



Perbandingan Algoritma K-Means dan K-Medoids Untuk Pemetaan Daerah Penanganan Diare Pada Balita di Kabupaten Kuningan

Tri Septiar Syamfithriani, Nita Mirantika, Ragel Trisudarmo*

Program Studi Sistem Informasi, Fakultas Ilmu Komputer,
Universitas Kuningan, Kuningan, Indonesia

Naskah Diterima : 7 September 2022; Diterima Publikasi : 24 Januari 2023

DOI: 10.21456/vol12iss2pp132-139

Abstract

Diarrhea is an endemic disease that contributes to the high mortality rate in Indonesia, especially among children under five. The Kuningan District Health Office had difficulties in monitoring and supervising the spread of diarrheal diseases. This study aims to produce a mapping scheme of priority areas in handling the prevention and control of the spread of diarrheal disease in children under five in Kuningan Regency. The method used is the Data Mining Clustering method by comparing two algorithms, namely the K-Means algorithm and the K-Medoids algorithm. Determination of the optimum number of clusters using the Elbow and Silhouette Coefficient methods. With this method, the result is that in the K-Means algorithm the optimum number of clusters is 3 clusters while the K-Medoids algorithm is 2 clusters. The best cluster evaluation uses the Davies-Bouldin Index (DBI) method and the results show that the K-Means DBI value is always smaller than the K-Medoids algorithm in either 2 clusters or 3 clusters, this shows that the K-Means algorithm is better than the K-Medoids algorithm. Based on these results, it is recommended to map priority areas for handling diarrheal diseases using the K-Means algorithm with 3 clusters, namely medium priority areas consisting of 9 regions, high priority areas consisting of 3 regions and low priority areas consisting of 25 regions. The results of the mapping can be used as input for the Kuningan District Health Office to develop strategies for preventing and preventing diarrheal diseases in children under five.

Keywords : Data Mining; K-Means; K-Medoids; Diarrhea.

Abstrak

Diare merupakan penyakit endemik yang memberikan kontribusi terhadap tingginya angka kematian di Indonesia khususnya pada balita. Dinas Kesehatan Kabupaten Kuningan mengalami kesulitan dalam pemantauan dan pengawasan penyebaran penyakit diare. Penelitian ini bertujuan untuk menghasilkan skema pemetaan daerah prioritas dalam penanganan pencegahan dan pengendalian penyebaran penyakit diare pada balita di Kabupaten Kuningan. Metode yang digunakan adalah metode *Data Mining Clustering* dengan membandingkan dua algoritma yakni algoritma K-Means dan algoritma K-Medoids. Penentuan jumlah *cluster optimum* menggunakan metode *Elbow* dan *Silhouette Coefficient*. Dengan metode tersebut diperoleh hasil bahwa pada algoritma K-Means jumlah *optimum cluster* adalah 3 *cluster* sedangkan algoritma K-Medoids adalah 2 *cluster*. Evaluasi *cluster* terbaik menggunakan metode *Davies-Bouldin Index* (DBI) dan diperoleh hasil bahwa nilai DBI K-Means selalu lebih kecil daripada algoritma K-Medoids baik pada 2 *cluster* ataupun 3 *cluster*, hal ini menunjukkan bahwa algoritma K-Means lebih baik daripada algoritma K-Medoids. Berdasarkan hasil tersebut, maka merekomendasikan pemetaan daerah prioritas penanganan penyakit diare menggunakan algoritma K-Means dengan jumlah 3 *cluster*, yaitu daerah prioritas sedang terdiri dari 9 daerah, daerah prioritas tinggi terdiri dari 3 daerah dan daerah prioritas rendah terdiri dari 25 daerah. Hasil pemetaan tersebut dapat menjadi masukan bagi Dinas Kesehatan Kabupaten Kuningan untuk membuat strategi penanggulangan dan pencegahan penyakit diare pada balita.

Katakunci: Data Mining; K-Means; K-Medoids; Diare

1. Pendahuluan

Menurut WHO (World Health Organization) diare adalah gejala infeksi yang ditimbulkan oleh berbagai mikroorganisme misalnya bakteri, virus dan

parasit, yang sebagian besar disebabkan melalui air yang tercemar oleh tinja (kotor). Diare adalah suatu keadaan pengeluaran tinja yang tidak normal atau tidak seperti biasanya. Perubahan yang terjadi berupa terdapatnya peningkatan volume, keenceran, dan

*) Penulis korespondensi: ragel.trisudarmo@uniku.ac.id

frekuensi dengan atau tanpa lendir darah, seperti lebih dari 3 kali/hari dan pada neonatus lebih dari 4 kali/hari (Selviana, 2017).

Penyakit diare tidak hanya menyerang orang dewasa, anak-anak juga dapat terkena. Berdasarkan data (Kemenkes, 2020), nilai persentase angka kematian pada kelompok anak usia 29 hari - 11 bulan yakni 14,5%, ini merupakan penyebab kematian kedua setelah *pneumonia*. Pada kelompok anak usia 12 sampai 59 bulan, diare merupakan penyebab kematian terbanyak sebesar 42,83%. Di Kabupaten Kuningan, kelompok umur dengan prevalensi diare tertinggi adalah kelompok umur balita 1 sampai 4 tahun, berdasarkan data yang diperoleh dari dinas kesehatan pada triwulan pertama tahun 2022 mencapai 36,06%.

Penyakit diare merupakan penyakit endemis di Jawa Barat, serta berpotensi Kejadian Luar Biasa (KLB) yang sering disertai dengan kematian. Berdasarkan persentase cakupan pelayanan penderita diare (semua umur) tahun 2020 sebesar 47,57 %, sementara cakupan pelayanan penderita diare (balita) tahun 2020 di Kabupaten Kuningan sebesar 34,1 % lebih besar dari Kabupaten Indramayu sebesar 31,9 %. (Dinkes, 2020). Hal ini terjadi karena kesulitan dalam pemantauan dan pengawasan penyebaran penyakit menular (diare), pemutakhiran data yang tidak akurat, berdampak kepada manajemen tidak dapat bertindak cepat dan tidak tepat sasaran.

Luasnya wilayah Kabupaten Kuningan cukup sulit mencari data individu, sehingga perlu dilakukan pemetaan daerah prioritas. Dengan menggabungkan data prevalensi diare dari setiap puskesmas maka dapat mencari data prevalensi diare di Kabupaten Kuningan berdasarkan hasil pemetaan daerah prioritas.

Metode *clustering* adalah suatu teknik memisahkan sekumpulan data menjadi kelompok atau *cluster* tertentu. Data yang terdapat kemiripan karakteristik yang satu sama dengan yang lain berada dalam satu klaster dan berbeda dengan data yang ada pada klaster lain (Supriyadi, 2021).

Algoritma K-Means merupakan metode analisis *cluster* yang bertujuan untuk memecah objek menjadi k *cluster* dengan menggunakan nilai rata-rata (*mean*) sebagai pusat *cluster* sedangkan algoritma K-Medoids menggunakan objek sebagai perwakilan (*medoid*) sebagai pusat *cluster* untuk setiap *cluster* (Riyanto, 2019).

Terdapat beberapa penelitian mengenai penerapan algoritma K-Means dan K-Medoids diantaranya penelitian yang dilakukan oleh (Farissa, 2021) mengenai evaluasi algoritma K-Means dan K-Medoids untuk pengelompokan obat di puskesmas karangsambung dengan menggunakan Silhouette Coefficient, didapatkan hasil bahwa algoritma K-Means mempunyai hasil cluster cukup baik dan masuk ke dalam kategori medium structure. Sedangkan penelitian menurut (Utomo, 2021) yaitu untuk

clustering penyebaran wabah covid-19 di Indonesia. Hasil penelitian menunjukkan bahwa berdasarkan nilai *Davies Boulden index* metode terbaik untuk mengelompokkan penyebaran wabah virus corona di Indonesia adalah metode K-Means.

Penelitian oleh (Farahdinna, 2019) dalam menentukan perbandingan klasterisasi produk asuransi perusahaan nasional menggunakan metode K-Means dan K-Medoid dengan eksperimen pembentukan cluster sebanyak 9 dengan nilai DBI terkecil pada metode K-Means nilai $k=5$ yaitu 0,018. Sedangkan metode K-Medoids memiliki nilai DBI terkecil pada $k=2$ sebesar 0,027. Menunjukkan bahwa pembentukan cluster yang paling optimal adalah menggunakan metode K-Means.

Penelitian lain yang dilakukan oleh (Aryuni, 2018) mengenai penerapan K-Means dan K-Medoids clustering pada data internet banking di bank XYZ. Menunjukkan bahwa metode *Elbow*, nilai k paling optimal baik pada K-Means maupun K-Medoids dengan 3 klaster. Algoritma K-Means memiliki performa yang lebih baik daripada K-Medoids, baik dari sisi nilai *average within centroid distance* dan kompleksitas waktu.

Berdasarkan penelitian terdahulu, penentuan jumlah cluster optimum sebagian menggunakan metode DBI pada tahap evaluasi. Lainnya menggunakan satu metode penentu jumlah cluster optimum yaitu menggunakan salah satu dari metode Silhouette Coefficient atau metode *Elbow*. Oleh karena itu, penelitian ini dilakukan menggunakan dua metode penentuan jumlah cluster optimum yaitu metode *Silhouette Coefficient* dan juga metode *Elbow*. Dengan menggunakan dua metode tersebut, diharapkan hasil clustering dengan algoritma K-Means dan K-Medoids lebih sensitif menghasilkan skema pemetaan daerah prioritas dalam penanganan pencegahan dan pengendalian penyebaran penyakit diare pada balita di Kabupaten Kuningan.

2. Kerangka Teori

2.1. Data Mining

Menurut (Vercellis, 2009) *data mining* adalah aktivitas yang menggambarkan sebuah proses analisis yang terjadi secara iteratif pada database yang besar, dengan tujuan mengekstrak informasi dan *knowledge* yang akurat dan berguna untuk pengambilan keputusan.

1) Operasi *Data Mining*

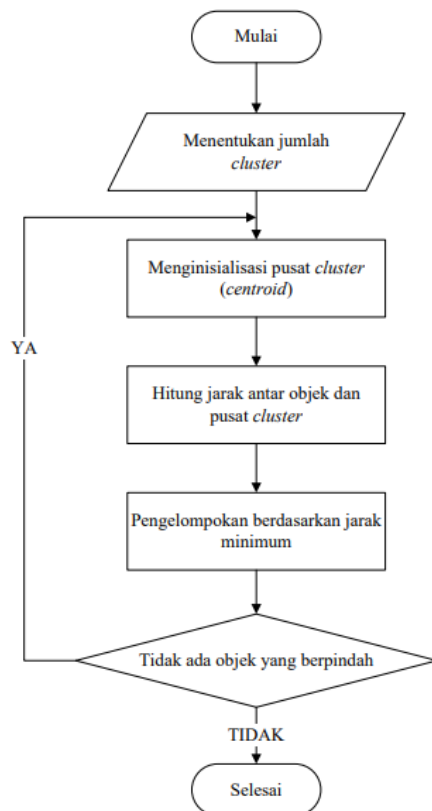
Menurut sifatnya, operasi *data mining* adalah sebagai berikut (Hermawati, 2013) :

- a) Prediksi atau *prediction driven* sifatnya menjawab pertanyaan apa dan sesuatu yang bersifat abstrak/transparan. Operasi prediksi dipakai untuk validasi hipotesis, *querying* serta pelaporan.
- b) Penemuan atau *discovery driven* bersifat transparan serta untuk menjawab pertanyaan

”mengapa?”. Operasi penemuan dipakai dalam analisis data eksplorasi, pemodelan prediktif, segmentasi database, analisis keterkaitan (*link analysis*) dan deteksi deviasi.

2.2. K-Means

Algoritma K-Means adalah algoritma klasik untuk memecahkan masalah *clustering*. Metode ini merupakan teknik penyekatan (*partition*) yang membagi atau memisahkan objek ke daerah yang terpisah (Adiana, 2018). Proses *clustering* menggunakan algoritma K-Means ditunjukkan oleh *flowchart* pada Gambar 1.



Gambar 1. Flowchart Algoritma K-Means
Sumber: (Adiana, 2018)

Langkah pada proses Algoritma K-Means:

- 1) Menentukan jumlah *cluster* yang akan dibentuk.
- 2) Menentukan pusat *cluster* awal (*centroid*) pada setiap *cluster* secara random.
- 3) Menghitung jarak setiap data observasi terhadap *centroid* terpilih menggunakan rumus *Euclidean distance* pada persamaan (1): (1)

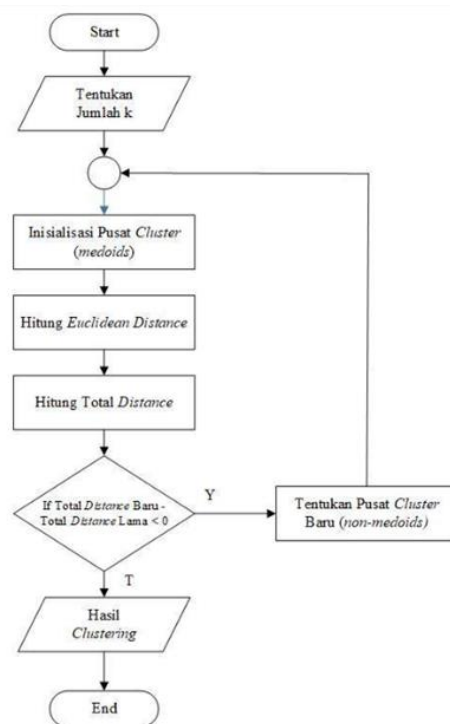
$$dist(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

- 4) Mengelompokkan semua data pengamatan berlandaskan kedekatannya dengan *centroid* atau jarak minimum

- 5) Perbaharui nilai hasil *centroid* baru yang di peroleh dari rata-rata (*means*) *cluster* yang sesuai
- 6) Ulangi langkah 3 hingga 5, sampai anggota setiap kelompok tidak berubah.

2.3. K-Medoids

Algoritma K-Medoids atau disebut juga dengan *Partitioning Around Medoids* (PAM) adalah algoritma *clustering* yang menggunakan objek sebagai perwakilan (*medoid*) sebagai pusat *cluster* untuk setiap *cluster*. Algoritma K-Medoids memiliki kelebihan untuk mengatasi kelemahan pada algoritma K-Means yang sensitive terhadap *noise* dan *outlier*, dimana objek dengan nilai yang besar yang memungkinkan menyimpang pada distribusi data. (Riyanto, 2019) Proses *clustering* menggunakan algoritma K-Medoids ditunjukkan oleh *flowchart* pada Gambar 2.



Gambar 1. Flowchart Algoritma K-Medoids
Sumber: (Andini, 2020)

Langka-langkah Algoritma K-Medoids:

- 1) Menentukan jumlah *cluster* yang akan dibentuk (*k*)
- 2) Menentukan pusat *cluster* awal (*medoids*) dari setiap *cluster* secara acak.
- 3) Tetapkan setiap objek pengamatan ke *cluster* terdekat memakai persamaan rentang *Euclidean Distance* pada persamaan (1).
- 4) Pilih secara acak objek disetiap *cluster* sebagai bakal *medoid* baru.
- 5) Hitung jarak dari setiap objek di setiap *cluster* dengan hasil bakal *medoid* baru.
- 6) Hitung pada simpangan total (S) dengan menghitung nilai jarak total *distance* baru – total

distance lama. Jika $S < 0$, maka menukar objek dengan data *cluster* untuk membentuk satu set k objek baru sebagai *medoid*

- 7) Ulangi langkah 4 sampai 6 sampai tidak terdapat perubahan *medoid*, sehingga dihasilkan berupa *cluster* dan masing-masing anggota *cluster*.

2.4. Metode Elbow

Metode *Elbow* merupakan sebuah metode visual untuk menentukan jumlah *cluster* (k) optimal yang akan dibentuk. Caranya dimulai dengan $k=2$ dan terus meningkat k di setiap langkah sebanyak 1. Nilai k akan menjadi titik di mana total jumlah kuadrat dalam *cluster* turun drastis. Setelah titik itu grafik mencapai dataran tinggi ketika k meningkat lebih jauh sehingga membentuk sebuah siku (Hung, 2019).

2.5 Silhouette Coefficient

Silhouette coefficient merupakan metode untuk menemukan jumlah klaster yang optimal dan interpretasi dan validasi konsistensi dalam *cluster* data. Metode *silhouette* menghitung koefisien siluet dari setiap titik yang mengukur seberapa mirip suatu titik dengan *clusternya* sendiri dibandingkan dengan *cluster* lain. dengan memberikan representasi grafis yang ringkas tentang seberapa baik setiap objek telah diklasifikasikan. *Silhouette coefficient* terletak pada rentang antara nilai -1 sampai 1. Rentang nilai tersebut digunakan untuk menunjukkan kedekatan kemiripan data yang dikelompokkan didalam suatu *cluster*. Jika nilai rata-rata *silhouette coefficient* mendekati nilai 1, maka *cluster* semakin baik. Sebaliknya jika nilai rata-rata *silhouette coefficient* mendekati -1, maka *cluster* tersebut tidak baik (Farissa, 2021).

2.6 Davies-Bouldin Index (DBI)

Davies-Bouldin Index (DBI) merupakan metode validasi *cluster* dari hasil *Clustering* dengan memaksimalkan jarak *inter cluster* dan meminimalkan jarak *intra cluster* (Kartikasari, 2021). DBI dapat dilihat dari nilai rasionya, model terbaik adalah model dengan nilai rasio paling kecil. Nilai ratio diperoleh dari hasil pembagian *average within* (SSW) dengan *average between* (SSB) dapat dilihat pada persamaan (2).

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i) \quad SSB_{i,j} = d(c_i, c_j) \quad (2)$$

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (2)$$

Dengan menggunakan DBI suatu *cluster* akan dianggap memiliki hasil *clustering* yang optimal jika memiliki DBI minimal (Gie, 2020).

2.7 Diare

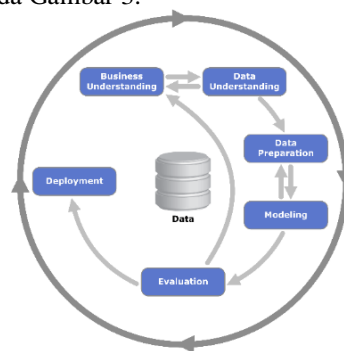
Diare merupakan gangguan Buang Air Besar (BAB) ditandai dengan BAB lebih dari 3 kali sehari

dengan konsisten tinja cair, dapat disertai dengan darah. (Riyanto, 2019).

Gejala diare bervariasi dan biasanya diawali dengan adanya perut kembung atau kram, tinja berair, mulas, atau terkadang mual dan muntah. Tergantung pada penyebab diare, penderita mungkin mengalami satu atau lebih gejala pada saat yang bersamaan.

3. Metode

Penelitian ini dilakukan menggunakan metode kuantitatif dengan tahapan penelitian menggunakan metode CRISP-DM. Dalam prosesnya CRISP-DM mempunyai enam tahapan pengembangan data mining (Chapman, 1999) yang menjadi acuan pada penelitian ini. Tahapan atau siklus proses CRISP-DM dapat dilihat pada Gambar 3.



Gambar 3. Siklus proses CRISP-DM
Sumber : (Chapman, 1999)

3.1. Business Understanding (Pemahaman Bisnis)

Tahap pertama adalah memahami tujuan dan kebutuhan dari sudut pandang bisnis, kemudian menterjemahkan pengetahuan ini ke dalam pendefinisian masalah dalam data mining. Dinas Kesehatan Kabupaten Kuningan bidang pencegahan dan pengendalian penyakit yang mengalami kesulitan dalam melakukan pemetaan daerah prioritas penanganan diare pada balita. Pemetaan daerah prioritas dapat dilakukan dengan menggunakan metode *clustering data mining*.

3.2. Data Understanding (Pemahaman Data)

Tahap ini dimulai dengan pengumpulan data yang kemudian dilanjutkan dengan proses untuk mendapatkan pemahaman yang mendalam tentang data. Data diambil dari Dinas Kesehatan Kabupaten Kuningan bidang pencegahan dan pengendalian penyakit pada periode triwulan pertama tahun 2022. Data tersebut terdiri dari 37 Puskesmas di Kabupaten Kuningan. Penelitian ini berupaya untuk melakukan pemetaan daerah prioritas penanganan diare pada balita, sehingga diperoleh tiga variabel yang relevan dengan tujuan tersebut, yaitu variabel usia balita kurang dari 6 bulan, usia balita 6 s/d kurang dari 12 bulan dan usia 1 s/d 5 tahun.

3.3. Data Preparation (Persiapan Data)

Tahap ini meliputi kegiatan untuk membangun kumpulan data akhir dari data mentah. Pada tahap ini juga mencakup pembersihan data (*data cleansing*), yaitu membersihkan data dari *noise* atau *missing value* dan *data transformation* yang dilakukan dengan cara normalisasi. Normalisasi dilakukan agar skala data tidak berbeda terlalu jauh (Monalisa, 2019) (Mirantika, 2021).

3.4. Modeling (Pemodelan)

Pemodelan dilakukan dengan membandingkan dua metode *clustering* yaitu K-Means dan K-Medoids. Data yang telah dikumpulkan selanjutnya diolah menggunakan algoritma K-Means dan K-Medoids dengan bantuan *software* R-Studio. Algoritma K-Means menggunakan nilai rata-rata (*mean*) sebagai pusat *cluster* sedangkan algoritma K-Medoids menggunakan objek (data observasi) sebagai perwakilan (*medoid*) menjadi pusat *cluster*. Hasil pengolahan data tersebut kemudian dilakukan uji validitas *cluster*.

3.5. Evaluation (Evaluasi)

Evaluasi hasil *clustering* yang digunakan adalah *Davies-Bouldin Index* (DBI). Pendekatan pengukuran DBI yaitu dengan memaksimalkan jarak *inter cluster* dan meminimalkan jarak *intra cluster*. Dengan menggunakan DBI suatu *cluster* akan dianggap memiliki hasil *clustering* yang optimal jika memiliki DBI minimal.

3.6. Deployment (Pengembangan)

Tahap *deployment* atau pengembangan, yaitu pengetahuan atau informasi yang telah diperoleh dalam hal ini adalah *cluster* atau pemetaan daerah dipresentasikan dalam bentuk laporan agar lebih mudah dipahami oleh Dinas Kesehatan. Laporan pemetaan daerah yang dihasilkan dapat menjadi masukan dalam membuat program atau strategi pencegahan dan penanggulangan penyakit.

4. Hasil dan Pembahasan

Dataset yang digunakan untuk pemodelan *clustering* adalah data laporan Dinas Kesehatan Kabupaten Kuningan bidang pencegahan dan pengendalian penyakit pada periode triwulan pertama tahun 2022. Dataset yang terdiri dari 3 variabel yaitu variabel usia kurang dari 6 bulan, usia 6 s/d kurang dari 12 bulan dan usia 1 s/d 5 tahun. Dataset penelitian dapat dilihat pada Tabel 1.

Tabel 1. Dataset Penelitian

Puskesmas	Usia <= 6	Usia <= 12	Usia 1 sd 5
	Bln	Bln	Thn
Darma	2	9	49
Kadugede	0	7	19
.....
Pancalang	1	2	13
Pasawahan	1	1	10

Setelah dataset disiapkan, langkah berikutnya adalah melakukan *data preparation* dengan melakukan *data cleansing* yaitu membersihkan data dari *noise* atau *missing value*. Untuk mengetahui *missing value* dilakukan dengan aplikasi Rstudio dengan *script* dan hasil berikut ini:

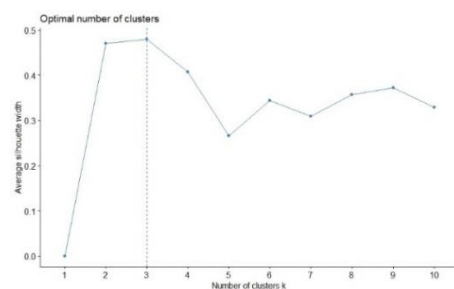
```
> #mengecek missing value
> anyNA(Diare1)
[1] FALSE
```

Hasil "FALSE" menunjukkan bahwa tidak ditemukan *missing value*, sehingga data dapat langsung digunakan untuk tahap selanjutnya yaitu *data transformation*. Pada *data transformation* dilakukan normalisasi yaitu proses transformasi dimana sebuah *atribut numeric* diskalakan dalam jarak lebih kecil. Normalisasi dilakukan dengan aplikasi Rstudio dan diperoleh hasil yang dapat dilihat pada Tabel 2.

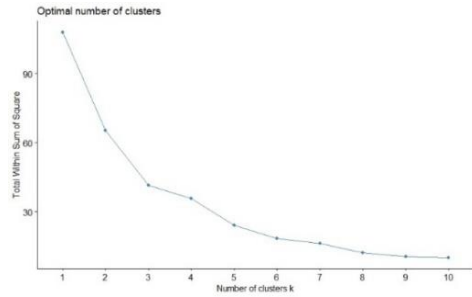
Tabel 2. Hasil Normlisasi

	Usia.Kurang. Dari.6.bulan.	Usia.6.sd.Kurang Dari.12.Bulan	Usia.1.sd. 5.Tahun
[1]	0.6435023	0.93545878	1.170783074
[2]	-0.8926000	0.53064036	-0.289407052
.....
[36]	-0.1245488	-0.48140569	-0.581445077
[37]	-0.1245488	-0.68381490	-0.727464090

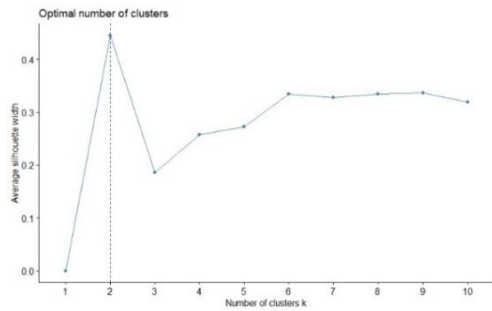
Setelah dilakukan *data preparation*, langkah selanjutnya adalah melakukan *modeling*, yaitu dengan menggunakan algoritma K-means dan K-Medoids. Dalam melakukan pemodelan K-Means dan K-medoids langkah pertama tentukan jumlah *cluster optimum* dengan menggunakan metode *silhouette* dan *elbow*. Hasil visualisasi metode tersebut dapat dilihat pada Gambar 4 s/d Gambar 7.



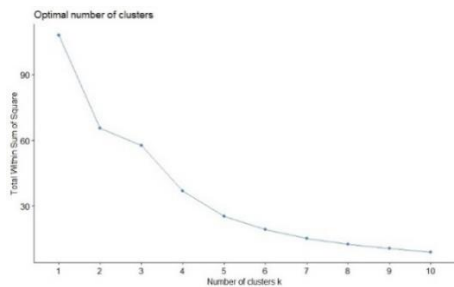
Gambar 4. Metode *Silhouette* Algoritma K-Means



Gambar 5. Metode *Elbow* Algoritma K-Means

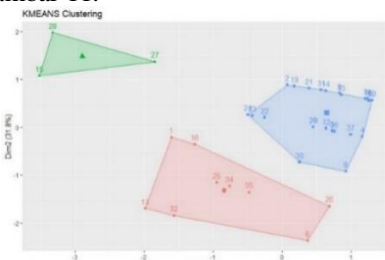


Gambar 6. Metode *Silhouette* Algoritma K-Medoids

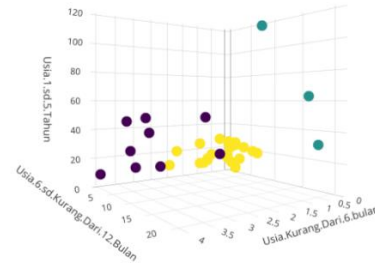


Gambar 7. Metode *Elbow* Algoritma K-Means

Dari gambar diatas, dapat diperoleh hasil bahwa pada algoritma K-means, jumlah *cluster optimum* baik pada metode *silhouette* maupun *elbow* adalah sama yaitu 3 *cluster* sedangkan pada metode K-Medoids, jumlah *cluster optimum* baik pada metode *silhouette* maupun *elbow* adalah 2 *cluster*. Dengan menggunakan jumlah *cluster optimum* tersebut, kemudian dilakukan proses *cluster* menggunakan algoritma K-Means dan K-Medoids. Visualisasi hasil *cluster* dengan algoritma K-Means dapat dilihat pada Gambar 8 dan Gambar 9 sedangkan untuk hasil *cluster* algoritma K-Medoids dapat dilihat pada Gambar 10 dan Gambar 11.



Gambar 8. Visualisasi *Cluster* 2D Algoritma K-Means



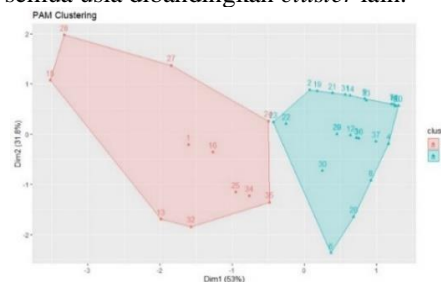
Gambar 9. Visualisasi *Cluster* 3D Algoritma K-Means

Algoritma K-Means dengan jumlah *cluster optimum* sebanyak 3 *cluster*, diperoleh 3 daerah pemetaan prioritas yaitu, daerah prioritas tinggi, daerah prioritas sedang dan daerah prioritas rendah yang dapat dilihat pada Tabel 3.

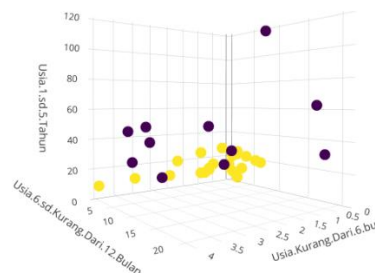
Tabel 3. Hasil *Cluster* Algoritma K-Means

<i>Cluster</i>	Usia <= 6 Bln	Usia <= 12 Bln	Usia 1 s/d 5 Thn	Jumlah Daerah	Pemetaan Prioritas
C1	3,11	6,11	30,89	9	Sedang
C2	0,33	15,67	70,67	3	Tinggi
C3	0,56	2,4	17,32	25	Rendah

Cluster 1 yang terdiri dari 9 daerah termasuk dalam daerah prioritas sedang yaitu dengan sebaran diare pada usia < 6 bulan yang tinggi tetapi sebaran pada usia 6 - <12 bulan dan usia 1-5 tahun yang sedang. *Cluster* 2 yang terdiri dari 3 daerah termasuk dalam daerah prioritas tinggi yaitu dengan sebaran usia 6- <12 bulan dan usia 1-5 tahun yang sangat tinggi, meskipun usia <6 bulan rendah sehingga *cluster* 2 ini menjadi daerah yang harus mendapat program penanggulangan lebih dahulu. Sedangkan *cluster* 3 yang terdiri dari 25 daerah termasuk dalam daerah prioritas rendah karena memiliki sebaran yang rendah pada semua usia dibandingkan *cluster* lain.



Gambar 10. Visualisasi *Cluster* 2D Algoritma K-Medoids



Gambar 11. Visualisasi *Cluster* 3D Algoritma K-Medoids

Algoritma K-Medoids dengan jumlah *cluster optimum* sebanyak 2 *cluster*, diperoleh 2 daerah pemetaan prioritas yaitu, daerah prioritas tinggi dan daerah prioritas rendah.

Tabel 4. Hasil *Cluster* Algoritma K-Medoids

Cluster	Usia	Usia	Usia 1	Jumlah Daerah	Pemetaan Prioritas
	<= 6 Bln	<= 12 Bln	s/d 5 Thn		
C1	2,09	9,82	45,36	11	Tinggi
C2	0,77	2,08	16,31	26	Rendah

Cluster 1 yang terdiri dari 11 daerah termasuk dalam daerah prioritas tinggi yaitu dengan sebaran diare pada semua usia mempunyai nilai yang tinggi sehingga *cluster* 1 ini menjadi daerah yang harus mendapat program penanggulangan lebih dahulu. Sedangkan *cluster* 2 yang terdiri dari 26 daerah termasuk dalam daerah prioritas rendah karena memiliki sebaran yang rendah pada semua usia dibandingkan *cluster* 1.

Evaluasi algoritma terbaik menggunakan metode *Davies-Bouldin Index* (DBI). Hasil *clustering* yang optimal diperoleh jika memiliki nilai DBI minimal. Dengan menggunakan aplikasi Rstudio diperoleh nilai DBI yang dapat dilihat pada Tabel 5.

Tabel 5. Nilai DBI

Metode	Jumlah Cluster	Nilai DBI
K-Means	2	1.22
	3	0.94
K-Medoids	2	1.25
	3	1.49

Nilai DBI algoritma K-Means selalu lebih kecil dari algoritma K-Medoids baik untuk jumlah *cluster* 2 ataupun jumlah *cluster* 3, hal ini menunjukkan bahwa algoritma K-Means lebih baik daripada algoritma K-Medoids. Hasil penelitian ini selaras dengan hasil penelitian yang dilakukan oleh (Utomo, 2021) bahwa dalam proses *clustering*, algoritma K-Means lebih baik daripada algoritma K-Medoids karena nilai DBI K-Means lebih kecil daripada DBI K-Medoids.

Pada jumlah *cluster* 2, nilai DBI K-Means adalah 1,22 lebih kecil sedikit dibandingkan K-Medoids 1,25. Pada jumlah *cluster* 3, nilai DBI K-Means adalah 0,94 jauh lebih kecil daripada K-Medoids 1,49. Jika dibandingkan jumlah *cluster* antara 2 *cluster* dan 3 *cluster*, maka jumlah 3 *cluster* jauh lebih kecil daripada 2 *cluster*. Oleh karena itu, penulis merekomendasikan pemetaan daerah prioritas penanganan penyakit diare menggunakan algoritma K-Means dengan jumlah 3 *cluster*, yaitu daerah prioritas sedang terdiri dari 9 daerah, daerah prioritas tinggi terdiri dari 3 daerah dan daerah prioritas rendah terdiri dari 25 daerah.

Hasil pemetaan daerah prioritas tersebut dapat menjadi masukan bagi Dinas Kesehatan Kabupaten Kuningan bidang pencegahan dan pengendalian

penyakit untuk membuat strategi dan program kerja yang tepat dalam melakukan penanggulangan dan pencegahan penyebaran penyakit diare pada balita.

5. Kesimpulan

Tujuan utama pada penelitian ini untuk mendapatkan pemetaan daerah prioritas penanganan penyakit diare pada balita di Kabupaten Kuningan dengan membandingkan algoritma K-Means dan K-Medoids. Penentuan *cluster* jumlah *optimum* dengan menggunakan metode *Shilouette* dan *Elbow* diperoleh hasil bahwa pada algoritma K-Means jumlah *optimum cluster* adalah 3 *cluster* sedangkan algoritma K-Medoids adalah 2 *cluster*. Evaluasi *cluster* terbaik menggunakan metode *Davies-Bouldin Index* (DBI) dan diperoleh hasil bahwa nilai DBI K-Means selalu lebih kecil daripada algoritma K-Medoids baik pada 2 *cluster* ataupun 3 *cluster*, hal ini menunjukkan bahwa algoritma K-Means lebih baik dari algoritma K-Medoids. Berdasarkan hasil tersebut, maka penulis merekomendasikan pemetaan daerah prioritas penanganan penyakit diare menggunakan algoritma K-Means dengan jumlah 3 *cluster*. Analisis algoritma K-Means dengan jumlah 3 *cluster* diperoleh hasil yaitu *Cluster* 1 termasuk dalam daerah prioritas sedang terdiri dari 9 daerah, *Cluster* 2 termasuk dalam daerah prioritas tinggi terdiri dari 3 daerah dan *cluster* 3 termasuk dalam daerah prioritas rendah terdiri dari 25 daerah. Dengan hasil pemetaan ini diharapkan Dinas Kesehatan Kabupaten Kuningan dapat membuat strategi dan program kerja yang tepat dalam melakukan penanganan dan pencegahan penyebaran penyakit diare pada balita.

Ucapan Terima Kasih

Penulis mengucapkan terimakasih kepada Lembaga Penelitian dan Pengabdian Kepada Masyarakat (LPPM) UNIKU dan Dinas Kesehatan Kabupaten Kuningan Bidang Pencegahan Dan Pengendalian Penyakit yang telah memberikan kesempatan dan dukungan hingga selesainya penelitian ini.

Daftar Pustaka

- Adiana, B.E., 2018. Analisis segmentasi pelanggan menggunakan kombinasi rfm model dan teknik clustering. *Jurnal Terapan Teknologi Informasi* 2(1), 23-32.
- Andini, A.D., 2020. Implementasi algoritma k-medoids untuk klusterisasi data penyakit pasien di RSUD Kota Bandung. *Jurnal Responsif: Riset Sains dan Informatika* 2(2), 128-138.
- Aryuni, M.M., 2018. Penerapan k-means dan k-medoids clustering pada data internet banking di Bank XYZ. *Jurnal Teknik dan Ilmu Komputer*.

- Chapman, P.C., 1999. The CRISP-DM user guide. In 4th CRISP-DM SIG Workshop in Brussels in March, Vol. 1999. Brussels: CRISP-DM SIG Workshop.
- Dinkes, P.J., 2020. Profil kesehatan Provinsi Jawa Barat tahun 2020. Bandung: Dinas Kesehatan Provinsi Jawa Barat.
- Farahdinna, F.N., 2019. Perbandingan algoritma k-means dan k-medoids dalam klasterisasi produk asuransi perusahaan nasional. *Jurnal Ilmiah Fifo* 11(2), 208-214.
- Farissa, R.A., 2021. Perbandingan algoritma k-means dan k-medoids untuk pengelompokan data obat dengan silhouette coefficient di Puskesmas Karangsembung. *Journal of Applied Informatics and Computing (JAIC)* 5(2), 109-116.
- Gie, W., Jollyta, D., 2020. Perbandingan euclidean dan manhattan untuk optimasi cluster menggunakan davies bouldin index: status covid-19 wilayah Riau. *Seminar Nasional Riset Information Science (SENARIS)* Vol. 2, 187-191.
- Hermawati, F.A., 2013. *data mining*. Yogyakarta: Andi Offset.
- Hung, P.D., 2019. Customer segmentation using hierarchical agglomerative clustering. *International Conference on Information Science and Systems*, 33-37.
- Kartikasari, M.D., 2021. Self-organizing map menggunakan davies-bouldin index dalam pengelompokan Wilayah Indonesia berdasarkan konsumsi pangan. *Jambura Journal of Mathematics* 3(2), 187-196.
- Kemenkes., 2020. Profil Kesehatan Indonesia. Jakarta: Kementerian Kesehatan Republik Indonesia.
- Mirantika, N., 2021. Penerapan algoritma k-means clustering untuk pengelompokan penyebaran covid-19 di Provinsi Jawa Barat. *Nuansa Informatika* 15(2), 92-98.
- Monalisa, S.N., 2019. Analysis for customer lifetime value categorization with RFM model. *Procedia Computer Science* 161, 834-840.
- Riyanto, B., 2019. Penerapan algoritma k-medoids clustering untuk pengelompokan penyebaran diare di Kota Medan (Studi Kasus: Kantor Dinas Kesehatan Kota Medan). *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)* 3(1), 562-568.
- Selviana, T.E., 2017. Faktor-faktor yang berhubungan dengan kejadian diare pada anak usia 4-6 Tahun. *Jurnal Vokasi Kesehatan* 3(1), 28-34.
- Supriyadi, A.T., 2021. Perbandingan algoritma k-means dengan k-medoids pada pengelompokan armada kendaraan truk berdasarkan produktivitas. *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)* 6(2), 229-240.
- Utomo, W., 2021. The comparison of k-means and k-medoids algorithms for clustering the spread of the covid-19 outbreak in Indonesia. *ILKOM Jurnal Ilmiah* 13(1), 31-35.
- Vercellis, C., 2009. *Business intelligence: data mining and optimization for decision making*. John Wiley & Sons.