menghasilkan simpangan tersebut. Proses ini diulang hingga tidak ada perubahan medoid yang memberikan hasil lebih baik, dan akhirnya dihasilkan medoid akhir sebagai representasi kluster.

2.7. Davies Bouldin Index (DBI)

Davies Bouldin Indeks merupakan metrik yang digunakan dalam analisis data untuk mengevaluasi kualitas hasil *clustering*. Metrik ini mengukur sejauh mana kluster yang terbentuk dapat memisahkan kelompok data yang berbeda serta seberapa dekat data dalam satu kluster terhadap pusat klasternya. Semakin

rendah nilai *Davies-Bouldin Index*, semakin baik kualitas *clustering* yang dihasilkan [9].

2.8. Silhouette Coefficient

Silhouette Coefficient adalah metode yang digunakan untuk mengukur dan menganalisis kualitas pengelompokan data. Metode ini menggabungkan dua perhitungan utama, yaitu *cohesion* dan *separation*. *Cohesion* berfungsi untuk mengukur tingkat kedekatan atau korelasi suatu objek dengan objek lain dalam *cluster* yang sama. Sementara itu, *separation* digunakan untuk menghitung sejauh mana perbedaan atau jarak antar-*cluster*, sehingga dapat menentukan seberapa baik data dikelompokkan. Sedangkan menurut [10]. Berikut nilai interpretasi *Silhouette Coefficient* pada Tabel 2. 1.

Tabel 2. 1 Rentang nilai *Silhouette Coefficient*

Rentang Nilai	Interpretasi
0,71 – 1,00	Struktur kuat
0,51 – 0,70	Struktur baik
0,26 – 0,50	Struktur lemah
$\leq 0,25$	Tidak terstruktur

Silhouette Coefficient merupakan metrik yang memiliki rentang nilai antara -1 hingga 1, dan digunakan untuk mengukur seberapa baik suatu objek berada dalam kluster yang sesuai. Nilai ini dihitung berdasarkan rata-rata *silhouette* dari setiap anggota kluster. Semakin mendekati angka 1, maka semakin baik kualitas pengelompokan data tersebut. Sebaliknya, jika nilainya mendekati -1, maka menunjukkan bahwa objek cenderung salah kluster dan kualitas pengelompokannya semakin buruk.

2.9. Metode Elbow

Metode *Elbow* adalah teknik yang digunakan untuk menentukan jumlah cluster (c) yang optimal dengan menghitung nilai *Sum of Square Error* (SSE) untuk setiap *cluster*. Semakin besar perbedaan nilai SSE antara satu *cluster* dengan *cluster* berikutnya hingga membentuk sudut siku, maka semakin baik jumlah cluster yang dipilih. Metode ini banyak digunakan untuk menentukan jumlah *cluster* yang optimal dengan cara menguji nilai SSE pada berbagai jumlah *cluster* dan memilih titik dengan selisih terbesar atau sudut paling tajam pada grafik *Elbow* [11].

2.10. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) merupakan salah satu metode paling umum yang digunakan dalam proses reduksi dimensi. Teknik ini termasuk dalam metode statistik standar yang berfungsi untuk menyederhanakan struktur data dengan mengurangi jumlah variabel tanpa kehilangan informasi penting. PCA bekerja melalui transformasi linear dan sering dimanfaatkan dalam analisis multivariat untuk mengekstraksi informasi utama dari data berukuran besar serta memahami struktur keterkaitan antar variabel [12].

2.11. StandardScaler

StandardScaler merupakan salah satu metode praproses data yang sering digunakan dalam analisis data dan *machine learning* untuk mengubah fitur numerik agar memiliki nilai rata-rata nol dan standar deviasi satu. Tujuan dari transformasi ini adalah untuk menyamakan skala antar fitur, sehingga algoritma pembelajaran mesin dapat berfungsi secara lebih optimal [13].

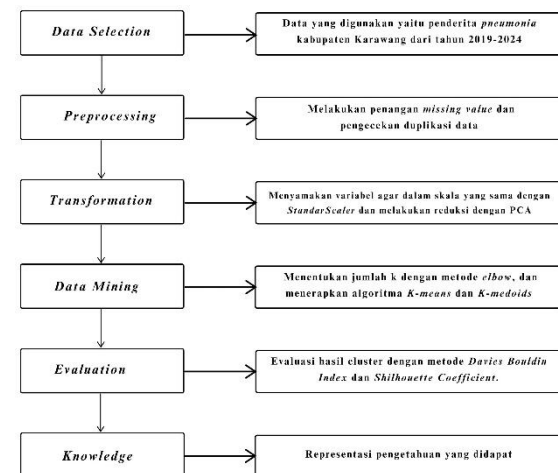
2.12. Google Colab

Google Colab adalah lingkungan pengembangan berbasis *cloud* yang memungkinkan pengguna menjalankan kode *Python* tanpa perlu instalasi perangkat lunak tambahan. Pemrosesan dilakukan di server Google dengan perangkat keras berperforma tinggi, sehingga pengguna dapat memanfaatkan sumber daya komputasi yang lebih besar [14]. Selain itu, Google Colab terintegrasi dengan Google Drive, memungkinkan penyimpanan dan akses data yang mudah. Dengan kemudahan penggunaan dan dukungan komputasi yang kuat, Google Colab menjadi pilihan utama bagi banyak praktisi data science dan analisis data [15].

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan *Knowledge Discovery in Databases* (KDD) untuk menganalisis dan mengelompokkan wilayah berdasarkan jumlah kasus pneumonia di Kabupaten Karawang pada periode 2019 hingga 2024. Proses KDD dilakukan melalui enam tahapan utama, yaitu *Data Selection*, *Data Preprocessing*, *Data Transformation*, *Data Mining*, *Evaluation*, dan *Knowledge*. Setiap tahapan dirancang untuk memastikan

kualitas data yang optimal, penerapan algoritma *clustering* yang tepat, serta interpretasi hasil yang relevan terhadap tujuan penelitian.



Gambar 3. 1 Alur Penelitian

3.1. Data Selection

Penelitian ini menggunakan data jumlah penderita *pneumonia* di Kabupaten Karawang tahun 2019 – 2023 yang diperoleh dari Dinas Kesehatan Kabupaten Karawang. Pada tahap *data selection*, dilakukan pemilahan data yang relevan dengan delapan atribut utama, seperti nama kecamatan, tahun, jumlah kasus berdasarkan usia dan jenis kelamin, serta total kasus dan kematian. Data yang telah diseleksi kemudian diolah menggunakan metode *clustering* untuk mengidentifikasi pola distribusi kasus *pneumonia*.

3.2. Pre-processing

Pada tahap ini, dilakukan proses pembersihan data dengan tujuan menghilangkan informasi yang tidak relevan atau tidak diperlukan dalam analisis. Beberapa atribut yang akan dihapus dalam proses ini antara lain nomor dan tahun. Selanjutnya, pada tahap transformasi data, dilakukan perubahan format data agar lebih sederhana dan mudah dianalisis. Proses ini mencakup standarisasi data untuk memastikan konsistensi dalam analisis. Selain itu, dilakukan data *scaling* guna mencegah ketidakakuratan dalam analisis akibat perbedaan rentang nilai yang terlalu jauh.

3.3. Transformation

Transformasi data dilakukan untuk mengatasi ketidaksempurnaan dalam data penderita pneumonia, seperti adanya data yang

hilang atau perbedaan skala antar variabel. Salah satu langkah dalam proses transformasi ini adalah normalisasi data, yang bertujuan untuk menyamakan skala sehingga setiap nilai dalam dataset memiliki tingkat yang seragam. Proses ini bertujuan agar analisis pengelompokan (*clustering*) dapat dilakukan dengan lebih akurat. Dalam penelitian ini, metode normalisasi yang diterapkan adalah *StandardScaler*, guna memastikan bahwa seluruh data berada dalam rentang yang sama tanpa mengubah pola distribusinya. Setelah proses normalisasi, dilakukan reduksi dimensi menggunakan metode *Principal Component Analysis* (PCA). Teknik ini digunakan untuk menyederhanakan kompleksitas data dengan mengubah sejumlah variabel yang saling berkorelasi menjadi sejumlah kecil komponen utama yang tetap mempertahankan informasi penting. Dengan penerapan PCA, proses *clustering* menjadi lebih efisien dan fokus terhadap pola dominan dalam data.

3.4. Data Mining

Data yang telah melalui tahap transformasi kemudian diolah menggunakan algoritma *K-means* dan *K-medoids* untuk mengelompokkan data berdasarkan pola distribusi kasus *pneumonia*. Penentuan jumlah kluster dilakukan melalui metode *elbow* untuk memperoleh jumlah kluster yang optimal. Hasil pengelompokkan kemudian dianalisis untuk mengidentifikasi pola persebaran kasus *pneumonia* di Kabupaten Karawang, sehingga dapat memberikan gambaran mengenai kelompok yang terbentuk berdasarkan karakteristik data.

3.5. Evaluation

Setelah dilakukan proses *data mining*, diperoleh hasil pengelompokan (*clustering*) dari dua algoritma, yaitu *K-means* dan *K-medoids*. Untuk menilai kualitas *cluster* yang terbentuk, evaluasi dilakukan dengan menggunakan metode *Davies Bouldin Index* (DBI) dan *Silhouette Coefficient* (SC).

3.6. Knowledge

Pada tahap akhir, data yang telah diproses akan menjalani evaluasi ulang untuk memastikan kualitas hasil *clustering*. Hasil yang diperoleh selanjutnya akan dianalisis dan

dirangkum agar dapat disajikan dengan jelas dan mudah dipahami. Penyajian hasil ini dilakukan agar informasi yang dihasilkan dapat dimanfaatkan secara optimal, khususnya dalam mengidentifikasi pola distribusi penderita *pneumonia* di kabupaten Karawang.

4. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan teknik *data mining* dengan menerapkan algoritma *K-means* dan *K-medoids* untuk melakukan proses *clustering*, guna menghasilkan pengelompokan wilayah berdasarkan jumlah penderita *pneumonia* di Kabupaten Karawang. Hasil pengelompokan kemudian dievaluasi menggunakan metrik *Silhouette Coefficient* dan *Davies-Bouldin Index* (DBI).

4.1. Data Selection

Pada tahap *data selection*, Penelitian ini menggunakan data jumlah penderita *pneumonia* yang diperoleh dari Dinas Kesehatan Kabupaten Karawang, mencakup seluruh kecamatan pada periode 2019 – 2023. Data bersifat kuantitatif dan merepresentasikan kondisi penyebaran kasus *pneumonia* selama lima tahun terakhir. Informasi ini menjadi dasar dalam proses *clustering* wilayah berdasarkan tingkat kerawanan *pneumonia* menggunakan teknik *data mining*. data tersebut dapat dilihat pada **Gambar 4. 1**.

No	Nama Kecamatan	Tahun	Pneumonia (L < 1 Tahun)					Pneumonia (1 < 5 Tahun)					Pneumonia (5 < 10 Tahun)					Pemeriksaan	Total Kasus	Rasio (L)	Rasio (P)	Rasio (T)	Total Rasio
			(L < 1 Tahun)	(1 < 5 Tahun)	(5 < 10 Tahun)	(10 < 15 Tahun)	(15 < 20 Tahun)	(20 < 25 Tahun)	(25 < 30 Tahun)	(30 < 35 Tahun)	(35 < 40 Tahun)	(40 < 45 Tahun)	(45 < 50 Tahun)	(50 < 55 Tahun)	(55 < 60 Tahun)	(60 < 65 Tahun)	(65 < 70 Tahun)						
1	ADIRAJA	2019-2023	78	63	240	172	3	4	12	10	333	249	582	33	47	80							
2	ANGGADITA	2019-2023	26	25	110	110	1	1	1	138	137	275	68	66	134								
3	BALONGGAGAH	2019-2023	17	19	127	128	0	0	0	144	147	291	69	52	121								
4	BATULARA	2019-2023	119	96	162	149	27	13	27	24	335	292	617	147	119	266							
5	BAKUR LOR	2019-2023	52	47	61	63	21	20	25	26	159	176	305	0	0	0							

Gambar 4. 1 Dataset *Pneumonia* Karawang 2019 -2023

Deskripsi dataset yang digunakan meliputi nama atribut serta penjelasan singkat mengenai masing-masing atribut. Rincian lebih lanjut mengenai struktur dataset kasus *pneumonia* ditampilkan pada **Tabel 4. 1**.

Tabel 4. 1 Deskripsi Dataset *Pneumonia*

Atribut	Keterangan
Nama Kecamatan	Nama Kecamatan yang ada di Kabupaten Karawang
<i>Pneumonia</i> (L < 1 Tahun)	Jumlah kasus <i>pneumonia</i> pada laki-laki usia di bawah 1 tahun

Atribut	Keterangan
<i>Pneumonia</i> (P < 1 Tahun)	Jumlah kasus <i>pneumonia</i> pada perempuan usia di bawah 1 tahun
<i>Pneumonia</i> (L 1 -< 5 Tahun)	Jumlah kasus <i>pneumonia</i> pada laki-laki usia 1 hingga kurang dari 5 tahun
<i>Pneumonia</i> (P 1 -< 5 Tahun)	Jumlah kasus <i>pneumonia</i> pada perempuan usia 1 hingga kurang dari 5 tahun
PB (L < 1 Tahun)	Jumlah kasus <i>pneumonia</i> berat pada laki-laki usia di bawah 1 tahun
PB (P < 1 Tahun)	Jumlah kasus <i>pneumonia</i> berat pada perempuan usia di bawah 1 tahun
PB (L 1 -< 5 Tahun)	Jumlah kasus <i>pneumonia</i> berat pada laki-laki usia 1 hingga kurang dari 5 tahun
PB (P 1 -< 5 Tahun)	Jumlah kasus <i>pneumonia</i> berat pada perempuan usia 1 hingga kurang dari 5 tahun
Laki-laki	Total kasus <i>pneumonia</i> pada pasien laki-laki (seluruh rentang usia)
Perempuan	Total kasus <i>pneumonia</i> pada pasien perempuan (seluruh rentang usia)
Total Kasus	Total kasus <i>pneumonia</i> pada pasien perempuan (seluruh rentang usia)
Meninggal (L)	Jumlah pasien laki-laki yang meninggal akibat <i>pneumonia</i>
Meninggal (P)	Jumlah pasien perempuan yang meninggal akibat <i>pneumonia</i>
Total Meninggal	Jumlah total kematian akibat <i>pneumonia</i>

4.2. Preprocessing

Setelah melalui tahapan *data selection* atau seleksi data, langkah selanjutnya adalah data *preprocessing* yang bertujuan untuk menangani *missing value* (data yang tidak memiliki nilai) serta mendeteksi adanya duplikasi data. Berdasarkan hasil pengecekan terhadap dataset *pneumonia* Karawang, tidak ditemukan adanya *missing value* pada atribut-atribut yang telah

dipilih. Proses pengecekan missing value dapat dilihat pada **Gambar 4. 2**.

```
print("Missing Values:")
print(df.isnull().sum())
```

```
Missing Values:
Nama Kecamatan          0
Pneumonia (L < 1 Tahun)  0
Pneumonia (P < 1 Tahun)  0
Pneumonia (L 1 -< 5 Tahun)  0
Pneumonia (P 1 -< 5 Tahun)  0
PB (L < 1 Tahun)         0
PB (P < 1 Tahun)         0
PB (L 1 -< 5 Tahun)      0
PB (P 1 -< 5 Tahun)      0
Laki-Laki                0
Perempuan                0
Total Kasus              0
Meninggal (L)            0
Meninggal (P)            0
Total Meninggal          0
dtype: int64
```

Gambar 4. 2 Pengecekan *Missing value*

Tahap selanjutnya adalah melakukan pengecekan duplikasi data pada dataset kasus *pneumonia* di Kabupaten Karawang. Berdasarkan hasil pengecekan, tidak ditemukan adanya data duplikat pada masing-masing dataset per tahun. Proses pengecekan duplikasi data dapat dilihat pada **Gambar 4. 3**.

```
# Cek duplikasi data

print("\nDuplicated Values:")
print(df.duplicated().sum())
```

```
Duplicated Values:
0
```

Gambar 4. 3 Pengecekan data duplikat

4.3. Transformation

Tahap selanjutnya adalah transformasi data untuk mempersiapkan proses *clustering*. Transformasi diperlukan agar data berada dalam bentuk dan skala yang sesuai, sehingga algoritma dapat bekerja secara optimal. Proses normalisasi data dilakukan dengan menggunakan metode standarisasi (*standardization*) melalui pendekatan *StandardScaler* dari pustaka *scikit-learn*. Metode ini bertujuan untuk menyamakan skala antar fitur numerik dalam dataset, sehingga setiap fitur memiliki rata-rata (*mean*) sebesar 0 dan standar deviasi sebesar 1. Berikut program

normalisasi dengan *StandardScaler* disajikan pada **Gambar 4. 4**.

```
from sklearn.preprocessing import StandardScaler

# Pastikan df sudah didefinisikan sebelumnya dan kolom numeriknya teridentifikasi
numerical_cols = df.select_dtypes(include=['int64', 'float64']).columns

# Standarisasi hanya kolom numerik
scaler = StandardScaler()
df_scaled_values = scaler.fit_transform(df[numerical_cols])

# Simpan hasil standarisasi ke DataFrame baru dengan nama kolom yang sesuai
df_scaled = pd.DataFrame(df_scaled_values, columns=numerical_cols, index=df.index)
```

Gambar 4. 4 Sintaks normalisasi untuk StandardScaler

Sementara itu, hasil dari proses normalisasi terhadap seluruh atribut numerik dapat dilihat pada **Gambar 4.5**, yang menunjukkan bahwa semua nilai telah berhasil diubah ke skala yang seragam.

	Prevalence ($P = 1$)	Prevalence ($P = 3$)	Prevalence ($P = 5$)	Prevalence ($P = 3 + 5$)	FB (L = 1) ($P = 1$)	FB (L = 3) ($P = 3$)	FB (L = 5) ($P = 5$)	FB (L = 3 + 5) ($P = 3 + 5$)	Logit	Parapetous	Total Logit	Herrington (%)	Herrington (%)	Total (%)
0	0.26202	0.23385	0.87291	0.42762	0.15720	0.15189	0.03006	0.42225	0.18106	0.23161	0.33502	0.23187	0.96860	0.23187
1	0.30545	0.28135	0.87545	0.40400	0.04482	0.43879	0.49986	0.47838	0.24640	0.17913	0.20210	0.16488	0.06486	0.06486
2	0.36338	0.33658	0.89418	0.37165	0.33035	0.53424	0.62026	0.53010	0.18106	0.15508	0.35876	0.05876	0.03563	0.03563
3	0.28328	0.44852	0.88125	0.26628	0.17301	0.03913	0.21780	0.18106	0.54504	0.05407	0.57200	0.31845	0.22267	0.37880
4	0.13147	0.12273	0.32644	0.19638	0.10741	0.10747	0.19624	0.13894	0.50850	0.04509	0.10400	0.39054	0.13905	0.13905

Gambar 4. 5 Hasil StandarScaler pada dataset

Setelah dilakukan normalisasi, tahap selanjutnya adalah reduksi dimensi menggunakan metode *Principal Component Analysis* (PCA). PCA digunakan untuk menyederhanakan jumlah atribut dalam dataset tanpa kehilangan informasi penting, dengan cara mengubah data berdimensi tinggi menjadi tiga komponen utama (PCA1, PCA2 dan PCA3) yang mewakili variasi terbesar dalam data. Pada **Gambar 4. 6** ditampilkan sintaks program yang digunakan untuk menerapkan PCA dengan tiga komponen.

```

# 1. Langsung gunakan df_scaled karena sudah berisi kolom numerik yang distandarisasi
data_numeric = df_scaled # tidak perlu filter ulang kolom numerik

# 2. Inisialisasi dan jalankan PCA dengan 3 komponen
pca = PCA(n_components=3)
pca_result = pca.fit_transform(data_numeric)

# 3. Simpan hasil PCA ke DataFrame baru
df_pca = pd.DataFrame(pca_result, columns=['PCA1', 'PCA2', 'PCA3'])

# 5. Tampilkan hasil PCA
df_pca.head()

```

Gambar 4. 6 Sintaks PCA pada Hasil *StandardScaler*

Sementara itu, hasil transformasi data setelah direduksi menjadi tiga dimensi ditampilkan pada **Gambar 4. 7**, yang menunjukkan nilai-nilai dari masing-masing data pada komponen PCA1, PCA 2 dan PCA3. Hasil ini selanjutnya digunakan dalam proses *clustering* agar analisis menjadi lebih efisien dan akurat.

	PCA1	PCA2	PCA3
0	0.816474	-0.374936	-0.055052
1	-0.858064	-0.214477	-0.604066
2	-0.870288	-0.136940	-0.839667
3	2.125395	0.949729	2.217627
4	1.039751	0.001667	3.109983

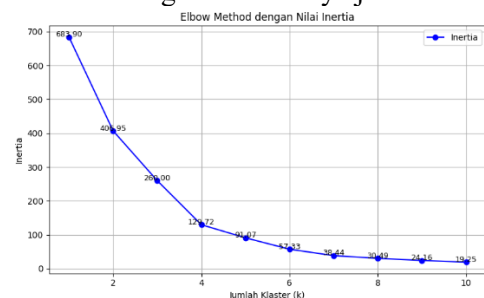
Gambar 4. 7 Hasil PCA pada data *StandardScaler*

4.4. Data Mining

Proses data mining dalam penelitian ini merupakan tahap inti yang bertujuan untuk mengelompokkan data berdasarkan pola tertentu yang tersembunyi. Data yang telah melalui proses transformasi sebelumnya kemudian dianalisis menggunakan dua algoritma *clustering*, yaitu *K-Means* dan *K-Medoids*. Kedua algoritma ini dipilih karena mampu mengelompokkan data tanpa memerlukan label atau kategori awal, sehingga sangat sesuai untuk tujuan eksplorasi pola distribusi kasus *pneumonia*. Dengan menggunakan pendekatan ini, diharapkan dapat terbentuk kluster-kluster wilayah yang memiliki karakteristik kasus *pneumonia* yang serupa, yang nantinya dapat digunakan sebagai dasar dalam penentuan prioritas penanganan kesehatan di Kabupaten Karawang.

4.4.1. Algoritma K-Means

Langkah awal dalam penerapan algoritma ini dimulai dengan menentukan nilai jumlah kluster (k) yang akan digunakan dalam proses pengelompokan. Untuk memperoleh nilai k yang optimal, dilakukan evaluasi menggunakan metode Elbow. Nilai *inertia* merepresentasikan total jarak kuadrat antar data terhadap pusat klasternya, di mana semakin kecil nilai *inertia*, maka semakin baik data terkelompokkan. Namun, penurunan nilai *inertia* akan semakin melambat seiring bertambahnya jumlah kluster.



Gambar 4. 8 Grafik Metode Elbow

Gambar 4. 8 merupakan visualisasi dari nilai-nilai *inertia* dalam bentuk grafik Elbow. Dari grafik tersebut, terlihat adanya titik siku (*elbow*) yang terjadi pada $K=2$, di mana setelah titik tersebut penurunan nilai *inertia* tidak lagi signifikan. Oleh karena itu, jumlah kluster optimal yang digunakan dalam penelitian ini ditentukan sebanyak 2 kluster.

Pada tahap selanjutnya, dilakukan penerapan algoritma *K-means* terhadap data yang telah melalui proses normalisasi dan reduksi dimensi menggunakan PCA. Proses *clustering* dilakukan dengan menetapkan jumlah kluster sebanyak dua berdasarkan hasil dari metode *Elbow*. Label kluster yang dihasilkan kemudian ditambahkan ke dalam dataset untuk keperluan analisis dan visualisasi lebih lanjut. Sintaks program yang digunakan ditampilkan pada **Gambar 4. 9** dan hasilnya pada **Gambar 4. 10**.

```
from sklearn.cluster import KMeans

# 1. Siapkan data PCA 3 komponen
X = df_pca[['PCA1', 'PCA2', 'PCA3']]

# 2. Tentukan jumlah kluster
k_optimal = 2

# 3. Jalankan K-Means
kmeans = KMeans(n_clusters=k_optimal, n_init=10, random_state=42)
kmeans.fit(X)

# 4. Tambahkan label kluster ke DataFrame PCA
df_pca['Cluster_Kmeans'] = kmeans.labels_

# 5. (Optional) Gabungkan hasil kluster ke DataFrame asli
df['Cluster_Kmeans'] = kmeans.labels_

# 6. Cek hasil
df_pca.head()
```

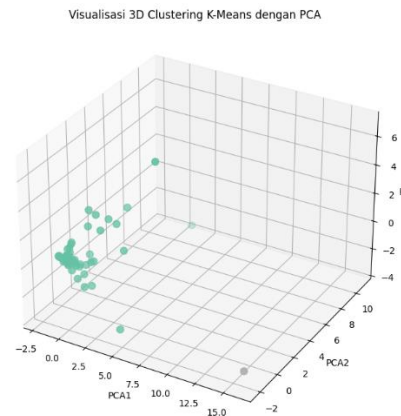
Gambar 4. 9 Sintaks Algoritma *K-Means*

	PCA1	PCA2	PCA3	Cluster_Kmeans
0	0.816474	-0.374936	-0.055052	0
1	-0.858064	-0.214477	-0.604066	0
2	-0.870288	-0.136940	-0.839667	0
3	2.125395	0.949729	2.217627	0
4	1.039751	0.001667	3.109983	0

Gambar 4. 10 Hasil *Clustering K-Means*

Selanjutnya dilakukan visualisasi hasil *clustering* dalam bentuk tiga dimensi berdasarkan dua komponen utama PCA. Masing-masing titik mewakili satu data dan diberi warna sesuai kluster yang terbentuk. Visualisasi ini memudahkan dalam memahami sebaran dan pemisahan antar kluster yang dihasilkan oleh algoritma *K-means*. Hasil

visualisasi tersebut ditunjukkan pada Gambar 4. 11.



Gambar 4. 11 Visualisasi 3D *Clustering K-Means*

4.4.1. Algoritma *K-Medoids*

Proses *clustering* selanjutnya dilakukan menggunakan algoritma *K-Medoids*. Algoritma ini serupa dengan *K-means* namun memiliki perbedaan dalam menentukan pusat kluster, yaitu menggunakan medoid sebagai pusatnya. Pada sintaks yang ditampilkan, jumlah kluster yang digunakan adalah dua, sesuai dengan hasil dari metode *Elbow*. Proses ini juga menggunakan metrik jarak *Euclidean* dan inisialisasi *build* sebagai nilai *default*. Sintaks program untuk penerapan algoritma *K-medoids* ditunjukkan pada **Gambar 4. 12**, hasilnya pada **Gambar 4. 13** dan visualisasi hasil *clustering K-medoids* ditampilkan pada **Gambar 4. 14**.

```
from pyclustering.cluster.kmedoids import kmedoids
from pyclustering.utils import calculate_distance_matrix
from pyclustering.cluster import cluster_visualizer
import numpy as np
import random

# Ambil data dari PCA
X = df_pca[['PCA1', 'PCA2', 'PCA3']].values

# 1. Hitung matriks jarak
distance_matrix = calculate_distance_matrix(X)

# 2. Pilih indeks medoid awal secara acak
k = 2 # ubah sesuai hasil elbow
initial_medoids = random.sample(range(len(X)), k)

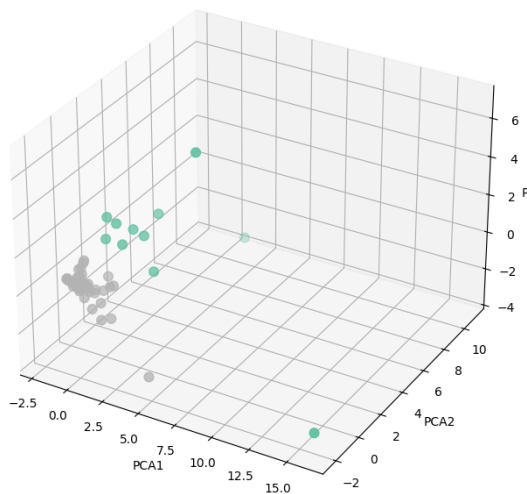
# 3. Buat dan jalankan algoritma K-Medoids
kmedoids_instance = kmedoids(distance_matrix, initial_medoids, data_type='distance_matrix')
kmedoids_instance.process()
```

Gambar 4. 12 Sintaks Algoritma *K-Medoids*

	PCA1	PCA2	PCA3	Cluster_Kmedoids
0	0.816474	-0.374936	-0.055052	1
1	-0.858064	-0.214477	-0.604066	1
2	-0.870288	-0.136940	-0.839667	1
3	2.125395	0.949729	2.217627	0
4	1.039751	0.001667	3.109983	0

Gambar 4. 13 Hasil *Clustering K-Medoids*

Visualisasi 3D Clustering K-Medoids dengan PCA

**Gambar 4. 14** Visualisasi 3D Clustering K-Medoids

4.5. Evaluation

Pada tahap evaluasi, digunakan dua metrik untuk menilai kualitas hasil *clustering*, yaitu *Silhouette Coefficient* dan *Davies-Bouldin Index* (DBI). *Silhouette Coefficient* mengukur seberapa mirip suatu objek dengan klasternya sendiri dibandingkan dengan klaster lainnya. Sementara *Davies-Bouldin Index* mengukur rata-rata kesamaan antar klaster, di mana nilai yang lebih rendah menandakan pemisahan klaster yang lebih baik. Evaluasi dilakukan terhadap hasil *clustering K-means* dan *K-medoids* menggunakan data hasil reduksi PCA. Nilai evaluasi metrik ditampilkan pada **Gambar 4. 15** (sintaks) dan **Gambar 4. 16** (hasil evaluasi), guna menilai kualitas hasil pengelompokan dari masing-masing algoritma.

```
from sklearn.metrics import silhouette_score, davies_bouldin_score

# Hitung Silhouette Score untuk K-Means
silhouette_kmeans = silhouette_score(df_pca[['PCA1', 'PCA2', 'PCA3']], df_pca['Cluster_Kmeans'])

# Hitung Davies-Bouldin Index untuk K-Means
dbi_kmeans = davies_bouldin_score(df_pca[['PCA1', 'PCA2', 'PCA3']], df_pca['Cluster_Kmeans'])

# Hitung Silhouette Score
silhouette_kmedoids = silhouette_score(df_pca[['PCA1', 'PCA2', 'PCA3']], df_pca['Cluster_Kmedoids'])

# Hitung Davies-Bouldin Index
dbi_kmedoids = davies_bouldin_score(df_pca[['PCA1', 'PCA2', 'PCA3']], df_pca['Cluster_Kmedoids'])
```

Gambar 4. 15 Sintaks evaluasi metrik

Evaluasi Klustering:

K-Means Silhouette Score: 0.8066
 K-Medoids Silhouette Score: 0.5847
 K-Means Davies-Bouldin Index: 0.1231
 K-Medoids Davies-Bouldin Index: 1.1531

Gambar 4. 16 Nilai evaluasi metrik

Hasil evaluasi klustering menggunakan dua metrik evaluasi, yaitu *Silhouette Score* dan *Davies-Bouldin Index*, terhadap algoritma *K-means* dan *K-medoids*. Berdasarkan nilai *Silhouette Score*, algoritma *K-means* menghasilkan skor sebesar 0.8066, sedangkan *K-medoids* memperoleh nilai 0.5847. Nilai *Silhouette Score* yang mendekati 1 menandakan bahwa objek-objek dalam satu klaster saling berdekatan dan berjauhan dari objek di klaster lain, sehingga *K-means* dapat disimpulkan memiliki kualitas klaster yang lebih baik dibandingkan *K-medoids* dalam hal ini.

Sementara itu, pada metrik *Davies-Bouldin Index*, yang semakin kecil nilainya menunjukkan klaster yang semakin optimal, *K-means* kembali menunjukkan performa yang lebih unggul dengan nilai sebesar 0.1231, dibandingkan *K-medoids* yang memiliki nilai sebesar 1.1531. Nilai indeks yang lebih rendah pada *K-means* mengindikasikan bahwa klaster yang terbentuk memiliki jarak yang lebih baik antar pusat klaster dan penyebaran dalam klaster yang lebih kecil. Secara keseluruhan, kedua metrik ini menunjukkan bahwa algoritma *K-means* memberikan hasil klasterisasi yang lebih baik dibandingkan dengan *K-medoids* pada data yang digunakan dalam penelitian ini.

4.6. Knowledge

Pada tahapan *Knowledge*, dilakukan interpretasi terhadap hasil klasterisasi guna memperoleh pengetahuan baru yang bermakna dari data yang telah diolah. Berdasarkan hasil evaluasi yang telah dilakukan pada tahap sebelumnya, algoritma *K-means* menunjukkan performa yang lebih baik dibandingkan *K-medoids*, ditinjau dari dua metrik evaluasi yaitu *Silhouette Score* dan *Davies-Bouldin Index*. Nilai *Silhouette Score* yang lebih tinggi dan nilai *Davies-Bouldin Index* yang lebih rendah pada *K-means* mengindikasikan bahwa pembentukan klaster lebih kompak dan terpisah dengan baik. Oleh karena itu, pada tahap ini, hasil klasterisasi yang digunakan untuk dianalisis lebih lanjut adalah hasil dari algoritma *K-means*.

Untuk mempermudah proses interpretasi, hasil pembagian klaster berdasarkan algoritma *K-means* disajikan secara rinci pada **Tabel 4. 2**.

Tabel 4. 2 Klasterisasi *Pneumonia* pada masing masing kecamatan

Nama Kecamatan	Klaster
Adiarsa, Anggadita, Balongsari, Batujaya, Bayur Lor, Ciampel, Cibuaya, Cicinde, Cikampek, Cikampek Utara, Cilamaya, Curug, Gempol, Jatisari, Jayakarta, Jomin, Kalangsari, Karawang, Krawangkulon, Kertamukti, Kota Baru, Kutamukti, Kutawaluya, Lemah Duhur, Lemah Abang, Loji, Majalaya, Medang Asem, Nagasari, Pacing, Pakisjaya, Pangkalan, Pasirukem, Pedes, Plawad, Purwasari, Rawamerta, Rngsdengklok, Sukatani, Sungai Buntu, Tanjungpura, Telagasari, Teluk Jambe, Tempuran, Tirtajaya, Tirtamulya, Tunggak Jati, Wadas, Wanakarta	Klaster 0
Klari	Klaster 1

Klaster 0 memiliki karakteristik data dengan nilai yang relatif rendah pada sebagian besar atribut, mencerminkan wilayah dengan tingkat kasus pneumonia yang lebih ringan. Sebaliknya, Klaster 1 menunjukkan karakteristik data dengan nilai yang cenderung tinggi, yang mengindikasikan wilayah dengan jumlah kasus pneumonia yang lebih besar serta tingkat keparahan yang lebih signifikan.

5. KESIMPULAN

a. Penelitian ini menunjukkan bahwa algoritma *K-means* dan *K-medoids* berhasil diterapkan dalam pengelompokan kasus *pneumonia* di Kabupaten Karawang melalui tahapan KDD, yang mencakup normalisasi data dan reduksi dimensi

menggunakan PCA. Hasil pengelompokan menghasilkan dua klaster wilayah berdasarkan kemiripan karakteristik kasus *pneumonia*.

- b. Berdasarkan evaluasi menggunakan *Silhouette Coefficient* dan *Davies Bouldin Index* (DBI), algoritma *K-means* menunjukkan performa yang lebih baik dibandingkan *K-medoids*, dengan skor *Silhouette* 0,8066 dan DBI 0,1231. Oleh karena itu, *K-means* dinilai lebih optimal dan direkomendasikan untuk pengelompokan wilayah berdasarkan tingkat kasus pneumonia di Karawang.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Ibu Betha Nurina Sari, M.Kom., dan Bapak Iqbal Maulana, S.Si., M.Sc., selaku dosen Program Studi Informatika Universitas Singaperbangsa Karawang, atas bimbingan dan masukan yang diberikan selama penyusunan penelitian ini. Dukungan yang diberikan sangat membantu dalam penyelesaian dan penyempurnaan artikel ini.

DAFTAR PUSTAKA

- [1] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan," *Jurnal Tekno Kompak*, vol. 15, no. 2, p. 25, 2021.
- [2] J. M. Moy, S. D. R. P. Santoso, and W. Paju, "Implementasi Fisioterapi Dada terhadap Masalah Bersihan Jalan Nafas Tidak Efektif pada Pasien Pneumonia," *Jurnal Keperawatan Sumba*, vol. 2, no. 2, pp. 58–69, 2024.
- [3] I. Ahmad, S. Samsugi, and Y. Irawan, "Implementasi Data Mining sebagai Pengolahan Data," *J. Teknoinfo*, vol. 16, no. 1, p. 46, 2022.
- [4] Y. B. Utomo, I. Kurniasari, and I. Yanuartanti, "Penerapan Knowledge Discovery In Database untuk Analisa Tingkat Kecelakaan Lalu Lintas," *JTIK (Jurnal Teknik Informatika Kaputama)*, vol. 7, no. 1, pp. 171–180, 2023.
- [5] R. D. Bkti, R. N. Zulfahmi, M. K. Daul, W. J. Pradnyaana, and E. Sutanta, "Sistem informasi berbasis website untuk pemetaan wilayah berdasarkan clustering kerentanan kriminalitas," *J. Informatika Teknologi dan Sains (Jinteks)*, vol. 6, no. 3, pp. 620–626, 2024.

- [6] E. Rahmah, "Penerapan algoritma K-Medoids clustering untuk menentukan strategi promosi pada data mahasiswa (studi kasus: STIKES Perintis Padang)," *J. Penerapan Algoritma K-Medoids Clustering*, vol. 5, no. 3, pp. 556–564, 2022.
- [7] F. Zahra, A. Khalif, dan B. N. Sari, "Pengelompokan tingkat kemiskinan di setiap provinsi di Indonesia menggunakan algoritma K-Medoids," *J. Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, 2024.
- [8] G. B. Kaligis, "Analisa perbandingan algoritma K-Means, K-Medoids, dan X-Means untuk pengelompokan kinerja pegawai (studi kasus: Sekretariat DPRD Provinsi Sulawesi Utara)," 2022.
- [9] I. T. Umagapi, B. Umaternate, H. Hazriani, and Y. Yuyun, "Uji kinerja K-Means clustering menggunakan Davies-Bouldin Index pada pengelompokan data prestasi siswa," *Prosiding Sisfotek*, vol. 7, no. 1, pp. 303–308, 2023.
- [10] R. D. Bkti, R. N. Zulfahmi, M. K. Daul, W. J. Pradnyaana, and E. Sutanta, "Sistem informasi berbasis website untuk pemetaan wilayah berdasarkan clustering kerentanan kriminalitas," *J. Inform. Tek. dan Sains (Jinteks)*, vol. 6, no. 3, pp. 620–626, 2024.
- [11] A. P. Riani, A. Voutama, and T. Ridwan, "Penerapan K-Means clustering dalam pengelompokan hasil belajar peserta didik dengan metode Elbow," *J. Teknol. Syst. Inform. dan Syst. Komput. TGD*, vol. 6, no. 1, pp. 164–172, 2023.
- [12] A. S. Ritonga and I. Muhandhis, "Teknik data mining untuk mengklasifikasikan data ulasan destinasi wisata menggunakan reduksi data principal component analysis (PCA)," *J. Ilm. Educic: Pendidik. dan Informatika*, vol. 7, no. 2, pp. 124–133, 2021.
- [13] M. F. R. Mahendra, S. Sumarno, and N. L. Azizah, "Implementasi Machine Learning untuk memprediksi cuaca menggunakan Support Vector Machine," *J. Ilm. Komputasi*, vol. 23, no. 1, pp. 45–50, 2024.
- [14] R. G. Guntara, "Pemanfaatan Google Colab untuk aplikasi pendeteksian masker wajah menggunakan algoritma deep learning YOLOv7," *J. Teknol. dan Sist. Inf. Bisnis*, vol. 5, no. 1, pp. 55–60, 2023.
- [15] R. Nazar, "Implementasi pemrograman Python menggunakan Google Colab," *J. Inform. dan Komput. (JIK)*, vol. 15, no. 1, pp. 50–56, 2024.