

Clustering Jumlah Tenaga Kesehatan Berdasarkan Kecamatan di Kabupaten Karawang Menggunakan *Algoritma K-Means*

Desi Kristina Sitinjak^{1*}, Bagus Aji Pangestu^{2*}, Betha Nurina Sari^{3*}

* Teknik Informatika, Universitas Singaperbangsa Karawang

desi.kristina18201@student.unsika.ac.id¹, bagus.pangestu18133@student.unsika.ac.id², betha.nurina@staff.unsika.ac.id³

Article Info

Article history:

Received 2022-01-23

Revised 2022-05-06

Accepted 2022-05-10

Keyword:

Tenaga Kesehatan,
Algoritma K-means,
Data Mining,

ABSTRACT

Pembangunan kesehatan merupakan bagian dari Pembangunan Nasional yang pada hakekatnya adalah penyelenggaraan upaya kesehatan untuk mencapai kemampuan hidup sehat bagi setiap penduduk agar dapat mewujudkan derajat kesehatan yang optimal, masalah kesehatan yang ada pada masyarakat di Indonesia yaitu masih minimnya tenaga kesehatan pada setiap wilayah. Salah satunya di Kabupaten Karawang, Tenaga kesehatan yang tidak tercukupi di beberapa kecamatan yang ada di Karawang akan membuat masyarakat di kecamatan tersebut kesulitan untuk hidup sehat dan mengobati penyakitnya. Penelitian ini bertujuan untuk melakukan pengelompokan terhadap Kecamatan yang memiliki tenaga kesehatan yang masih kurang sehingga data tersebut dapat digunakan untuk peningkatan kualitas kesehatan. Penelitian ini menggunakan metode *K-Means Clustering*. Hasil pengolahan dataset tenaga kesehatan yang ada di Kabupaten Karawang menghasilkan 3 *cluster*, yaitu *cluster* 1 dengan tenaga kesehatan sedikit sebanyak 24 kecamatan, *cluster* 2 dengan tenaga kesehatan sedang sebanyak 4 kecamatan dan *cluster* 3 dengan tenaga kesehatan terbanyak yaitu 2 kecamatan.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

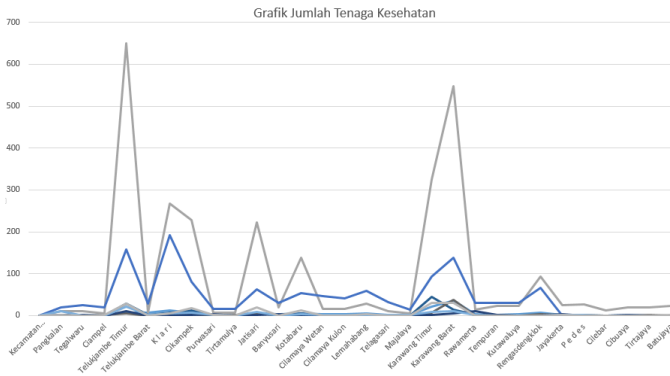
Pembangunan kesehatan sebagai bagian integral dari Pembangunan Nasional pada hakekatnya adalah penyelenggaraan upaya kesehatan untuk mencapai kemampuan hidup sehat bagi setiap penduduk agar dapat mewujudkan derajat kesehatan yang optimal dan besar artinya bagi pengembangan dan pembinaan sumber daya manusia sebagai modal Pembangunan Nasional. Perencanaan kebutuhan tenaga kesehatan secara nasional disesuaikan dengan masalah kesehatan, kemampuan daya serap dan kebutuhan pengembangan program pembangunan kesehatan. Pengadaan tenaga kesehatan sesuai dengan perencanaan kebutuhan tersebut diselenggarakan melalui pendidikan dan pelatihan baik oleh Pemerintah dan atau oleh masyarakat termasuk swasta sedangkan pendayagunaannya diselenggarakan secara efektif dan merata [1].

Pembangunan di sektor kesehatan merupakan salah satu upaya pemenuhan hak dasar rakyat, yaitu hak untuk

memperoleh pelayanan kesehatan. Hal ini menjelaskan bahwa pemerintah memiliki tugas dan tanggung jawab untuk mencari dan melaksanakan upaya-upaya pembangunan di sektor kesehatan sehingga tuntutan UUD 1945 dapat terpenuhi dan tujuan pembangunan nasional [2] [3] di sektor kesehatan yaitu “meningkatkan derajat kesehatan masyarakat yang setinggi-tingginya” dapat tercapai.

Puskesmas merupakan fasilitas pelayanan kesehatan yang menyelenggarakan upaya kesehatan masyarakat dan upaya perseorangan tingkat pertama dengan mengutamakan upaya *promotif* dan *preventif* untuk mencapai derajat kesehatan masyarakat. Tenaga kesehatan yang tidak tercukupi di beberapa Kecamatan yang ada di Karawang akan membuat masyarakat di Kecamatan tersebut kesulitan untuk hidup sehat dan mengobati penyakitnya. Seperti contoh di saat pandemi awal jumlah pasien yang sakit lebih banyak dari pada tenaga kesehatan yang menangani, dampaknya tenaga kesehatan lelah dan angka covidnya naik kembali. Jumlah

tenaga kesehatan di Kabupaten Karawang dapat dilihat pada Gambar 1.



Gambar 1. Grafik Jumlah Tenaga Kesehatan (Sumber: karawangkab.bps.go.id)

Penelitian ini bertujuan untuk melakukan pengelompokan jumlah tenaga kesehatan yang ada di Kabupaten Karawang. Diharapkan data tersebut dapat digunakan sebagai evaluasi Dinas Kesehatan Kabupaten Karawang. Pengelompokan ini di ambil berdasarkan kecamatan [3] di Kabupaten Karawang untuk mengetahui jumlah tenaga kesehatannya, jumlah tenaga kesehatan berdampak pada kualitas pelayanan kepada masyarakat, semakin lengkap tenaga kesehatan maka pelayanan atau kepuasan masyarakat akan lebih baik. Clustering dipilih untuk bisa mengelompokkan jumlah tenaga kesehatan di kecamatan mana yang sudah lengkap, sedang, dan kurang lengkap, dalam penelitian ini menggunakan algoritma K-Means [4].

II. TINJAUAN PUSTAKA

A. Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan dalam database [5]. Data mining merupakan proses menemukan korelasi baru yang bermanfaat, pola dengan menambang sejumlah repositori data dalam jumlah besar, menggunakan teknologi pengenalan pola seperti statistik dan teknik matematika. Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (Machine Learning) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*), secara otomatis [6]. Dari beberapa pemaparan tentang data mining maka dapat dikatakan bahwa data mining merupakan sebuah proses mengolah suatu data menjadi informasi. Data mining mempunyai fungsi yang penting untuk membantu mendapatkan informasi yang berguna serta meningkatkan pengetahuan bagi pengguna.

Teknik data mining biasanya terbagi dalam dua kategori, prediksi dan deskripsi. Teknik prediksi menggunakan data historis untuk menyimpulkan sesuatu tentang kejadian di masa depan. Sedangkan teknik deskripsi bertujuan untuk menemukan pola dalam data yang menyediakan beberapa informasi tentang hubungan interval yang tersembunyi [4].

B. Clustering

Analisis Pengelompokan/*Clustering* merupakan proses membagi data dalam suatu himpunan ke dalam beberapa kelompok yang kesamaan datanya dalam suatu kelompok lebih besar daripada kesamaan data tersebut dengan data dalam kelompok lain [7]. Analisis *cluster* merupakan pengelompokan objek-objek data hanya berdasarkan pada informasi yang terdapat pada data, yang menjelaskan objek dan relasinya [8]. Tujuan utama dari metode *clustering* adalah pengelompokan sejumlah data atau objek ke dalam *cluster* atau grup sehingga dalam setiap *cluster* akan berisi data yang semirip mungkin. *Clustering* melakukan pengelompokkan data yang didasarkan pada kesamaan antar objek, oleh karena itu klusterisasi digolongkan sebagai metode *unsupervised learning* [9] [10].

C. Algoritma K-Means

Algoritma K-Means merupakan metode yang digunakan untuk mengelompokkan suatu data yang dilakukan berdasarkan pencarian pusat cluster secara iteratif dimana letak cluster masing-masing data ditemukan jarak minimum setiap data pada pusat cluster [8]. Algoritma K-Means termasuk dalam metode nonhierarchial yang mempartisi data ke dalam bentuk satu atau lebih cluster, sehingga data yang mempunyai karakteristik yang sama dikelompokkan dalam satu cluster yang sama dan data yang memiliki karakteristik berbeda dikelompokkan ke dalam cluster lain [3]. Algoritma K-Means ini dalam penelitian mereka untuk meningkatkan kecepatan metode clustering dengan cara membuat penentuan centroid yang lebih sederhana [11]. K-means memiliki beberapa keuntungan yang berbeda dibandingkan dengan algoritma pengelompokan lainnya. Artinya, K-means merupakan algoritma yang sangat sederhana dan kuat, sangat efisien, dan dapat digunakan untuk berbagai macam jenis data. Untuk menghitung pusat cluster terbaru menggunakan rumus jarak Euclidean.

$$d(x, y) = \sqrt{\sum_{i=0}^n (xi - yi)^2}$$

Dimana:

d(x,y) = jarak antara data pada titik x dan y

x = Titik data objek

y = titik data *centroid*

i = jumlah atribut data

Rumus pengelompokan data ke dalam *cluster* dengan jarak terdekat.

$$C_k = (\frac{1}{nk}) \sum d_i$$

Dimana:

Ck = *centroid cluster*

nk = jumlah data dalam *cluster* K

di = jumlah dari nilai jarak yang masuk dalam masing masing *cluster*.

D. Davies-Bouldin Index (DBI)

Davies-Bouldin Index (DBI) adalah salah satu metode validitas internal dalam melakukan evaluasi terhadap suatu cluster [11]. Metrik DBI diperkenalkan oleh David L Davies dan Donald W. Bouldin (1979), Validitas Internal yang dilakukan DBI adalah seberapa baik cluster sudah dilakukan dengan menghitung kuantitas dan fitur turunan dari set data. Sum of square within cluster (SSW) sebagai metrik kohesi dalam sebuah cluster ke- i diformulasikan pada persamaan

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i)$$

Dimana:

SSW = Sum of square within cluster

m_i = Jumlah data dalam cluster i

(x_j, c_i) = Jarak data x ke centroid c

Nilai $d()$ dalam persamaan (2.2) bisa menggunakan formula ketidakmiripan (jarak) yang digunakan ketika proses pengelompokannya sehingga validasi yang diberikan juga mempunyai maksud yang sama terhadap proses pengelompokannya. Sementara metrik untuk separasi antara dua cluster digunakan formula Sum of Square Between Cluster (SSB) dengan mengukur jarak antar centroid c_i dan c_j seperti pada persamaan berikut.

$$SSB_{i,j} = d(c_i, c_j)$$

Dimana:

SSB = Sum Of Square Between Cluster

(x_j, c_i) = Jarak centroid c_i dengan centroid c_j

Langkah selanjutnya adalah menghitung nilai $R_{i,j}$ adalah ukuran rasio seberapa baik nilai perbandingan antara cluster ke- i dan cluster ke- j . Nilainya 21 didapatkan dari komponen kohesi dan separasi. Cluster yang baik adalah yang mempunyai kohesi yang sekecil mungkin dan separasi yang sebesar mungkin. $R_{i,j}$ di formulasikan dalam persamaan berikut.

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}}$$

Dimana:

$R_{i,j}$ = Nilai Perbandingan cluster i dengan cluster j .

Setelah kita mendapatkan nilai SSW, SSB dan R . Nilai Davies-Bouldin Index (DBI) dapat dihitung menggunakan formula dalam persamaan berikut.

$$DBI = \frac{1}{K} \sum_{j=1}^K \max(R_{i,j})$$

Dimana:

K = Jumlah cluster yang digunakan

Dari formula diatas dapat kita amati bahwa semakin rendah nilai DBI maka semakin baik cluster yang didapatkan, artinya kemiripan antar data dalam satu cluster akan semakin mirip. DBI banyak digunakan untuk membantu clustering berbasis

non hierarki seperti *K-means* atau *K-Modes* untuk menentukan berapa jumlah cluster yang tepat untuk digunakan.

III. METODE PENELITIAN

Metode analisis data yang digunakan pada penelitian ini adalah *Knowledge Discovery Database* (KDD). Adapun tahapan dari metode Knowledge Discovery in Database yang digunakan pada penelitian ini adalah sebagai berikut.

1) Data Selection

Data selection merupakan salah satu proses dimana pengambilan data/pengumpulan data, pemilihan data dan pengecekan data.

2) Pre-Processing

Tahap *pre-processing* merupakan salah satu tahap pembersihan data dengan membuang duplikasi data, mengatasi missing value, memeriksa data yang tidak konsisten.

3) Data Transformation

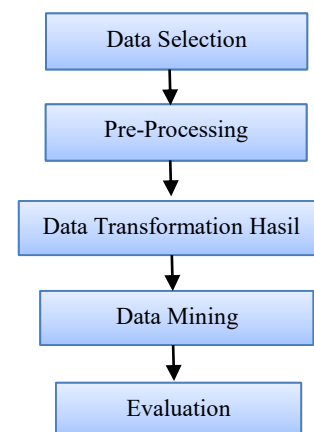
Tahap ini data akan dilakukan transformasi sebelum dilakukannya data mining. Hal tersebut bertujuan supaya data dapat menyesuaikan pada saat diolah berdasarkan algoritma dan tools yang akan digunakan dalam pengolahan data.

4) Data Mining

Data mining merupakan salah satu proses pengolahan data berdasarkan algoritma sesuai dengan teknik data mining. Algoritma yang digunakan pada penelitian ini adalah algoritma *k-means* yang merupakan salah algoritma *clustering*.

5) Knowledge Interpretation/Evaluation.

Tahap ini merupakan proses dari penginterpretasian dan evaluasi hasil dari data mining. Sehingga hasil yang didapat menjadi informasi.



Gambar 2 Alur Penelitian

IV. HASIL DAN PEMBAHASAN

1) Data Selection

Data yang diseleksi yaitu data tenaga kesehatan yang ada di Kabupaten Karawang yang diperoleh dari (karawangkab.bps.go.id). Data tersebut terdiri dari 30 data kecamatan dan akan diambil 3 variabel yang akan diujikan (Dokter, Perawat, dan Bidan). Data yang akan digunakan pada penelitian ini ditunjukkan pada tabel 1:

TABEL 1
JUMLAH TENAGA KESEHATAN DI SETIAP KECAMATAN
PADA KABUPATEN KARAWANG

No	Kecamatan	Dokter Umum	Perawat	Bidan
1	Pangkalan	7	11	20
2	Tegalwaru	3	10	25
3	Ciampel	2	6	19
4	Telukjambe Timur	191	650	158
5	Telukjambe Barat	7	8	28
6	Klari	76	267	192
7	Cikampek	86	228	80
8	Purwasari	2	7	17
9	Tirtamulya	2	7	17
10	Jatisari	65	222	63
11	Banyusari	5	20	31
12	Kotabaru	30	139	54
13	Cilamaya Wetan	5	17	46
14	Cilamaya Kulon	6	17	41
15	Lemahabang	36	29	60
16	Telagasari	3	11	33
17	Majalaya	2	6	14
18	Karawang Timur	162	325	93
19	Karawang Barat	172	547	138
20	Rawamerta	5	14	30
21	Tempuran	5	24	31
22	Kutawaluya	5	23	31
23	Rengasdengklok	44	94	66
24	Jayakarta	6	25	0
25	Pedes	9	27	0
26	Cilebar	2	13	0
27	Cibuaya	3	19	0
28	Tirtajaya	0	19	0
29	Batujaya	5	23	0
30	Pakisjaya	3	13	0

2) Data Pre-Processing

Setelah dilakukan data selection atau pemilihan data kemudian dilakukan penanganan *missing value*, duplikasi data dan data yang tidak sesuai atau *inkonsisten* data. Dari dataset dilakukannya proses pembersihan data atau pemilihan

data, dimana hanya data-data tertentu yang akan digunakan pada proses data mining.

3) Data Transformation

Pada tahap ini data ditransformasikan dan atribut-atribut yang akan digunakan Tabel atribut yang digunakan pada dataset jumlah tenaga kesehatan di Kabupaten Karawang dapat dilihat di tabel 2:

TABEL 2
ATRIBUT DATASET

Atribut	Tipe Data	Keterangan
Kecamatan	Polynomial	Nama Kecamatan
Dokter	Integer	Jumlah Dokter
Dokter Gigi	Integer	Jumlah Dokter Gigi
Perawat	Integer	Jumlah Perawat
Bidan	Integer	Jumlah Bidan
Tenaga Kefarmasian	Integer	Jumlah Tenaga Kefarmasian
Tenaga Kesehatan Masyarakat	Integer	Jumlah Tenaga Kesehatan Masyarakat
Tenaga Kesehatan Lingkungan	Integer	Jumlah Tenaga Kesehatan Lingkungan
Tenaga Gizi	Integer	Jumlah Tenaga Gizi
Ahli Teknologi Laboratorium Medik	Integer	Jumlah Ahli Teknologi Laboratorium Medik

4) Data Mining

Dataset yang telah diperoleh, selanjutnya diproses datanya. Proses *clustering* pada data ini menggunakan metode *K-Means Clustering*. Dalam pengolahan datasetnya dilakukan 2 kali iterasi atau perulangan, karena pada iterasi kedua galatnya sudah bernilai 0 atau sudah sama dengan iterasi sebelumnya. Langkah pertama dalam metode *K-Means Clustering* yaitu menentukan jumlah *cluster* terlebih dahulu, yaitu sebanyak 3 cluster. *Cluster* yang ditentukan yaitu untuk *cluster* dengan jumlah tenaga kesehatan yang sedikit (C1), *cluster* dengan jumlah tenaga kesehatan yang sedang (C2), *cluster* dengan jumlah tenaga kesehatan yang banyak (C3). Selanjutnya menentukan pusat *cluster* atau *centroid* dengan acak. Penentuan pusat *cluster* sebagai berikut.

- *Centroid* 1, diambil dari kecamatan Cilebar dengan *variable* Dokter, Perawat, dan Bidan (2, 13, 0).
- *Centroid* 2, diambil dari kecamatan Cikampek dengan *variable* Dokter, Perawat, dan Bidan (86, 228, 80).
- *Centroid* 3, diambil dari kecamatan Telukjambe Timur dengan *variable* Dokter, Perawat, dan Bidan (191, 650, 158).

Selanjutnya menghitung jarak antara data dengan pusat *cluster* awal menggunakan persamaan *Euclidean distance*. Kemudian mengelompokkan data ke dalam *cluster* berdasarkan data yang mempunyai jarak yang paling minim

akan masuk ke dalam *cluster* 1, 2, atau 3. Jarak data ke pusat *cluster* pada iterasi 1 dapat dilihat pada tabel 3.

TABEL 3
JARAK DATA KE PUSAT CLUSTER PADA ITERASI 1

No	Kecamatan	dc1	dc2	dc3	class
1	Pangkalan	23.4946	239.6622	680.449	1
2	Tegalwaru	25.298	240.8111	681.284	1
3	Ciampel	20.3960	246.1219	687.009	1
4	Telukjambe Timur	684.654	442.565	0	3
5	Telukjambe Barat	29.748	240.497	681.950	1
6	Klari	327.333	119.920	403.022	2
7	Cikampek	245.479	0.00000	442.565	2
8	Purwasari	18.466	245.656	686.378	1
9	Tirtamulya	18.055	245.845	686.542	1
10	Jatisari	228.411	28.460	457.040	2
11	Banyusari	32.372	229.682	670.677	1
12	Kotabaru	140.730	108.761	546.879	2
13	Cilamaya Wetan	46.540	229.660	670.783	1
14	Cilamaya Kulon	41.593	230.1477	671.423	1
15	Lemahabang	71.182	207.0338	648.987	1
16	Telagasari	33.271	238.153	679.312	1
17	Majalaya	15.842	247.457	688.041	1
18	Karawang Timur	367.586	129.622	333.775	2
19	Karawang Barat	580.496	580.496	338.594	3
20	Rawamerta	32.062	235.414	676.408	1
21	Tempuran	33.120	226.068	666.931	1
22	Kutawaluya	32.954	226.993	667.910	1
23	Rengasdeng klock	112.898	142.961	584.311	1
24	Jayakarta	13.379	233.463	672.247	1
25	Pedes	15.716	230.848	669.673	1
26	Cilebar	0	245.479	684.654	1
27	Cibuaya	6.4807	239.837	678.763	1
28	Tirtajaya	6.4031	240.937	679.628	1
29	Batujaya	10.448	235.648	674.484	1
30	Pakisjaya	2.8284	245.002	684.2967	1

Selanjutnya mengelompokan data ke dalam *cluster*, data yang mempunyai nilai yang minim atau kecil akan masuk ke dalam *cluster* yang tersedia. Hasil pengelompokan *cluster* dapat dilihat pada tabel 4.

TABEL 4
PENGELOMPOKAN CLUSTER PADA ITERASI 1

No	Kecamatan	C1	C2	C3
1	Pangkalan	X		
2	Tegalwaru	X		
3	Ciampel	X		
4	Telukjambe Timur			X
5	Telukjambe Barat	X		
6	Klari		X	
7	Cikampek		X	
8	Purwasari	X		
9	Tirtamulya	X		
10	Jatisari		X	
11	Banyusari	X		
12	Kotabaru	X		
13	Cilamaya Wetan	X		
14	Cilamaya Kulon	X		
15	Lemahabang	X		
16	Telagasari	X		
17	Majalaya	X		
18	Karawang Timur		X	
19	Karawang Barat			X
20	Rawamerta	X		
21	Tempuran	X		
22	Kutawaluya	X		
23	Rengasdengklok	X		
24	Jayakarta	X		
25	Pedes	X		
26	Cilebar	X		
27	Cibuaya	X		
28	Tirtajaya	X		
29	Batujaya	X		
30	Pakisjaya	X		

Langkah Selanjutnya hitung pusat cluster terbaru, dari hasil pengelompokan iterasi pertama kemudian hitung rata-ratanya dari setiap cluster pada masing-masing variable. Pada tahap ini hasilnya akan menjadi pusat cluster pada iterasi kedua, seperti yang terlihat pada tabel 5 berikut.

TABEL 5
JARAK DATA KE PUSAT CLUSTER PADA ITERASI 2

No	Kecamatan	dc1	dc2	dc3
1	Pangkalan	13.45587201	251.158595313798	627.8913123
2	Tegalwaru	10.68601545	251.895692698387	629.1251863
3	Ciampel	14.71525577	257.476678555554	634.2779359
4	Telukjambe Timur	672.4638795	432.587840790746	57.94393842
5	Telukjambe Barat	13.57650032	251.414478501140	629.1156491

6	Klari	308.4934917	101.782316735 276	352.6946271
7	Cikampek	231.5823068	19.0955492196 48	389.7402981
8	Purwasari	14.46042772	257.125728000 914	633.697483
9	Tirtamulya	14.61147748	257.332936096 412	633.8647332
10	Jatisari	215.8623143	42.1691830606 19	404.3173259
11	Banyusari	9.606137767	240.516610652 986	617.9000728
12	Kotabaru	126.6898629	119.586119595 880	494.26764
13	Cilamaya Wetan	24.168243	239.574289104 653	618.0481373
14	Cilamaya Kulon	19.16562968	240.390182827 835	618.6885323
15	Lemahabang	48.80546624	216.746026491 837	596.0364922
16	Telagasari	14.39412819	248.792363226 848	626.572821
17	Majalaya	16.45466936	259.072267909 940	635.3546254
18	Karawang Timur	354.6704773	123.269785430 169	281.0524862
19	Karawang Barat	568.0307857	328.641202529 446	57.94393842
20	Rawamerta	13.21460211	246.274318596 154	623.6974427
21	Tempuran	10.43278966	236.949024053 698	614.2674499
22	Kutawaluya	10.16043328	237.845832420 919	615.1889954
23	Rengasdengklok	94.30625211	152.688702921 991	531.3365224
24	Jayakarta	23.01170903	246.078117678 106	619.5865557
25	Pedes	23.68396201	243.580869527 966	617.0668521
26	Cilebar	23.79659393	257.877955630 178	632.0652656
27	Cibuaya	22.58355626	252.301089969 901	626.1481454
28	Tirtajaya	23.48856457	253.344113805 709	627.0546228
29	Batujaya	22.77335524	248.237466954 527	621.9087554
30	Pakisjaya	23.42647267	257.406759817 997	631.6268677

Selanjutnya mengelompokkan *cluster* berdasarkan nilai yang minim atau paling kecil, seperti yang sudah dijelaskan sebelumnya, hasilnya dapat dilihat pada tabel 6.

TABEL 6
PENGELOMPOKAN CLUSTER PADA ITERASI 2

No	Kecamatan	C1	C2	C3
1	Pangkalan	X		
2	Tegalwaru	X		
3	Ciampel	X		
4	Telukjambe Timur			X
5	Telukjambe Barat	X		

6	Klari		X	
7	Cikampek		X	
8	Purwasari	X		
9	Tirtamulya	X		
10	Jatisari		X	
11	Banyusari	X		
12	Kotabaru	X		
13	Cilamaya Wetan	X		
14	Cilamaya Kulon	X		
15	Lemahabang	X		
16	Telagasari	X		
17	Majalaya	X		
18	Karawang Timur		X	
19	Karawang Barat			X
20	Rawamerta	X		
21	Tempuran	X		
22	Kutawaluya	X		
23	Rengasdengklok	X		
24	Jayakarta	X		
25	Pedes	X		
26	Cilebar	X		
27	Cibuaya	X		
28	Tirtajaya	X		
29	Batujaya	X		
30	Pakisjaya	X		

Proses *K-Means clustering* akan terus melakukan iterasi hingga data hasil *clustering* akan sama dengan hasil iterasi sebelumnya [12]. Pada iterasi kedua hasil pengelompokan data sama dengan hasil pengelompokan iterasi pertama, maka proses pengolahan dataset berhenti sampai disini. Hasil pengolahan dari dataset didapatkan *cluster* 1 sebanyak 24 kecamatan, *cluster* 2 sebanyak 4 kecamatan dan *cluster* 3 sebanyak 2 kecamatan.

Implementasi *K-Means* dengan *Rapidminer* Pengolahan dataset jumlah tenaga kesehatan di Kab. Karawang telah didapatkan, lalu dilakukan pengolahan dataset kembali menggunakan *Rapid Miner* untuk perbandingan. Hasil akhir yang didapatkan dari hasil model *K-Means* ditunjukkan oleh Gambar 3.

Cluster Model

```
Cluster 0: 24 items
Cluster 1: 2 items
Cluster 2: 4 items
Total number of items: 30
```

Gambar 3 Hasil Model *K-Means* dengan *Rapidminer*

Hasil pengolahan data Jumlah Tenaga kesehatan di Kabupaten Karawang dengan *Algoritma K-Means* yang terbagi menjadi 3 *cluster* yaitu *cluster* 0 sebanyak 24

kecamatan, *cluster* 1 sebanyak 2 kecamatan dan *cluster* 3 sebanyak 4 kecamatan .

Berdasarkan hasil pengolahan data baik menggunakan *Microsoft excel* maupun *RapidMiner*, menghasilkan *cluster* yang sama pada setiap kelompoknya dengan *Cluster* 1 dengan jumlah tenaga kesehatan rendah sebanyak 24 yaitu, Pangkalan, Tegalwaru, Ciampel, Telukjambe Barat, Purwasari, Tirtamulya, Banyusari, Kotabaru, Cilamaya Wetan, Cilamaya Kulon, Lemahabang, Telagasari, Majalaya, Rawamerta, Tempuran, Kutawulya, Rengasdengklok, Jayakarta, Pedes, Cilebar, Cibuya, Tirtajaya, Batujaya, Pakisjaya. Selanjutnya, *Cluster* 2 dengan jumlah tenaga kesehatan sedang sebanyak 4 yaitu Klari, Cikampek, Jatisari, Karawang Timur. Sedangkan untuk *Cluster* 3 dengan jumlah tenaga kesehatan terbanyak adalah 2 yaitu, Telukjambe Timur dan Karawang Barat.

5) Knowledge Interpretation/Evaluation

Parameter dalam mengukur kinerja atau akurasi algoritma *K-means* dilakukan dengan menghitung *Davies Bouldin Index (DBI)* dan *Avg. within centroid distance*. DBI adalah algoritma yang memiliki hasil *cluster* berdasarkan jarak *inter-cluster* rendah dan jarak antar *cluster* tinggi akan mempunyai nilai DBI rendah. Kumpulan *cluster* dengan nilai DBI terendah akan dianggap sebagai algoritma terbaik.

Sedangkan *Avg. within centroid distance* adalah rata-rata dalam jarak *cluster* yang dihitung berdasarkan rata-rata jarak antara titik pusat *cluster* dan semua contoh *cluster*. Hasil nilai DBI dan *Avg. within centroid distance* dapat dilihat pada Gambar 4.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -2041.839
Avg. within centroid distance_cluster_0: -1372.184
Avg. within centroid distance_cluster_1: -2842.500
Avg. within centroid distance_cluster_2: -5659.438
Davies Bouldin: -0.370
```

Gambar 4 Hasil DBI dan Avg. Within Centroid Distance

TABEL 7
HASIL CENTROID AKHIR

Atribut	Cluster 0	Cluster 1	Cluster 2
Total	51.625	619.1667	5057
Jumlah	999	394	15

IV. KESIMPULAN

Penerapan algoritma *K-Means Clustering* menggunakan *tools Rapidminer*, menghasilkan 3 *cluster* dengan nilai DBI sebesar -0.370, yaitu dengan *cluster* 1 dengan jumlah tenaga kesehatan rendah sebanyak 24 yaitu, Pangkalan, Tegalwaru,

Ciampel, Telukjambe Barat, Purwasari, Tirtamulya, Banyusari, Kotabaru, Cilamaya Wetan, Cilamaya Kulon, Lemahabang, Telagasari, Majalaya, Rawamerta, Tempuran, Kutawulya, Rengasdengklok, Jayakarta, Pedes, Cilebar, Cibuya, Tirtajaya, Batujaya, Pakisjaya. Selanjutnya, *Cluster* 2 dengan jumlah tenaga kesehatan sedang sebanyak 4 yaitu Klari, Cikampek, Jatisari, Karawang Timur. Sedangkan untuk *Cluster* 3 dengan jumlah tenaga kesehatan terbanyak adalah 2 yaitu, Telukjambe Timur dan Karawang Barat.

Dalam hasil tersebut masih ada beberapa jumlah tenaga kesehatann di Kabupaten Karawang yang jumlahnya masih rendah. Dengan demikian hal tersebut dapat dijadikan sebagai acuan bagi pemerintah untuk mengevaluasi dalam menentukan jumlah tenaga kesehatan yang sesuai standar agar masyarakat dapat terjamin kesehatannya, dapat sejahtera kehidupannya. Untuk penelitian lebih lanjut dapat melakukan pengoptimalan dengan teknik *preprocessing* data untuk menghasilkan hasil yang berkualitas dan menggunakan algoritma yang lain agar mendapatkan perbandingan.

UCAPAN TERIMA KASIH

Terimakasih kepada program studi Teknik Informatika, Universitas Singaperbangsa Karawang atas dukungan dalam terlaksananya penelitian ini, serta kepada BPS Kabupaten Karawang yang telah mengizinkan meneliti data yang ada sebagai objek penelitian.

DAFTAR PUSTAKA

- [1] K. K. Republik Indonesia, 'Peraturan Pemerintah Republik Indonesia no 32 tahun 1996 tentang tenaga kesehatan'. <https://kemkes.go.id/> (accessed May 10, 2022).
- [2] A. Ariyanto, D. E. Kurniawan, and A. Fatulloh, 'Rancang Bangun Aplikasi WebGIS untuk Pemetaan Kondisi Sosial Ekonomi Kota Batam', *J. Appl. Inform. Comput. JAIC*, vol. 2, no. 1, Art. no. 1, 2018, doi: 10.30871/jaic.v2i1.904.
- [3] D. E. Kurniawan and A. Fatulloh, 'Clustering of Social Conditions in Batam, Indonesia Using K-Means Algorithm and Geographic Information System', *Int. J. Earth Sci. Eng. IJEE*, vol. 10, no. 5, pp. 1076–1080, 2017.
- [4] F. Nasution, 'Penerapan Metode K-Means Clustering untuk Mengelompokkan Ketahanan Tanaman Pangan Kabupaten/Kota di Provinsi Sumatera Utara', 2019
- [5] Y. Fadhillah, M. N. H. Siregar, and O. Siagian, 'Perancangan Aplikasi Data Mining Untuk Menentukan Pasien Menderita Tifoid Dengan Metode Algoritma C4.5', *Explorer (Hayward)*, vol. 1, no. 2, pp. 63–70, 2021.
- [6] M. S. Mustafa, M. R. Ramadhan, and A. P. Thenata, 'Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier', *Creat. Inf. Technol. J.*, vol. 4, no. 2, pp. 151–162, 2018.
- [7] L. Maulida, 'Penerapan Data Mining dalam Mengelompokkan Kunjungan Wisatawan ke Objek Wisata Unggulan di Prov. DKI Jakarta dengan K-Means', *JISKA J. Inform. Sunan Kalijaga*, vol. 2, no. 3, pp. 167–174, 2018.
- [8] M. G. Sadewo, A. Eriza, A. P. Windarto, and D. Hartama, 'Algoritma K-Means Dalam Mengelompokkan Desa/Kelurahan Menurut Keberadaan Keluarga Pengguna Listrik dan Sumber Penerangan Jalan Utama Berdasarkan Provinsi', 2019, vol. 1, no. 1.
- [9] N. Y. S. Munti, 'Analisis Dan Penerapan Data Mining Untuk Menentukan Gaji Karyawan Tetap Dan Karyawan Kontrak Menggunakan Algoritma K-Means Clustering', *J. Inov. Tek. Inform.*, vol. 1, no. 1, pp. 1–11, 2018.

- [10] P. Prasetyawan, I. Ahmad, R. I. Borman, Y. A. Pahlevi, and D. E. Kurniawan, 'Classification of the Period Undergraduate Study Using Back-propagation Neural Network', 2018, pp. 1–5.
- [11] F. Mahmuda, M. A. R. Sitorus, H. Widyastuti, and D. E. Kurniawan, 'Clustering Profil Pengunjung Perpustakaan Menggunakan Algoritma K-Means', *J. Appl. Inform. Comput.*, vol. 1, no. 1, Art. no. 1, 2017, doi: 10.30871/jaic.v1i1.476.
- [12] R. Prasajo, Y. R. W. Utami, and R. T. Vlandari, 'Implementasi K-Means Clustering Pada Pengelompokan Potensi Kerjasama Pelanggan', *J. Teknol. Inf. Dan Komun. TIKomSiN*, vol. 7, no. 2, Art. no. 2, Dec. 2019, doi: 10.30646/tikomsin.v7i2.435.