IJAI (Indonesian Journal of Applied Informatics)
Vol. 9 No. 2 Tahun 2025 pISSN: 2548-3846, eISSN: 2598-5981

Optimalisasi Klasterisasi Tenaga Kesehatan Menggunakan K-Means dan Davies Bouldin Indexs

Ayura Yufita[1]*, Rudi Kurniawan[1], Yudhistira Arie Wijaya[1] ,Tati Suprapti[1]
[1]Program Studi S1Teknik Informatika, STMIK IKMI Cirebon, Indonesia
*Email: ayurayufita1413@gmail.com

Info Artikel

Kata Kunci :
tenaga kesehatan, optimalisasi, davies bouldin indeks, knowledge discovery in database, k-means clustering

Keywords :
health workers, optimisation, davies bouldin index, knowledge discovery in databases, k-means clustering

Tanggal Artikel
Dikirim : 18 Desember 2024
Direvisi : 25 Januari 2025
Diterima : 10 Februari 2025

Abstrak

Optimalisasi model pengelompokan data tenaga kesehatan adalah langkah strategis untuk memahami pola dan karakteristik kelompok data tertentu. Tujuan dari penelitian ini adalah untuk mendapatkan nilai K optimal *menurut Davies Bouldin Indeks* (DBI), mendapatkan nilai iterasi yang diperlukan oleh algoritma *K-Means* Clustering untuk mencapai hasil yang optimal, dan menentukan jenis metrik apa yang akan menghasilkan nilai (DBI) yang paling kecil. Hal ini penting karena penelitian ini membantu perencanaan distribusi tenaga kesehatan yang lebih efisien di wilayah Jawa Barat denfan menghasilkan klaster optimal berbasis K-Means dan Optimize Parameter Grid. Penggunaan metode *Knowledge Discovery in Database* (KDD), yang mencakup proses pemilihan, praproses, transformasi, data mining, dan interpretasi/ evaluasi hasil. Hasil penelitian ditunjukkan pada iterasi 1-10 menggunakan K=2 dengan nilai DBI terendah sebesar 0,377.

Abstract

Optimisation of health worker data clustering model is a strategic step to understand the patterns and characteristics of certain data groups. The objectives of this study are to obtain the optimal K value according to the Davies Bouldin Index (DBI), obtain the iteration value required by the K-Means Clustering algorithm to achieve optimal results, and determine what type of metric will produce the smallest (DBI) value. This is important because this research helps to plan a more efficient distribution of health workers in the West Java region by producing optimal clusters based on K-Means and Optimise Parameter Grid. The use of Knowledge Discovery in Database (KDD) method, which includes the process of selection, preprocessing, transformation, data mining, and interpretation/evaluation of results. The results showed in iterations 1-10 using K=2 with the lowest DBI value of 0.377.

1. PENDAHULUAN

Metode pengelompokan data telah diubah secara signifikan oleh kemajuan pesat dalam informatika, terutama dalam hal data tenaga kesehatan. Integrasi catatan kesehatan elektronik (EHR) telah memungkinkan pengumpulan kumpulan data yang sangat besar [1]. Pengelompokan K-Means membantu menemukan pola dan perbedaan dalam distribusi tenaga kesehatan di berbagai wilayah, yang memungkinkan pengoptimalan alokasi sumber daya dan peningkatan layanan kesehatan [2]. Selain bidang kesehatan, kemajuan di bidang informatika ini berdampak pada teknologi, bisnis, dan pendidikan. Munculnya analitik data besar dalam teknologi telah menghasilkan penciptaan alat yang lebih canggih untuk pemrosesan data. Alat-alat ini memungkinkan perusahaan untuk mendapatkan pemahaman yang dapat diandalkan tentang kumpulan data yang rumit [3].

Salah satu masalah penting yang harus dipertimbangkan adalah analisis masalah utama dalam analisis data tenaga kerja kesehatan. Masalah seperti distribusi tenaga kerja kesehatan yang tidak merata, kurangnya data terintegrasi antar wilayah, atau kesulitan Mengelompokan untuk memahami kebutuhan tenaga kerja, hal ini sesuai dengan hasil penelitian [4]. Selain itu kurangnya data terintegrasi antar wilayah menyulitkan analisis data tenaga kerja kesehatan. Ini dapat menyulitkan pengawasan dan evaluasi yang menyeluruh atas pemerataan layanan kesehatan [5]. Dampak pandemi COVID-19 terhadap kesejahteraan mental tenaga kesehatan adalah tantangan tambahan yang dihadapi. Ini dapat menyebabkan kekurangan tenaga kerja dan kelelahan karyawan di layanan kesehatan pedesaan [6].

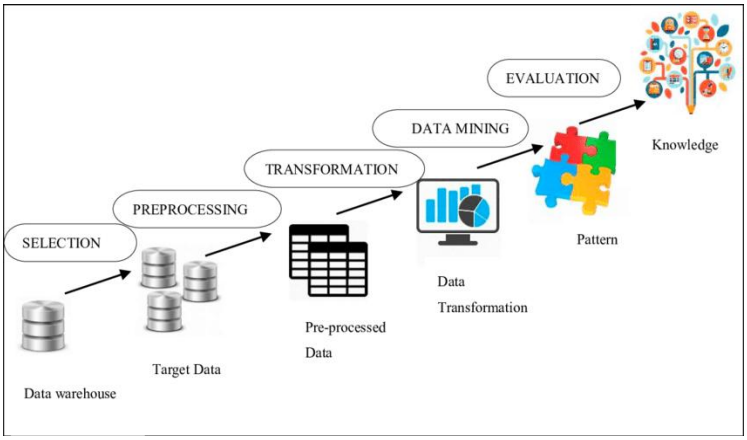
Sejumlah penelitian telah membahas penggunaan algoritma K-Means, dalam analisis tenaga kesehatan. Penerapan algoritma K-means digunakan untuk mengevaluasi distribusi sumber daya dan aksesibilitas layanan kesehatan. Sebagai contoh, analisis aksesibilitas spasial di Cina telah dilakukan oleh Yue *et al*., yang menemukan distribusi fasilitas kesehatan yang tidak merata dan mendukung alokasi sumber daya yang masuk akal berdasarkan kepadatan penduduk [7]. Hal ini terbukti bahwa algoritma K-Means mampu mengelompokan data tenaga kesehatan ditingkat regional, dan sangat penting untuk memastikan bahwa setiap wilayah memiliki akses yang memadai terhadap tenaga kesehatan, guna meningkatkan derajat kesehatan masyarakat secara keseluruhan[8].

Tujuan penelitian ini adalah mengoptimalkan model pengelompokan melalui penerapan *Optimize Parameter Grid* untuk meningkatkan efisiensi dan akurasi hasil pengelompokan [9]. Dengan metode KDD (*Knowledge Discovery in Database*) digunakan untuk mengelola dan menganalisis data tenaga kesehatan secara efektif dalam penelitian ini [10]. Penelitian ini menggunakan metode *Parameter Grid* untuk menentukan nilai beberapa parameter, seperti jumlah cluster, tipe pengukuran jarak, dan jumlah iterasi, sehingga menghasilkan hasil penggelompokan yang sangat baik. Hasil pengelompokan dievaluasi menggunakan Davies Bouldin Index. Penggunaan *Parameter Grid* dalam penelitian ini akan membuat pemilihan parameter lebih optimal, sehingga tidak lagi bergantung pada penilaian manual yang memakan waktu.

Penelitian ini juga berfokus pada mengidentifikasi masing-masing kelompok untuk membantu proses pengambilan keputusan kesehatan yang lebih baik. Penelitian ini juga dapat memberikan manfaat untuk menganalisis data dengan lebih efisien, meningkatkan kinerja operasional, dan mendukung pengembangan kebijakan yang berbasis data [11].

2. METODE PENELITIAN

Proses analisis data menggunakan tahapan *Knowledge Discovery in Database* (KDD), yang bertujuan menentukan pola atau informasi penting dari kumpulan data. pada Gambar 3.2



Gambar 1. Tahapan Proses KDD

Berdasarkan gambar 3.2 tahapan penelitian terdiri dari tahapan-tahapan sebagai berikut:

(1) *Selection*

Selection data merupakan proses memilih dan menyeleksi data dengan memisahkan atribut data dengan mengambil atribut yang akan digunakan pada penelitian. Proses ini adalah merupakan langkah pertama dalam tahapan *Knowledge Discovery in Database* (KDD). Pada tahap ini data mengenai distribusi tenaga kesehan berdasarkan kecamatan di berbagai kabupaten di Jawa Barat. Meliputi nama kecamatan, jumlah tenaga kesehatan dan jenis tenaga kesehatan.

(2) *Preprocessing*

Pada *preprocessing* adalah proses membersihkan data atau *missing value*. Bertujuan untuk menghilangkan data yang salah input, data tidak relevan, data tidak memiliki nilai *(null),* data duplikat dan data tidak konsisten. Ini karena keberadaannya dapat mengurangi mutu atau akurasi hasil nantinya.

(3) *Transformation*

Transformation merupakan proses sebelum dilakukannya *data mining*. Tujuan transformasi data adalah untuk meningkatkan kualitas, konsistensi, dan relevansi data sehingga lebih mudah diolah berdasarkan algoritma dan *tools* yang akan digunakan dalam pengolahan data. Proses ini mencakup normalisasi (mengubah data ke skala tertentu), standardisasi (menyetarakan data dengan rata-rata nol dan standar deviasi satu), diskretisasi (mengubah data kontinu menjadi kategori), dan pengkodean (mengubah data kategori menjadi numerik). Data numerik lebih mudah diolah oleh algoritma karena komputasinya efisien, cocok dengan metrik jarak seperti Euclidean, dan lebih fleksibel dalam merepresentasikan informasi. Penggunaan metrik seperti Bregman Divergence memungkinkan hasil analisis lebih akurat dengan menyesuaikan distribusi data dan memengaruhi pusat cluster. Transformasi dan metrik yang tepat meningkatkan kualitas serta relevansi hasil analisis.

(4) *Data Mining*

Data mining merupakan salah satu proses pengelolaan data berdasarkan algoritma sesuai dengan teknik *data mining*. Pada tahap ini algoritma K-Means ditetapkan untuk mengelompokkan tenaga kesehatan berdasarkan atribut data yang ada. Pengujian hasil clustering menggunakan Indeks Validasi Davies Bouldien Index (DBI).

(5) *Evaluation*

Pada tahap evaluasi, akan diketahui apakah hasil daripada tahap *data mining* dapat menjawab tujuan yang telah ditetapkan. Hal ini dilakukan untuk mengetahui apakah pola atau model yang di temukan bermanfaat dan akurat.

(6) *Knowledge Discovery in Database*

Knowledge Discovery in Database (KDD) adalah proses menemukan informasi atau pola penting dari data besar dan rumit. Dalam proses KDD, tahap pertama adalah memahami masalah untuk menentukan tujuan analisis dan mengumpulkan informasi. Kedua, pilihan data dilakukan dengan memilih kumpulan data yang relevan dari berbagai sumber yang tersedia. Selanjutnya, proses *preprocessing* dilakukan, yang mencakup pembersihan, pengisian data yang tidak relevan, dan penghilangan anomali untuk memastikan kualitas data. Terakhir, transformasi data dilakukan dengan mengubah format atau struktur data menjadi yang sesuai untuk analisis, seperti *encoding* atau normalisasi. Kelima, data yang telah siap dianalisis dengan menggunakan algoritma *data mining* yang tepat, seperti clustering, klasifikasi, atau asosiatif, untuk menemukan pola tersembunyi. Terakhir, temuan analisis dievaluasi dan ditafsirkan pada tahap evaluasi dan interpretasi. Ini dilakukan untuk memastikan bahwa hasilnya relevan, akurat, dan dapat digunakan sebagai dasar pengambilan keputusan. Termasuk memahami masalah yang dihadapi, memilih kumoulan data terkait, memanipulasi dan mengubah data, menerapkan algoritma yang sesuai, dan melakukan analisis yang menyeluruh terhadap temuan yang dihasilkan.

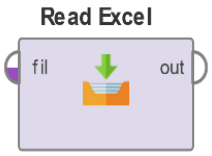
3. HASIL DAN PEMBAHASAN

3.1 Hasil penelitian

pengelompokkan menggunakan algoritma K-Means *Clustering* dalam mengolah data jumlah tenaga kesehatan perkecamatan dari beberapa kabupaten di provinsi Jawa Barat sebanyak 273 *record*. Data ini di proses menggunakan *tools* Rapidminer dengan tahapan komputasi sebagai berikut :

3.1.1 *Selection*

Langkah pertama dalam proses pemilihan menggunakan operator *Read Excel,* yang berfungsi untuk membaca file excel. Data yang digunakan memiliki 12 atribut, yaitu Nama Kecamatan, Jumlah Perawat, Jumlah Bidan, Tenaga Kefarmasiaan, Tenaga Kesehatan Masyarakat, Tenaga Kesehatan Lingkungan, Tenaga Gizi, Tenaga Medis, Tenaga Kesehatan Psikologi Klinis, Tenaga Keterapian Fisik, Tenaga Keteknisan Medis, dan Tenaga Teknik Biomedika. Semua data ini akan diproses lebih lanjut pada tahap data preprocessing. Dataset yang digunakan berbentuk file Excel menggunaka *read excel.* Seperti yang ditunjukan pada Gambar 2 sebagai berikut.



Gambar 2. Operator *Read Excel* pada Rapidminer

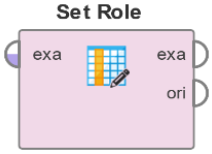
Pada Gambar 2 *Read Excel* adalah salah satu operator yang digunakan untuk membaca data dari file Excel. Operator ini mengimport dataset dari file excel dengan mengatur parameter seperti lokasi file, pemilihan halaman, rentang sel yang diimport, baris yang digunakan sebagai *header.* Menunjukkan hasil yang diperoleh dari proses *Read Excel.* Gambar ini menggambarkan data terkait yang akan digunakan untuk analisis selanjutnya dalam konteks penelitian.

Pada Gambar 3 menunjukkan hasil yang diperoleh dari proses *Read Excel.* Dengan menggambarkan data terkait yang akan digunakan untuk analisis selanjutnya dalam konteks penelitian.

Kecamatan polynominal	Tenaga Kes... integer	Tenaga Kes... integer	Tenaga Kes... integer	Tenaga Kes... integer	Tenaga Kes... integer	Tenaga Kes... integer	Jumlah Tena integer
Banjarsari	33	38	9	5	1	3	8
Banjaranyar	11	17	2	1	1	0	2
Lakbok	20	29	4	3	2	1	4
Purwadadi	18	25	2	2	1	1	2
Pamarican	32	42	16	7	0	5	6
Cidolog	13	16	3	2	0	1	3
Cimaragas	17	18	4	2	1	1	3
Cijeungjing	46	42	16	7	3	3	21
Cisaga	11	18	2	4	1	1	4
Tambaksari	13	13	3	2	1	1	2
Rancah	23	24	5	3	2	1	8
Rajadesa	18	31	3	1	2	1	3
Sukadana	13	15	5	3	1	2	3
Ciamis	543	134	83	35	8	16	93
Baregbeg	179	45	26	2	2	4	36
Cikoneng	22	21	6	1	1	1	5
Sindangkasih	11	31	5	2	1	1	6
Cihaurbeuti	37	28	10	4	2	3	10

Gambar 3. Hasil Read Excel pada Rapidminer

Pada Gambar 4 menunjukan bahwa langkah selanjutnya yaitu penggunaan operator Set Role yang bertujuan Untuk mengganti atribut kecamatan menjadi ID.



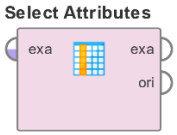
Gambar 4. Operator Set Role pada Rapidminer

Dari hasil pembacaan operator Set Role didapatkan informasi hasil data statistik Gambar 5

Kecamatan	Nominal	0	Weru (1)	Bungursari (2)	Bungursari (2), Cibitung (2), ... [260 more]
Tenaga Kesehatan - Perawat	Integer	0	Min 4	Max 1313	Average 66.788
Tenaga Kesehatan - Bidan	Integer	0	Min 8	Max 230	Average 37.875
Tenaga Kesehatan - Tenaga K...	Integer	0	Min 0	Max 170	Average 10.531
Tenaga Kesehatan - Tenaga K...	Integer	0	Min 0	Max 46	Average 2.586
Tenaga Kesehatan - Tenaga K...	Integer	0	Min 0	Max 17	Average 1.916
Tenaga Kesehatan - Tenaga Gizi	Integer	0	Min 0	Max 35	Average 3.147
Jumlah Tenaga Medis	Integer	0	Min 0	Max 185	Average 9.960
Jumlah Tenaga Kesehatan Psi...	Integer	0	Min 0	Max 2	Average 0.011
Jumlah Tenaga Keterapian Fisik	Integer	0	Min 0	Max 14	Average 0.256
Jumlah Tenaga Keteknisan Me...	Integer	0	Min 0	Max 83	Average 3.571
Jumlah Tenaga Teknik Biomed...	Integer	0	Min 0	Max 66	Average 2.249

Gambar 5. Hasil Data Statistik

Proses selanjutnya yaitu *select attribute* untuk menentukan data yang paling penting yang mendukung pengelompokan. Data ini dianggap signifikan karena membantu membedakan masing-masing kelompok. Proses ini dilakukan menggunakan fitur *select attribute* di aplikasi Rapidminer seperti yang ditunjukan pada Gambar 6



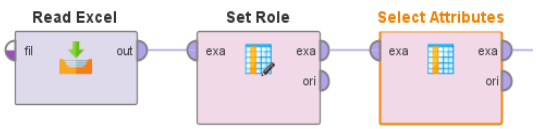
Gambar 6 .Operator *Select attribute* pada *Rapidminer*

Dari hasil penggunaan operator *Select Attribute* hasil tampak pada Gambar 7

Row No.	Kecamatan	Tenaga Kes...	Tenaga Kes...	Tenaga Kes...	Tenaga Kes...	Tenaga Kes...	Tenaga Kes...	Jumlah Tena...	Jumlah T
1	Banjarsari	33	38	9	5	1	3	8	0
2	Banjaranyar	11	17	2	1	1	0	2	0
3	Lakbok	20	29	4	3	2	1	4	0
4	Purwadadi	18	25	2	2	1	1	2	0
5	Pamarican	32	42	16	7	0	5	6	0
6	Cidolog	13	16	3	2	0	1	3	0
7	Cimaragas	17	18	4	2	1	1	3	0
8	Cijeungjing	46	42	16	7	3	3	21	0
9	Cisaga	11	18	2	4	1	1	4	0
10	Tambaksari	13	13	3	2	1	1	2	0
11	Rancah	23	24	5	3	2	1	8	0
12	Rajadesa	18	31	3	1	2	1	3	0
13	Sukadana	13	15	5	3	1	2	3	0
14	Ciamis	543	134	83	35	8	16	93	1
15	Baregbeg	179	45	26	2	2	4	36	0
16	Cikoneng	22	21	6	1	1	1	5	0
17	Sindangkasih	11	31	5	2	1	1	6	0

Gambar 7. Operator *Select Attribute*

Model proses *Selection* pada Rapidminer yang terlihat pada Gambar 7



Gambar 8. Model proses *Selection*

3.1.2 *Preprocessing*

Processing dilakukan untuk memeriksa apakah ada data yang *missing*, *noise*, atau duplikat data yang akan digunakan agar tidak ada kesalahan atau masalah saat proses. Pada hasil result dari statistik dataset seperti pada Gambar 5 diketahui bahwa masing-masing atribut tidak memiliki nilai *missing*. menunjukkan bahwa proses *Preprocessing* tidak dilakukan karena semua atribut sudah terbebas dari ketidak konsistenan data. Maka disimpulkan data sudah layak digunakan untuk tahap berikutnya, yaitu transformasi data.

.

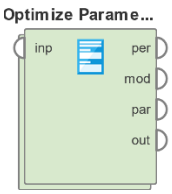
3.1.3 *Transformation*

Pada tahap ini semua data akan diolah untuk menghasilkan kelompok atribut yang akan digunakan dalam proses transformasi dengan mengubah tipe data nominal ke tipe data numerik. Dikarenakan algoritma K-Means hanya dapat mengolah data dalam bentuk numerik saja, maka data kategori atau teks perlu diubah menjadi numerik. Transformasi ini juga bertujuan membantu menyederhanakan data sehingga model dapat lebih mudah di pahami. Pada Gambar 5 diatas menunjukan bahwa penelitian ini tidak dilakukan proses *transformation*. Dikarenakan data sudah *Integer* atau bersifat numerik.

.

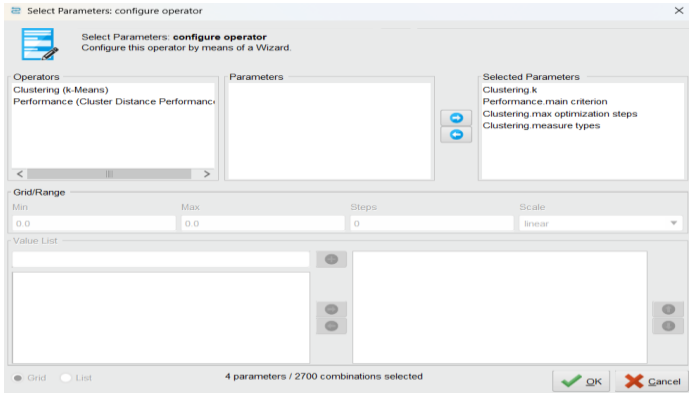
3.1.4 *Data Mining*

Untuk mendapatkan nilai K, tipe *meansure type*, dan nilai iterasi yang menghasilkan performa paling optimal diimplementasikan operator *Optimize Parameter Grid*. Dimana *Optimize Parameter Grid* adalah operator yang menjalankan proses didalamnya unuk mencoba semua kombinasi nilai parameter. operator ini bekerja secara *iterative* untuk membuat jarak antar data dalam satu cluster sekecil mungkin. Dengan cara ini, analisis lebih lanjut menjadi lebih mudah karena kelompok data tertentu dapat dilihat dengan jelas.



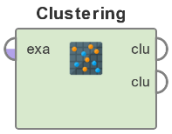
Gambar 9. Penerapan operator *Optimize Parameter Grid*

Adapun *select* parameter yang digunakan pada operator *Optimize Parameter Grid* dapat dilihat pada Gambar



Gambar 10 . *Select Parameter* pada *Optimize Parameter Grid*

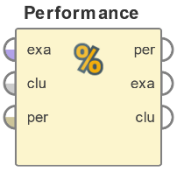
Pada tahap ini penerapan algoritma K-Means *Clustering* ditunjukan pada Gambar 11 operator ini secara efisien mengelompokkan data yang mirip satu sama lain menggunakan data yang berbentuk angka. Algoritma K-means sangat efektif untuk mengelompokkan banyak data berdasarkan kesamaan atau jarak antar data. Dengan cara ini, data yang tidak memiliki label bisa dikelompokkan kedalam beberapa cluster, dimana setiap cluster berisi data yang karakteristiknya sama/mirip.



Gambar 11. Operator *Clustering* pada Rapidminer

Parameter yang digunakan pada operator *Clustering* sudah diatur pada *operator Optimize Parameter Grid* dapat dilihat pada Gambar 10 adapun parameter lainnya dibuat secara default.

Proses selanjutnya yaitu menggunakan operator *Cluster Distance Performance* atau disebut sebagai operator kinerja kelompok jarak jauh. Operator ini berfungsi untuk menilai validitas pengelompokkan data. Metode ini memastikan kedekatan titik data ke centroid di dalam cluster, sementara secara bersamaan menjamin maksimalisasi jarak yang memisahkan cluster. Penerapan operator *Cluster Distance Performance* diilustrasikan pada 12



Gambar 12. Operator *Cluster Distance Performance*

Parameter yang digunakan pada operator *Cluster Distance Performance* sudah diatur pada operator *Optimize Parameter Grid* dapat dilihat pada Gambar 10 adapun parameter lainnya dibuat secara *default*.

Dari penerapan operator *Optimize Parameter Grid* Hasil parameter dapat dilihat pada Tabel yang menjelaskan Hasil *Select Parameter* Hasil dari penerapan *Optimize Parameter Grid* dapat dilihat pada Tabel 1.

Tabel 1 Hasil *Select Parameter* [1]

No	Clustering	*Performance*	Iterasi	*Measure Type*	DBI
1	2	Davies Bouldin	1	BregmanDivergences	0,37749
2	2	Davies Bouldin	2	BregmanDivergences	0,37749
3	2	Davies Bouldin	3	BregmanDivergences	0,37749
…..	…..	…..	…..	…..	…..
2700	9	Davies Bouldin	44	NumericalMeasures	1,00823

Dari hasil penggunaan operator *Optimize Parameter Grid* Algoritma K-Means Clustering menghasilkan Model Cluster yang dapat dilihat pada Tabel 2.

Tabel 2 Model *Cluster* [2]

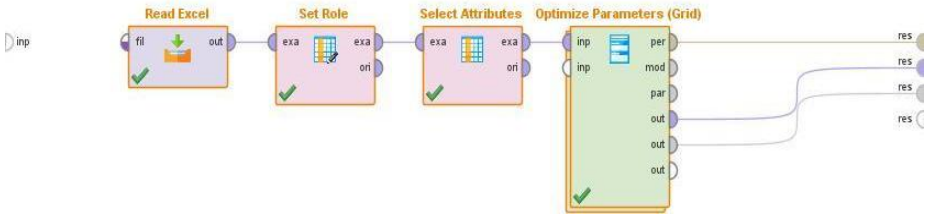
No	Model cluster	Jumlah Record
1.	Cluster_0	263 item
2.	Cluster_1	10 item
Total Number of items = 273		

Adapun hasil dari penggunaan parameter pada operator *Optimize Parameter Grid* menghasilkan operator *Cluster Distance Performance* dapat dilihat pada Tabel 3 berikut ini

Tabel 3 Hasil Performance Vector pada operator *Cluster Distance Performance* [3]

No	PerformanceVector:	Hasil
1.	Avg. within centroid distance:	8356.771
2.	Avg. within centroid distance_cluster_0	6306.722
3.	Avg. within centroid distance_cluster_1	62273.050
4.	Davies Bouldin	0.377

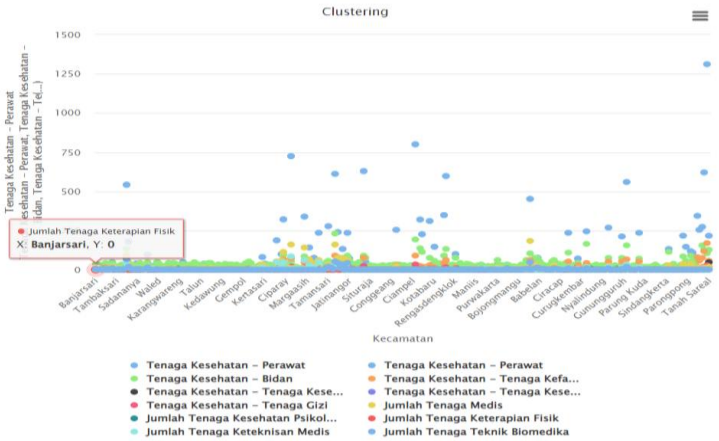
Model proses *clustering* pada Rapidminer yang terlihat pada Gambar 13.



Gambar 13. Model Proses Clustering

3.1.5 *Evaluation*

Hasil penelitian juga menunjukkan bahwa percobaan pada K=2, rata-rata jarak dalam *cetroid* menghasilkan 8356,771 dengan nilai DBI terendah sebesar 0,377 yang menunjukkan kualitas *cluster* terbaik. Namun, ketika nilai K di *maximize* menjadi 8, rata-rata jarak dalam *cetroid* menurun menjadi 821,986, sementara nilai DBI meningkat menjadi 1,008. Hal ini mengidentifikasi bahwa K=2 dipilih sebagai nilai optimal dalam clustering dikarenakan nilai DBI merupakan yang paling rendah. Hasil uji nilai K dapat dilihat pada Tabel 4.8 Model *Cluster*. Adapun hasil evaluasi *Meansure Type* dengan nilai k=2 menggunakan *Breagman Divergen*ce dengan 10 kali iterasi memberikan hasil terbaik. Dengan nilai *Davies Bouldin Indeks* (DBI) sebesar 0,377, menunjukkan bahwa jarak antar cluster optimal dan jarak antar data dalam setiap cluster tersusun dengan konsisten Adapun hasil visualisasinya dalam bentuk grafik *scatter/ bubble* yang dapat dilihat pada Gambar 14



Gambar 14. Hasil Visualisasi

Gambar 14 menjelaskan bentuk visualisasi scatter/bubble yang merupakan titik-titik atau "bubble" yang masing-masing mempresentasikan satu atau lebih variabel dalam suatu dataset. Sumbu x dan y menunjukkan variabel-variabel yang akan dibandingkan, sedangkan ukuran setiap titik menunjukkan nilai variable kelima. Dimana hasil visualisasi dapat dijelaskan sebagai berikut:

1. Sumbu X menunjukkan kategori "Tenaga kesehatan-Tenaga gizi" yaitu jumlah tenaga kesehatan yang bekerja di bidang gizi disetiap cluster.
2. Sumbu Y mencakup jenis tenaga kesehatan lainnya, seperti perawat, bidan, atau tenaga kefarmasian.
3. Warna setiap titik menggambarkan cluster yang dihasilkan algoritma clustering berbeda warna.
4. Kelompok biru (Cluster_0) menunjukkan area yang memiliki ciri khusus, sedangkan kelompok hijau (Cluster_1) menunjukkan area lain dengan pola yang berbeda.
5. Dalam bagian "Ukuran" tidak ada variable tambahan yang mempengaruhi ukuran titik.
6. Setiap titik data menunjukkan wilayah tertentu, yaitu jumlah tenaga kesehatan dalam cluster dibidang tertentu..

3.2 Pembahasan

Penelitian ini membahas hasil pengelompokan data tenaga kesehatan menggunakan algoritma K-Means dengan menerapkan analisis lebih mendalam. Adapun analisis dilakukan untuk:

3.2.1 Mendapatkan nilai K optimal berdasarkan dbi

Hasil pengelompokkan data tenaga kesehatan menggunakan Rapidminer AI Studio 2024 dengan algoritma K-Means. Hasil pengelompokan ini telah dianalisis lebih lanjut pada bagian diskusi. Karena memiliki nilai DBI terendah, yaitu 0,377, yang menunjukkan pemisahan cluster yang jelas, eksperimen menunjukkan bahwa nilai K=2 adalah nilai yang paling ideal untuk pengelompokan data tenaga kesehatan, hal itu ditunjukkan pada Tabel 2. Untuk mengetahui jarak yang ada di antara data dalam cluster, Teknik jarak geometris sesuai dengan karakteristik data, dan menghasilkan hasil pengelompokan yang lebih konsisten. Penelitian ini sejalan dengan penelitian Sembiring Brahmana yang mendapatkan jumlah cluster (k) yang optimal untuk algoritma K-Means clustering yang menggunakan nilai Davies-Bouldin Index (DBI) telah menjadi subjek yang sangat dibahas[12]. Metode ini sering digunakan bersama dengan indeks validitas internal seperti Davies-Bouldin Index (DBI) untuk menilai kualitas pengelompokan, dimana hasil sejalan dengan penelitian [13]

3.2.2 mendapatkan nilai iterasi yang di butuhkan oleh algoritma K-Means

Untuk mendapatkan niilai iterasi yang dibutuhkan oleh algoritma K-Means digunakan *Optimize Parameter Grid*. Dengan hasil eksperimen pada data tenaga kesehatan menunjukkan bahwa iterasi 1-10, dengan percobaan K=2, menghasilkan nilai DBI yang sama, 0,377, yang menunjukkan bahwa cluster terbentuk lebih baik dan terpisah. Dimana pendekatan *Optimalisasi Parameter Grid* digunakan untuk meningkatkan performa dan konvergensi algoritma K-Means untuk mendapatkan nilai iterasi yang diperlukan untuk mencapai hasil yang optimal.

Penelitian ini sejalan dengan penelitian Del Giorgio Solfa & Simonato membahas jumlah cluster (k) dan jumlah iterasi adalah beberapa kombinasi parameter algoritma yang dipelajari secara sistematis dalam teknik optimize parameter grid. Metode ini bertujuan untuk menemukan konfigurasi terbaik yang akan memberikan hasil pengelompokan yang optimal [14]. Jumlah iterasi yang dibutuhkan dapat dipengaruhi oleh berbagai faktor. Penggunaan teknik optimasi lain yang dikombinasikan dengan K-Means, seperti algoritma genetika, *optimization particle swa*rm, dan *optimization parameter grid*, adalah salah satu contohnya. [15]

3.2.3 *Meansure Type* yang menghasilkan nilai DBI paling kecil

Hasil eksperimen data tenaga kesehatan menunjukkan bahwa parameter pengelompokan data terbaik adalah menggunakan jenis meansure Bregman Divergence. Ini karena model K-Means menjadi lebih akurat dan efisien. Setelah mencoba tiga jenis parameter perbedaan perbedaan perbedaan *(Bregman Divergence, Numerical Meansure, Dan Mixed Meansure*). Hasil penelitian ini dapat membantu menyebarluaskan tenaga medis ke wilayah yang paling membutuhkan. Untuk menentukan jumlah cluster (K) yang tepat, daerah diklasifikasikan menurut berbagai hal. Ini termasuk jumlah penduduk, rasio tenaga kesehatan, tingkat penyakit, dan akses ke fasilitas kesehatan. Hasilnya membantu menentukan wilayah mana yang paling membutuhkan tenaga kesehatan. Model K-Means dapat menemukan pola dalam data dan mengelompokkan daerah dengan kebutuhan yang sama karena teknik pencarian grid membantu menemukan parameter yang ideal. Metode ini memastikan bahwa sumber daya kesehatan dialokasikan dengan tepat, sehingga mengurangi kesenjangan layanan.

Penelitian ini sejalan dengan penelitian Idrus membahas bahwa penggunaan *Bregman Divergence* yang telah digeneralisasi menunjukkan kinerja pengelompokan yang lebih unggul berdasarkan pengukuran *Davies-Bouldin Index* (DBI) [16].

4. KESIMPULAN

Dengan menggunakan algoritma K-Means, penelitian ini menunjukkan bahwa *Optimize Parameter Grid* meningkatkan akurasi klasterisasi dan mengurangi kebutuhan pemilihan parameter manual. Maka, hasil menunjukkan kualitas dari pengelompokkan yaitu nilai K optimal sebesar 2, dan nilai Davies Bouldin Index (DBI) sebesar 0,377. Jumlah iterasi yang ideal untuk menghasilkan clustering yang optimal adalah antara 1-10. Untuk meningkatkan efisiensi, metode Optimalisasi *Parameter Grid* digunakan. Selain itu, untuk pengelompokan yang lebih akurat, perbedaan Bregman adalah metrik terbaik yang digunakan. Diharapkan dalam proses pengelompokkan pada penelitian lebih lanjut harus membandingkan kinerja algoritma K-Means dengan algoritma clustering lainnya seperti DBSCAN, *Mean-Shift*, atau *Hierarchical Clustering*. Supaya memberikan wawasan lebih luas mengenai metode yang paling sesuai untuk data tenaga kesehatan.

DAFTAR PUSAKA

[1] H. Zhong, G. Loukides, and R. Gwadera, "Clustering datasets with demographics and diagnosis codes," *J. Biomed. Inform.*, vol. 102, no. January, 2020, doi: 10.1016/j.jbi.2019.103360.

[2] V. Mhasawade, Y. Zhao, and R. Chunara, "Machine learning and algorithmic fairness in public and population health," *Nat. Mach. Intell.*, vol. 3, no. 8, pp. 659–666, 2021, doi: 10.1038/s42256-021-00373-4.

[3] R. Hammad, M. Barhoush, and B. H. Abed-Alguni, "A semantic-based approach for managing healthcare big data: A survey," *J. Healthc. Eng.*, vol. 2020, 2020, doi: 10.1155/2020/8865808.

[4] K. Wynter, S. Holton, J. Considine, and A. M. Hutchinson, "Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ' s public news and information ," no. January, 2020.

[5] C. Jongen, J. McCalman, S. Campbell, and R. Fagan, "Working well: Strategies to strengthen the workforce of the Indigenous primary healthcare sector," *BMC Health Serv. Res.*, vol. 19, no. 1, pp. 1–12, 2019, doi: 10.1186/s12913-019-4750-5.

[6] A. C. Cardoso-dos-Santos *et al.*, "The importance of geographic and sociodemographic aspects in the characterization of mucopolysaccharidoses: a case series from Ceará state (Northeast Brazil)," *J. Community Genet.*, pp. 573–580, 2024, doi: 10.1007/s12687-024-00718-7.

[7] J. Yue *et al.*, "Evaluating the accessibility to healthcare facilities under the chinese hierarchical diagnosis and treatment system," *Geospat. Health*, vol. 16, no. 2, 2021, doi: 10.4081/gh.2021.995.

[8] D. K. Sitinjak, B. A. Pangestu, and B. N. Sari, "Clustering Tenaga Kesehatan Berdasarkan Kecamatan di Kabupaten Karawang Menggunakan Algoritma K-Means," *J. Appl. Informatics Comput.*, vol. 6, no. 1, pp. 47–54, 2022, doi: 10.30871/jaic.v6i1.3855.

[9] H. Rosyid, R. Mailok, and M. M. Lakulu, "Optimizing k-means initial number of cluster based heuristic approach: Literature review analysis perspective," *Int. J. Artif. Intell.*, vol. 6, no. 2, pp. 120–124, 2019, doi: 10.36079/LAMINTANG.IJAI-0602.40.

[10] M. Rochmawati *et al.*, "Implementasi Algoritma K-Means dalam Klasterisasi Penjualan pada Sebuah Perusahaan menggunakan Metodologi KDD Implementation of the K-Means Algorithm in Sales Clustering at a Company using the KDD Methodology," vol. 13, pp. 54–62, 2024.

[11] F. Leoni, M. Carraro, E. McAuliffe, and S. Maffei, "Data-centric public services as potential source of policy knowledge. Can 'design for policy' help?," *Transform. Gov. People, Process Policy*, vol. 17, no. 3, pp. 399–411, 2023, doi: 10.1108/TG-06-2022-0088.

[12] R. W. Sembiring Brahmana, F. A. Mohammed, and K. Chairuang, "Customer Segmentation Based on RFM Model Using K-Means, K-Medoids, and DBSCAN Methods," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 11, no. 1, p. 32, 2020, doi: 10.24843/lkjiti.2020.v11.i01.p04.

[13] D. H. Khoiriyah and R. Ambarwati, "Dynamic Segmentation Analysis for Expedition Services: Integrating K-Means and Decision Tree," *J. Inf. Syst. Informatics*, vol. 6, no. 1, pp. 363–377, 2024, doi: 10.51519/journalisi.v6i1.666.

[14] F. Del Giorgio Solfa and F. R. Simonato, "Big Data Analytics in Healthcare: Exploring the Role of Machine Learning in Predicting Patient Outcomes and Improving Healthcare Delivery," *Int. J. Comput. Inf. Manuf.*, vol. 3, no. 1, pp. 1–9, 2023, doi: 10.54489/ijcim.v3i1.235.

[15] M. Ezar, A. Rivan, R. A. Sonaru, and K. Kunci-Algoritma, "Perbandingan Metode K-Means Dan GA K-Means Untuk Clustering Dataset Heart Disease Patients hasil intra cluster dari GA K-Means lebih baik dibandingkan dengan K-Means dan untuk inter cluster sangat kecil perbedaannya, dimana rata-rata inter cluster metode ," *J. Tek. Inform. dan Sist. Inf.*, vol. 9, no. 3, p. 2585, 2022, [Online]. Available: http://jurnal.mdp.ac.id

[16] A. Idrus, N. Tarihoran, U. Supriatna, A. Tohir, S. Suwarni, and R. Rahim, "Distance Analysis Measuring for Clustering using K-Means and Davies Bouldin Index Algorithm," *TEM J.*, vol. 11, no. 4, pp. 1871–1876, 2022, doi: 10.18421/TEM114-55.