

Clustering Fasilitas Kesehatan Berdasarkan Kecamatan Di Karawang Dengan Algoritma K-Means

Bagus Muhammad Islami^{1,*}, Cepy Sukmayadi¹, Tesa Nur Padilah¹

¹:¹ Ilmu Komputer; Universitas Singaperbangsa Karawang; Jl. HS. Ronggo Waluyo, Teluk Jambe Timur, Karawang, Jawa Barat, Indonesia – 41361; telp/fax: 0812-95094228; e-mail: bagus.muhammad17065@student.unsika.ac.id, cepy.sukmayadi17009@student.unsika.ac.id, tesa.nurpadilah@staff.unsika.ac.id

* Korespondensi: e-mail: bagus.muhammad17065@student.unsika.ac.id

Diterima: 20 Mei 2021; Review: 14 Juni 2021; Disetujui: 25 Juni 2021

Cara sitasi: Islami BM, Sukmayadi C, Padilah TS. 2021. Clustering Fasilitas Kesehatan Berdasarkan Kecamatan Di Karawang Dengan Algoritma K-Means. Bina Insani ICT Journal. Vol. 8 (1): 83-92.

Abstrak: Masalah kesehatan yang ada di dalam masyarakat terutama di negara-negara berkembang seperti Indonesia dipengaruhi oleh dua faktor yaitu aspek fisik dan aspek non fisik. Berdasarkan data yang diperoleh dari karawangkab.bps.go.id data dibagi menjadi 3 cluster yaitu sedikit, sedang dan terbanyak. Algoritma yang digunakan adalah K-Means cluster yang diimplementasikan menggunakan Microsoft Excel dan Rapidminer Studio. Hasil pengolahan data fasilitas kesehatan di karawang menghasilkan 3 cluster dengan cluster 1 yang mempunyai fasilitas kesehatan sedikit sebanyak 23 kecamatan, cluster 2 yang mempunyai fasilitas kesehatan sedang sebanyak 5 kecamatan dan cluster 3 yang mempunyai fasilitas kesehatan terbanyak terdapat 2 kecamatan. Kinerja yang dihasilkan dari algoritma K-means menghasilkan nilai Davies Boildin Index sebesar 0,109.

Kata kunci: *clustering*, data mining, fasilitas kesehatan, K-Means.

Abstract: Health problems that exist in society, especially in developing countries like Indonesia, are built by two factors, namely physical and non-physical aspects. Based on data obtained from karawangkab.bps.go.id the data is divided into 3 clusters, namely the least, medium and the most. The algorithm used is the K-Means cluster which is implemented using Microsoft Excel and Rapidminer Studio. The results of data processing of health facilities in Karawang produce 3 clusters with cluster 1 which has 23 sub-districts of health facilities, cluster 2 which has medium health facilities as many as 5 districts and cluster 3 which has the most health facilities in 2 districts. The performance resulting from the K-means algorithm results in a Davies Boildin Index value of 0.109.

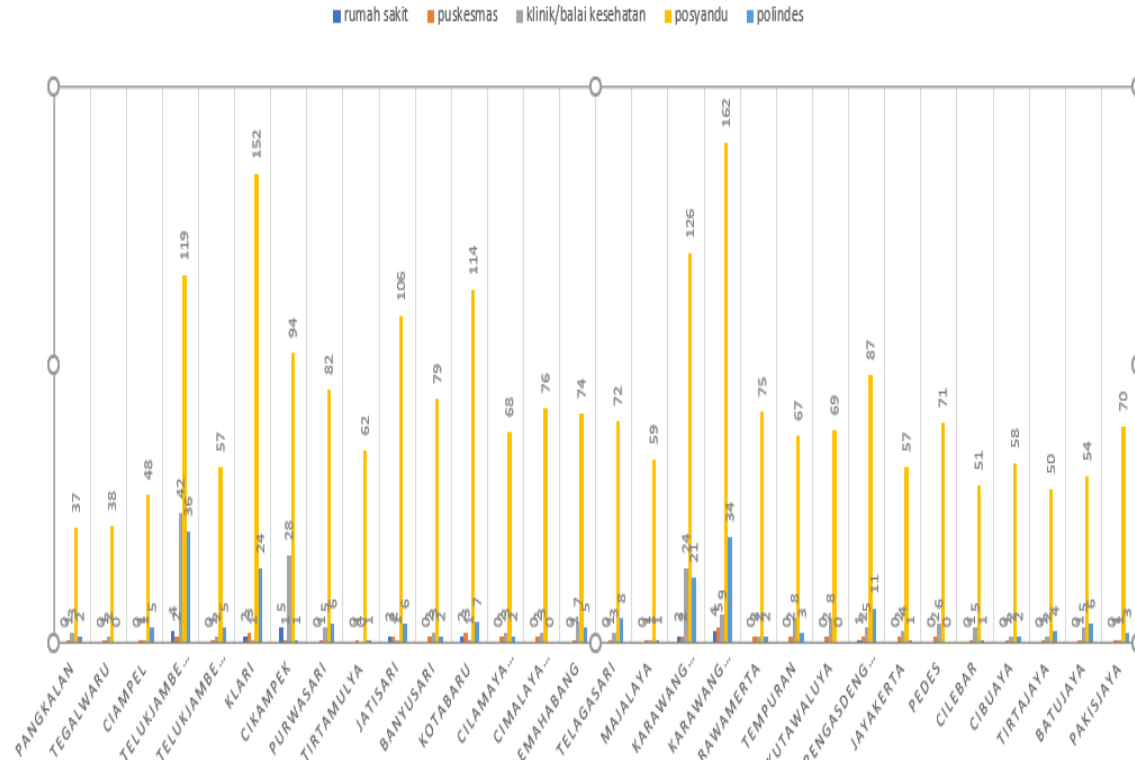
Keywords: *clustering*, data mining, health facilities, K-Means.

1. Pendahuluan

Masalah kesehatan yang ada di dalam masyarakat terutama di negara-negara berkembang seperti Indonesia dipengaruhi oleh dua faktor yaitu aspek fisik dan aspek non fisik. Aspek fisik seperti sarana kesehatan dan pengobatan penyakit, sedangkan yang kedua adalah aspek non fisik yang menyangkut masalah kesehatan seperti perilaku Manusia dalam bentuk pengetahuan, sikap dan tindakan [1]. Pelayanan kesehatan merupakan hak dasar masyarakat yang harus dipenuhi dalam pembangunan kesehatan [2]. Pembangunan kesehatan diselenggarakan untuk meningkatkan kesadaran, kemauan, dan kemampuan hidup sehat bagi setiap orang agar terwujud derajat kesehatan masyarakat yang setinggi-tingginya.

Fasilitas kesehatan yang ada di Indonesia sangat perlu diperhatikan oleh pemerintah khususnya di kecamatan yang ada di Karawang ini masih tergolong kurang. Contohnya seperti

puskesmas, rumah sakit, klinik, maupun tempat pengobatan lainnya. Sarana 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 jika ada ibu hamil yang ingin bersalin akan kesulitan apabila sarannya tidak memadai. Oleh sebab itu, pemerintah diharapkan dapat memperbaiki kondisi sarana kesehatan yang kurang di Kabupaten Karawang. Penelitian ini bertujuan untuk melakukan pengelompokan terhadap Kecamatan yang memiliki sarana kesehatan dengan cepat dan efektif.



Sumber: karawangkab.bps.go.id (2021)

Gambar 1. Jumlah Fasilitas Kesehatan Di Kabupaten Karawang

Clustering merupakan sebuah metode pengelompokan suatu obyek kedalam sejumlah kelompok (*cluster*) yang sesuai. Prinsip dari *clustering* adalah memaksimalkan kesamaan antar anggota satu *cluster* dan meminimumkan kesamaan antar anggota *cluster* yang berbeda. Analisis *cluster* berfungsi sebagai pemisah obyek kedalam beberapa kelompok yang memiliki perbedaan karakteristik antar kelompok. K-Means merupakan teknik pengelompokan *hard partition* yang efisien dalam mengelompokkan data besar, namun terbatas hanya pada data numerik [3].

2. Metode Penelitian

Penelitian ini menggunakan dataset fasilitas kesehatan di kabupaten karawang. Dataset tersebut kemudian dikelompokkan menggunakan algoritma K-Means. Pengelompokan Fasilitas Kesehatan dengan algoritma k-means ini dilakukan dengan beberapa tahapan, yaitu: a) Pengumpulan data, tahapan ini dilakukan untuk mencari data-data yang akan digunakan dalam penelitian terkait clustering fasilitas Kesehatan data diambil dari situs resmi Badan Pusat Statistik Karawang (karawangkab.bps.go.id), b) Pengolahan dataset menggunakan *microsoft excel*, merupakan tahap pengolahan data menggunakan algoritma K-Means untuk mendapatkan kelompok data fasilitas kesehatan. Aplikasi yang digunakan tahapan ini adalah *microsoft excel*, c) Implementasi K-Means *cluster* dengan *Rapidminer*, sebagai bahan perbandingan pengolahan data yang sudah dilakukan menggunakan *microsoft excel* juga dilakukan pengolahan data kembali menggunakan aplikasi *rapidminer studio*, d) Analisa hasil pengolahan data, berdasarkan hasil pengolahan data yang telah dilakukan menggunakan *microsoft excel* dan *rapidminer studio*, kemudian data tersebut dianalisa dalam bentuk *cluster model* dan *performance vector* atas *clustering* fasilitas kesehatan.



Sumber: Hasil Penelitian (2021)

Gambar 2. Alur penelitian

Data Mining

Kegiatan mencari knowledge terpendam dari himpunan data yang amat banyak adalah definisi dari data mining [8]. Output atau hasil dari proses data mining bisa dijadikan bahan untuk menentukan langkah di waktu yang akan datang [9]. Berdasarkan tugasnya data mining terbagi menjadi 6 metode yaitu description, estimation, prediction, classification, clustering, serta association [9]. Data mining berfungsi untuk menambang pengetahuan berharga yang tersembunyi di dalam sebuah data [10].

Clustering

Clustering merupakan salah satu teknik data mining yang digunakan untuk mendapatkan kelompok-kelompok dari objek objek yang mempunyai karakteristik yang umum di data yang cukup besar [4]. Cluster berbeda dari klasifikasi karena cluster tidak memiliki variabel target. Tujuan cluster bukan untuk mengklasifikasikan, memperkirakan, atau memprediksi nilai variabel target [1].

K-Means Clustering

K-means merupakan suatu algoritma pengklasteran yang cukup sederhana yang mempartisi dataset kedalam beberapa kluster k [5]. Tahapan yang dilakukan dalam algoritma K-Means adalah sebagai berikut [6]: Tentukan jumlah *cluster* dan tetapkan secara acak pusat *cluster*. Hitung jarak data dengan pusat *cluster*.

$$d(x, y) = \sqrt{\sum_{i=0}^n (x_i - y_i)^2} \quad (1)$$

Kelompokan data ke dalam *cluster* dengan jarak terdekat. Perbarui nilai pusat *cluster*.

$$C_k = \left(\frac{1}{n_k}\right) \sum d_i \quad (2)$$

Ulangi langkah 3 dan 5 hingga data pada tiap cluster tidak berubah.

Davies-Bouldin Index (DBI)

DBI Merupakan salah satu uji validitas internal pada metode. Pengelompokan basis partisi berdasarkan jumlah dari kedekatan data terhadap *centroid* dari *cluster* yang diikutinya dan diantara dua *cluster* diukur dengan kedekatan dua centroid cluster. DBI bertujuan untuk memaksimalkan jarak *inter-cluster*, Dengan menggunakan *Davies Bouldin Index* suatu cluster akan dianggap memiliki skema *clustering* yang optimal [11] semakin kecil nilai DBI maka semakin baik pula hasil dari cluster tersebut.

Min-Max Normalization

Salah satu dari banyak metode pada normalisasi data adalah metode *min-max normalization*. metode ini mengubah nilai pada dataset kedalam skala atau range yang lebih kecil yang bernilai 0 sampai dengan 1 [12]. Berikut merupakan persamaan pada *min-max normalization*.

$$d = \frac{d - \min(p)}{\max(p) - \min(p)} \quad (3)$$

3. Hasil dan Pembahasan

Dataset yang digunakan pada penelitian ini adalah data fasilitas kesehatan di Kabupaten Karawang. Data diambil dari situs resmi Badan Pusat Statistik Kabupaten Karawang (karawangkab.bps.go.id). Data tersebut terdiri dari 30 kecamatan dengan 5 variabel. Berikut data dari BPS Karawang yang ditampilkan pada tabel 1.

Tabel 1. Data Fasilitas Kesehatan yang ada di Kabupaten Karawang

No	Kabupaten	Rumah Sakit	Puskesmas	Balai Kesehatan	Posyandu	Polindes	Apotek
1	Pangkalan	0	1	3	37	2	4
2	Tegalwaru	0	1	2	38	0	1
3	Ciampel	0	1	1	48	5	3
4	Telukjambe Timur	4	2	42	119	36	8
5	Telukjambe Barat	0	1	2	57	5	3
6	Klari	2	3	1	152	24	9
7	Cikampek	5	1	28	94	1	6
8	Purwasari	0	1	5	82	6	4
9	Tirtamulya	0	1	0	62	1	2
10	Jatisari	2	2	1	106	6	3
11	Banyusari	0	2	3	79	2	3
12	Kotabaru	2	3	1	114	7	8
13	Cilamaya Wetan	0	2	3	68	2	9
14	Cilamaya Kulon	0	2	3	76	0	5
15	Lemahabang	0	1	7	74	5	6
16	Telagasari	0	1	3	72	8	6
17	Majalaya	0	1	1	59	1	2
18	Karawang Timur	2	2	24	126	21	7
19	Karawang Barat	4	5	9	162	34	6
20	Rawamerta	0	2	2	75	2	3
21	Tempuran	0	2	8	67	3	6
22	Kutawaluya	0	2	8	69	0	0
23	Rengasdengklok	1	2	5	87	11	5
24	Jayakarta	0	2	4	57	1	1
25	Pedes	0	2	6	71	0	4
26	Cilebar	0	1	5	51	1	2
27	Cibuaya	0	1	2	58	2	2
28	Tirtajaya	0	1	2	50	4	3
29	Batujaya	0	1	5	54	6	3
30	Pakisjaya	0	1	1	70	3	1

Sumber: karawangkab.bps.go.id (2021)

Pengolahan Dataset

Dataset yang telah didapatkan kemudian dilakukan pengolahan untuk dikelompokkan datanya. Clustering data ini menggunakan algoritma K-Means. Proses pengolahan dataset menggunakan Microsoft Excel untuk mendapatkan hasil cluster dilakukan dengan dua kali perulangan atau iterasi, karena pada iterasi kedua hasil clustering data sama dengan hasil iterasi pertama dengan: 1) Menentukan jumlah *cluster*, jumlah *cluster* yang digunakan pada penelitian ini berjumlah 3 *cluster* yang terdiri dari *cluster* 1 (C1) merupakan kecamatan dengan fasilitas kesehatan sedikit, *cluster* 2 (C2) merupakan kecamatan dengan fasilitas sedang dan *cluster* 3 (C3) merupakan kecamatan dengan fasilitas kesehatan terbanyak. Pusat cluster iterasi pertama meliputi: a) *Cluster* 1 (sedikit) yang terdiri dari: Rumah Sakit=0, Puskesmas=1, Balai Kesehatan=1, Posyandu=48, Polindes=5, dan Apotek=3, b) *Cluster* 2 (sedang) yang terdiri dari: Rumah Sakit=2, Puskesmas=2, Balai Kesehatan=24, Posyandu=126, Polindes=21, dan Apotek=7, dan c) *Cluster* 3 (banyak) yang terdiri dari: Rumah Sakit=4, Puskesmas=5, Balai Kesehatan=9, Posyandu=162, Polindes=34, dan Apotek=6; 2) Menghitung jarak setiap data ke pusat *cluster*. Jarak setiap data ke pusat cluster ditentukan dengan rumus *Euclidian Distance* dijelaskan pada tabel 2.

Tabel 2. Jarak Data ke Pusat *Cluster* Iterasi Pertama

No	Kabupaten	C1	C2	C3	Jarak Terdekat
1	Pangkalan	135	8735	16709	135
2	Tegalwaru	130	8708	16626	130
3	Ciampel	0	6888	13930	0

No	Kabupaten	C1	C2	C3	Jarak Terdekat
4	Telukjambe Timur	7713	601	2955	601
5	Telukjambe Barat	82	5520	11944	82
6	Klari	11219	1219	279	279
7	Cikampek	2875	1445	6091	1445
8	Purwasari	1174	2534	7224	1174
9	Tirtamulya	214	5100	11206	214
10	Jatisari	3368	1170	4004	1170
11	Banyusari	975	3029	7971	975
12	Kotabaru	4391	871	3107	871
13	Cilamaya Wetan	450	4172	9918	450
14	Cimalaya Kulon	818	3388	8602	818
15	Lemahabang	721	3253	8609	721
16	Telagasari	598	3530	8832	598
17	Majalaya	138	5446	11798	138
18	Karawang Timur	6888	0	1702	0
19	Karawang Barat	13930	1702	0	0
20	Rawamerta	740	3464	8664	740
21	Tempuran	424	4064	10000	424
22	Kutawaluya	525	3997	9855	525
23	Rengasdengklok	1579	1987	6183	1579
24	Jayakarta	111	5599	12177	111
25	Pedes	581	3801	9463	581
26	Cilebar	42	6414	13462	42
27	Cibuaya	111	5497	11925	111
28	Tirtajaya	6	6568	13522	6
29	Batujaya	53	5789	12493	53
30	Pakisjaya	492	4028	9534	492

Sumber: Hasil Penelitian (2021)

3) Pengelompokan data ke dalam *cluster*, data yang mempunyai jarak terdekat akan masuk kedalam cluster tersebut yang dijelaskan pada tabel 3.

Tabel 3. Pengelompokan Data Iterasi Pertama

No	Kabupaten	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	Cimalaya 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		

Sumber: Hasil Penelitian (2021)

4) Hitung pusat cluster yang baru, hasil pengelompokan iterasi pertama kemudian dihitung rata-rata dari setiap cluster pada masing-masing variabel. Hasilnya dijadikan pusat *cluster* pada iterasi kedua.

Iterasi Kedua: 1) Menentukan jumlah *cluster*, hasil perhitungan pusat cluster yang baru pada iterasi pertama dijadikan sebagai pusat cluster pada iterasi kedua. Pusat *cluster* iterasi kedua meliputi: a) *Cluster* 1 (sedikit) yang terdiri dari: Rumah Sakit=0,043, Puskesmas=1,391, Balai Kesehatan=3,522, Posyandu=63,522, Polindes=3,043, dan Apotek=3,391, b) *Cluster* 2 (sedang) yang terdiri dari: Rumah Sakit=3, Puskesmas=2, Balai Kesehatan=19,2, Posyandu=111,8, Polindes=14,2, dan Apotek=6,4, dan c) *Cluster* 3 (banyak) yang terdiri dari: Rumah Sakit=3, Puskesmas=4, Balai Kesehatan=5, Posyandu=157, Polindes=29, dan Apotek=7,5; 2) Menghitung jumlah setiap data ke pusat *cluster*, jarak setiap data ke pusat cluster ditentukan dengan rumus *Euclidian Distance*.

Tabel 4. Jarak Data ke Pusat *Cluster* Iterasi Kedua

No	Kabupaten	C1	C2	C3	Jarak Terdekat
1	Pangkalan	705,331	6016,08	15157,25	705,331
2	Tegalwaru	668,853	5977,08	15065,25	668,853
3	Ciampel	251,461	4501,88	12505,25	251,461
4	Telukjambe Timur	5670,113	1050,48	2867,25	1050,48
5	Telukjambe Barat	49,026	3399,08	10617,25	49,026
6	Klari	8309,94	2052,08	70,25	70,25
7	Cikampek	1544,2	571,68	5295,25	571,68
8	Purwasari	352,94	1166,68	6178,25	352,94
9	Tirtamulya	21,026	3046,28	9876,25	21,026
10	Jatisari	1821,983	444,68	3171,25	444,68
11	Banyusari	241,505	1501,68	6844,25	241,505
12	Kotabaru	2595,853	392,48	2351,25	392,48
13	Cilamaya Wetan	53,287	2339,48	8663,25	53,287
14	Cimalaya Kulon	168,244	1750,68	7419,25	168,244
15	Lemahabang	132,722	1666,48	7483,25	132,722
16	Telagasari	103,722	1889,08	7684,25	103,722
17	Majalaya	33,113	3316,68	10446,25	33,113
18	Karawang Timur	4660,679	272,28	1391,25	272,28
19	Karawang Barat	10710,07	3026,28	70,25	70,25
20	Rawamerta	135,722	1813,48	7489,25	135,722
21	Tempuran	39,374	2261,08	8794,25	39,374
22	Kutawaluya	71,244	2202,88	8657,25	71,244
23	Rengasdengklok	620,635	830,88	5236,25	620,635
24	Jayakarta	53,07	3440,48	10834,25	53,07
25	Pedes	72,113	2049,28	8257,25	72,113
26	Cilebar	165,287	4095,88	12062,25	165,287
27	Cibuaya	36,026	3362,48	10581,25	36,026
28	Tirtajaya	186,418	4234,68	12115,25	186,418
29	Batujaya	101,94	3625,28	11170,25	101,94
30	Pakisjaya	54,244	2237,08	8315,25	54,244

Sumber: Hasil Penelitian (2021)

3) Pengelompokan data ke dalam *cluster* berdasarkan jarak terdekat, data yang mempunyai jarak terdekat akan masuk kedalam *cluster* tersebut dijelaskan pada tabel 5. Kabupaten Pangkalan dikelompokkan pada C1, Tegalwaru dikelompokkan pada C1, Ciampel dikelompokkan pada C1, Telukjambe Timur dikelompokkan pada C2, Telukjambe Barat dikelompokkan pada C1, Klari dikelompokkan pada C3, Cikampek dikelompokkan pada C2, Purwasari dikelompokkan pada C1, Tirtamulya dikelompokkan pada C1, Jatisari dikelompokkan pada C2, Banyusari dikelompokkan pada C1, Kotabari dikelompokkan pada C2, Cilamaya Wetan dikelompokkan pada C1, Cimalaya Kulon dikelompokkan pada C1, Lemahabang dikelompokkan pada C1, Tegalsari dikelompokkan pada C1, Majalaya dikelompokkan pada C1, Karawang Timur dikelompokkan pada C2, Karawang Barat dikelompokkan pada C3, Rawamerta dikelompokkan pada C1, Tempuran dikelompokkan pada C1, Kutawaluya dikelompokkan pada C1, Rengasdengklok dikelompokkan pada C1, Jayakarta dikelompokkan pada C1, Pedes dikelompokkan pada C1, Cilebar dikelompokkan pada C1, Cibuaya dikelompokkan pada C1, Tirtajaya dikelompokkan pada C1, Batujaya dikelompokkan pada C1 dan Pakisjaya dikelompokkan pada C1 yang dijelaskan pada tabel 5.

Tabel 5. Pengelompokan Data Iterasi Kedua

No	Kabupaten	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	Cimalaya 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		

Sumber: Hasil Penelitian (2021)

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 23 kecamatan, cluster 2 sebanyak 5 kecamatan dan cluster 3 sebanyak 2 kecamatan.

K-Means

Clustering memiliki konsep dimana Elemen data K dipilih sebagai center, kemudian jarak semua data elemen dihitung menggunakan rumus Euclidean. Elemen data yang memiliki jarak kurang ke centroid dipindahkan ke cluster yang sesuai. Proses dilanjutkan hingga tidak ada lagi perubahan yang terjadi dalam cluster [k-1] [14]. Metode ini merupakan metode pengelompokan nonhirarki yang bertujuan untuk mengelompokkan objek sehingga jarak-jarak tiap objek ke pusat kelompok di dalam satu kelompok bernilai minimum [13]. Data diproses dengan menggunakan algoritma k-means clustering. Adapun proses yang dilewati pada algoritma K-means adalah sebagai berikut: 1. Menentukan jumlah cluster yang akan digunakan. Penentuan cluster atau K merupakan hal pertama yang harus dilakukan pada algoritma K-Means, dimana nilai K merupakan bilangan bulat positif. Untuk penelitian ini nilai K yang digunakan adalah cluster K = 2 sampai dengan K = 10. 2. Kemudian menentukan centroid, penentuan centroid dilakukan dengan memberikan nilai secara acak pada tiap atribut 3. Menghitung jarak setiap data dengan titik pusat setiap cluster centroid awal. Hal ini untuk mengetahui cluster mana yang paling dekat dengan data. 4. Menghitung kembali pusat cluster yang baru berdasarkan rata-rata anggota yang terdapat pada cluster tersebut setelah semua data ditempatkan ke dalam cluster yang terdekat, pada tahap ini iterasi terus dilakukan jika centroid belum konvergen dan iterasi hanya dapat dihentikan jika centroid yang baru bersifat konvergen dengan centroid yang lama. Pada penelitian ini, implementasi algoritma k-means menggunakan alat bantu/tools RapidMiner. Percobaan dilakukan pada data dengan menggunakan nilai K = 2 sampai dengan K = 10. Kemudian dievaluasi dengan menggunakan uji validasi clustering untuk mengidentifikasi cluster terbaik.

Implementasi K-Means dengan *Rapidminer*

Pengolahan dataset fasilitas kesehatan di Kabupaten Karawang telah didapatkan, kemudian akan dilakukan pengolahan dataset kembali menggunakan Rapidminer sebagai bahan perbandingan. Hasil yang didapatkan berdasarkan Rapidminer.

Cluster Model

```
Cluster 0: 23 items
Cluster 1: 2 items
Cluster 2: 5 items
Total number of items: 30
```

Sumber: Hasil Penelitian (2021)

Gambar 3. Hasil *Cluster* Algoritma K-Means

Gambar 3 menunjukkan hasil pengolahan data fasilitas kesehatan di Kabupaten Karawang dengan Algoritma K-Means yang terbagi menjadi 3 cluster yaitu cluster 0 sebanyak 23 kabupaten, cluster 1 sebanyak 2 kabupaten dan cluster 3 sebanyak 5 kabupaten yang dijelaskan pada gambar 3.

Analisa Hasil Pengolahan Data

Clustering data fasilitas kesehatan di Kabupaten Karawang dengan algoritma K-Means yang menggunakan *micorosoft excel* maupun *rapidminer* menghasilkan pengelompokan yang sama pada setiap kelompoknya. Berdasarkan hasil pengolahan data didapatkan cluster 1 yang mempunyai fasilitas kesehatan sedikit sebanyak 23 kecamatan yaitu, Pangkalan, Tegalwaru, Ciampel, Telukjambe Barat, Purwasari, Tirtamulya, Banyusari, Cilamaya Wetan, Cilamaya Kulon, Lemahabang, Telagasari, Majalaya, Rawamerta, Tempuran, Kutawaluya, Rengasdengklok, Jayakarta, Pedes, Cilebar, Cibuaya, Tirtajaya, Batujaya, dan Pakisjaya. Cluster 2 yang mempunyai fasilitas kesehatan sedang sebanyak 5 kecamatan yaitu, Telukjambe Timur, Cikampek, Jatisari, Kotabaru Dan Karawang Timur. Cluster 3 yang mempunyai fasilitas kesehatan terbanyak terdapat 2 kecamatan yaitu, karawang barat dan klari. Parameter untuk mengukur kinerja algoritma K-Means dilakukan dengan menghitung *Davies Bouldin Index* dan *Avg. within centroid distance*. *Davies Bouldin Index* merupakan algoritma yang menghasilkan *cluster* dengan jarak *inter-cluster* rendah (kesamaan *inter-cluster* tinggi) dan jarak antar *cluster* tinggi (kemiripan antar *cluster* rendah) akan memiliki *Davies-Bouldin Index* yang rendah, algoritma *cluster* yang menghasilkan kumpulan *cluster* dengan indeks *Davies-Bouldin* terkecil dianggap sebagai algoritma terbaik berdasarkan kriteria ini. Sedangkan *Avg. within centroid distance* merupakan rata-rata dalam jarak cluster dihitung dengan rata-rata jarak antara centroid dan semua contoh *cluster*. Gambar 4 menunjukkan hasil kinerja dari pengelompokan fasilitas kesehatan kabupaten karawang dengan nilai *Davies Bouldin Index* sebesar 0,109 dan *Avg. within centroid distance* sebesar 39,947.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 39.947
Avg. within centroid distance_cluster_0: 31.278
Avg. within centroid distance_cluster_1: 11.708
Avg. within centroid distance_cluster_2: 91.120
Davies Bouldin: 0.109
```

Sumber: Hasil Penelitian (2021)

Gambar 4. Hasil kinerja

Perbandingan nilai Davies Bouldin Index berdasarkan jumlah *cluster* yang ditentukan. Dapat dilihat pada nilai Davies Bouldin Index dengan jumlah cluster 3, nilai yang dihasilkan bukan merupakan yang terbaik. Jumlah cluster terbaik berada pada jumlah k sebanyak 6 dengan Nilai Davies Bouldin Index sebesar 0,081. Untuk K pada Rumah Sakit=2, Puskesmas=3, Balai Kesehatan=4, Posyandu=5, Polindes=6, dan Apotek=7, dan DBI pada pada Rumah Sakit=0,098, Puskesmas=0,109, Balai Kesehatan=0,111, Posyandu=0,107 Polindes=0,081 dan Apotek=0,099.

4. Kesimpulan

Hasil pengolahan dataset fasilitas kesehatan yang ada di kabupaten karawang menghasilkan 3 cluster, yaitu cluster 1 dengan fasilitas kesehatan sedikit sebanyak 23 kecamatan, cluster 2 dengan fasilitas kesehatan sedang sebanyak 5 kecamatan dan cluster 3 dengan fasilitas kesehatan sebanyak 2 kecamatan. Ternyata masih banyak kecamatan yang mempunyai fasilitas kesehatan sedikit di karawang. Hal ini bisa dijadikan acuan bagi pemerintah dalam membangun fasilitas kesehatan agar masyarakat dapat berobat tanpa berpergian jauh dari tempat tinggal. Untuk penelitian lebih lanjut dapat melakukan pengoptimalan dengan teknik preprocessing data untuk menghasilkan hasil yang berkualitas dan menggunakan algoritma yang lain agar mendapatkan perbandingan.

Referensi

- [1] Mayangsarim, Sekar D.S., Solikhun & Irawan. Pengelompokkan Jumlah Desa/Kelurahan yang Memiliki Sarana Kesehatan menurut Provinsi dengan Menggunakan Metode K-Means Cluster. *Konferensi Nasional Teknologi Informasi dan Komputer (KOMIK)*, vol. 3, pp. 370-377, 2019.
- [2] Sadewo & Gading M. Pemanfaatan Algoritma Clustering dalam Mengelompokkan Jumlah Desa/Kelurahan yang Memiliki Sarana Kesehatan Menurut Provinsi dengan K-Means. *Konferensi Nasional Teknologi Informasi dan Komputer (KOMIK)*, vol. 1, pp. 124-131, 2017.
- [3] Han, J., Kamber, M. & Pei, J. Data mining: concepts and techniques Third Edition. Morgan Kaufmann Publishers, San Francisco, 2011.
- [4] Hantoro, Kusdarnowo., Andi S. & Siti A. Pencarian Lokasi Perumahan Berdekatan dengan Fasilitas Kesehatan dan Belanja menggunakan Algoritma K-Means. *Jurnal IKRA-ITH Informatika*, vol 4, pp.25-34, 2020.
- [5] Silitonga, Parasian D.P & Damanik R. Implementasi Algoritma K-Means Clustering pada Analisis Penyebaran Penyakit Pasien Pengguna Badan Penyelenggara Jaminan Sosial Kesehatan (BPS). *Jurnal Teknik Informatika Unika St. Thomas (JTIUST)*, vol. 1, pp. 37-41, 2016.
- [6] Supriyatna A., Carolina I., Widiati W. & Nuraeni C. Rice Productivity Analysis by Province Using K-Means Cluster Algorithm. *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 771, p. 012025, 2020.
- [7] Supriyatna A., Carolina I., Janti S., & Haidir A. Clustering Koridor Transjakarta Berdasarkan Jumlah Penumpang Dengan Algoritma K-Means. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 4, pp. 682-693, 2020.
- [8] R. Setiawan, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus: Politeknik LP3I Jakarta)," *J. Lentera ICT*, vol. 3, no. 1, pp. 76–92, 2016.
- [9] Y. Mardi, "Data Mining: Klasifikasi Menggunakan Algoritma C4.5," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017.
- [10] A. F. Lestari and M. Hafiz, "Penerapan Algoritma Apriori Pada Data Penjualan Barbar Warehouse," *J. INOVTEK Polbeng Seri Inform.*, vol. 5, no. 1, pp. 96–105, 2020.
- [11] Mustofa Z, Suasana IS. Algoritma Clustering K-Medoids pada E-Government Bidang Information and Communication Technology Dalam Penentuan Status EDGI. *Jurnal Teknologi Informasi dan Komunikasi*, vol. 9, no. 1, pp. 1-10, 2018
- [12] Jain Y K, Bhandare S K. Min Max Normalization Based Data Perturbation Method for Privacy Protection. *International Journal of Computer & Communication Technology (IJCCT)*, vol. 3, no. 4, pp. 45-50, 2014.
- [13] Aditya KB, Diyah, Setiawan Y. Sistem Informasi Geografis Pemetaan Faktor-Faktor Yang Mempengaruhi Angka Kematian Ibu (AKI) Dan Angka Kematian Bayi (AKB) Dengan

- Metode K-Means Clustering (Studi Kasus: Provinsi Bengkulu). Jurnal Teknik Informatika, vol. 10, no. 1, pp. 59-66, 2017
- [14] Raval, UR. and Chaita J. Implementing & Improvisation of K-means Clustering Algorithm. International Journal of Computer Science and Mobile Computing, vol.5 no.5, pp. 191-203, 2016.