

Implementasi Metode K-Medoids *Clustering* Dalam Pengelompokan Data Penyakit Alergi Pada Anak

Haryati Ningrum¹, Eka Irawan², Muhammad Ridwan Lubis³

^{1,2}STIKOM Tunas Bangsa, Pematangsiantar, Sumatera Utara, Indonesia

³AMIK Tunas Bangsa, Pematangsiantar, Sumatera Utara, Indonesia

Jln. Sudirman Blok A No. 1-3 Pematangsiantar, Sumatera Utara

¹haryatiningrum123@gmail.com, ²eka.irawan@amiktunasbangsa.ac.id

Abstract

Allergies are an abnormal response from the immune system. People who experience allergies have an immune system that reacts to a substance that is usually harmless in the environment. There are two limitations in this study, namely, seafood allergy and air allergy. In this study, the data used were sourced from the National Statistics Agency in 2011-2019. This study uses data mining techniques in data processing with the k-medoids clustering method. The k-medoids method is a clustering method that functions to split the dataset into several groups. The advantages of this method are able to overcome the weaknesses of the k-means method which is sensitive to outliers. Another advantage of this method is that the results of the clustering process do not depend on the order in which the dataset is entered. This method can be applied to data on the percentage of children affected by allergies by province, so that it can be seen the grouping of provinces based on this data. From this grouping data obtained 3 clusters namely low cluster (2 provinces), medium cluster (30 provinces) and high cluster (2 provinces) from the percentage of allergy immunization under five in each province. It is hoped that this research can provide information to the health department, especially the public health center regarding data grouping of Allergic Diseases in children in Indonesia which has an impact on equity in giving anti-allergic immunization to children in Indonesia.

Keywords: Data Mining, K-Medoids Clustering, Allergy

Abstrak

Alergi merupakan respon abnormal dari sistem kekebalan tubuh. Orang yang mengalami alergi memiliki sistem kekebalan tubuh yang bereaksi terhadap suatu zat yang biasanya tidak berbahaya di lingkungan. Pembatasan pada penelitian ini terdapat dua penyebab alergi yaitu, alergi makanan laut dan alergi udara. Pada penelitian ini data yang digunakan bersumber dari Badan Pusat Statistik Nasional tahun 2011-2019. Penelitian ini menggunakan teknik data mining dalam proses pengolahan data dengan metode k-medoids clustering. Metode k-medoids merupakan metode clustering yang berfungsi untuk memecah dataset menjadi beberapa kelompok. Kelebihan dari metode ini mampu mengatasi kelemahan dari metode k-means yang sensitive terhadap outlier. Kelebihan lain dari metode ini yaitu hasil proses clustering tidak bergantung pada urutan masuk dataset. Metode ini dapat diterapkan pada data persentase anak yang terkena penyakit Alergi berdasarkan provinsi, sehingga dapat diketahui pengelompokan provinsi berdasarkan data tersebut. Dari data pengelompokan tersebut diperoleh 34 provinsi dengan menghasilkan 21 provinsi yakni cluster rendah, cluster sedang sebanyak 12 provinsi, dan cluster tinggi sebanyak 1 provinsi dari persentase imunisasi Alergi pada setiap provinsi. Diharapkan penelitian ini dapat memberikan informasi kepada pihak dinas kesehatan terutama puskesmas tentang pengelompokan data Penyakit Alergi pada anak di Indonesia yang berdampak pada pemerataan dalam pemberian imunisasi anti alergi pada anak di Indonesia.

Kata Kunci: Data Mining, K-Medoids Clustering, Alergi

1. PENDAHULUAN

Penyakit alergi merupakan sistem kekebalan tubuhnya bereaksi secara berlebihan terhadap zat atau benda tertentu yang sebetulnya tidak berbahaya.



Kondisi ini cukup sering dialami oleh anak-anak dan jenisnya bisa bermacam-macam, misalnya alergi kacang, alergi debu, alergi obat, atau alergi susu.

World Health Organization (WHO), sebagai organisasi kesehatan dunia di bawah PBB, memprediksi bahwa tingkat kejadian asma pada tahun 2025 akan mencapai angka 400 juta orang, dan kira-kira 50% diperkirakan akan menderita alergi makanan. Angka kejadian kasus alergi pada anak terus meningkat dalam beberapa tahun terakhir. Suatu riset yang dipublikasikan pada tahun 2019 menunjukkan bahwa sekitar 30–40% orang di seluruh dunia memiliki alergi dan kebanyakan kasus alergi ini ditemukan pada anak-anak. Peningkatan angka kejadian kasus alergi diduga disebabkan oleh beberapa faktor, antara lain keturunan atau riwayat alergi dalam keluarga, pengaruh lingkungan, serta pola makan. Dalam hal ini pemerintah harus berperan aktif dan berkomitmen untuk mengurangi penyakit alergi pada anak.

Oleh karena itu, untuk dapat memutuskan mata rantai penyakit alergi khususnya pada anak diperlukan imunisasi minimal 95%. Untuk bisa mewujudkan hal ini pemerintah memerlukan data yang akurat untuk melakukan pengelompokan data penderita penyakit alergi pada anak. Data diperoleh dari BPS mengenai data persentasi penyakit alergi pada anak berdasarkan provinsi di Indonesia pada tahun 2011-2019. Salah satu cara mengatasi masalah tersebut adalah dengan melakukan pengklasteran terhadap data anak penderita Alergi. Tujuannya adalah untuk mengetahui Provinsi yang paling terbanyak penderita alergi pada anak. Pengelompokan tersebut dapat menggunakan metode pengelompokan dengan algoritma *K-Medoids*.

Berdasarkan uraian diatas banyak cabang kecerdasan buatan dalam ilmu komputer yang dapat menyelesaikan permasalahan tersebut secara kompleks diantaranya sistem pendukung keputusan, sistem pakar, data mining dan lain sebagainya. Beberapa penelitian tentang data mining seperti: [1] menunjukkan bahwa dengan penggunaan 2 *cluster* menghasilkan kelompok data dengan potensi yang dimiliki yaitu *cluster* 1 termasuk dalam potensi tinggi dengan hasil rata-rata *brightness* sebesar 344.470K dengan rata-rata *confidence* 87.08% dan *cluster* 2 masuk dalam potensi sedang dengan hasil rata-rata *brightness* sebesar 318.800K dengan rata-rata *confidence* sebesar 58,73%. Penelitian lainnya oleh (Defiyanti, Jajuli & W 2017). *K-medoids* merupakan salah satu metode pengelompokan dalam data mining yang merupakan bagian dari *partitional clustering*. Metode ini menggunakan objek pada kumpulan objek untuk mewakili sebuah *cluster*. "Kelebihan dari metode ini mampu mengatasi kelemahan dari metode *k-medoids* yang sensitive terhadap *outlier* dan hasil proses *clustering* tidak bergantung pada urutan masuk dataset "[1].

Penerapan Algoritma *K-medoids* dapat membantu dalam mengelompokkan data penderita penyakit alergi pada anak. Dengan demikian pihak Dinas Kesehatan terutama puskesmas di provinsi tersebut dapat memberi perhatian lebih terhadap provinsi yang terbanyak penyakit alergi pada anak. Diharapkan penelitian ini dapat membantu pihak Penyedia Layanan Kesehatan untuk mengurangi jumlah anak-anak yang menderita Alergi Kedepannya.ui.

2. METODOLOGI PENELITIAN

2.1. Pengumpulan Data

Penelitian ini dilakukan dengan mengambil data sebagai Sumber data penelitian yang diperoleh dari Badan Pusat Statistik Nasional dengan situs <https://www.bps.go.id>. Data yang digunakan dalam penelitian ini adalah data persentase Data penyakit alergi pada anak pada setiap provinsi dari tahun 2011-2019 yang terdiri dari 34 provinsi.

Tabel 1. Data Persentase Penyakit Alergi Pada Anak

Provinsi	Penyakit Alergi Pada Anak								
	2011	2012	2013	2014	2015	2016	2017	2018	2019
Aceh	3,13	2,74	3,66	3,38	2,98	2,28	2,09	2,26	1,85
Sumatera Utara	2,47	2,40	2,85	2,36	1,95	1,31	1,18	1,12	1,04
Sumatera Barat	2,79	2,60	3,41	2,97	2,37	1,47	1,29	1,19	1,07
Riau	1,69	1,49	2,19	2,01	1,93	1,16	1,06	0,93	0,79
Jambi	3,93	3,67	3,99	3,57	2,91	1,99	1,92	1,99	1,72
Sumatera Selatan	2,47	2,34	3,00	2,81	2,51	1,71	1,58	1,54	1,35
Bengkulu	4,46	4,15	4,32	3,80	3,18	2,27	2,10	2,25	1,96
Lampung	4,95	4,75	4,44	4,35	3,78	3,12	2,98	3,22	2,79
Kep. Bangka Belitung	4,13	4,12	3,98	3,73	3,27	2,16	2,11	2,34	2,03
Kep. Riau	3,54	2,51	2,16	2,01	1,90	1,23	1,08	1,16	1,12
Dki Jakarta	0,99	0,81	1,10	0,86	0,82	0,43	0,39	0,36	0,32
Jawa Barat	3,56	3,38	3,62	3,39	2,95	1,84	1,78	1,78	1,62
Jawa Tengah	9,36	8,98	8,58	8,45	7,82	6,27	6,18	6,70	6,01
Di Yogyakarta	9,02	8,38	7,76	7,30	6,62	5,12	5,05	5,41	4,91
Jawa Timur	10,99	10,53	10,32	9,65	8,88	7,77	7,70	8,41	7,47
Banten	3,56	3,40	3,35	3,07	3,05	2,51	2,33	2,45	2,21
Bali	11,52	10,51	9,73	8,81	8,28	6,70	6,52	7,18	6,41
Nusa Tenggara Barat	17,20	16,51	14,56	14,14	13,34	11,36	11,34	12,94	11,47
Nusa Tenggara Timur	10,34	9,84	10,80	9,70	8,37	7,62	7,27	8,48	7,25
Kalimantan Barat	9,06	8,57	8,76	7,79	7,69	6,79	6,71	7,61	6,75
Kalimantan Tengah	2,32	2,22	2,83	2,25	1,88	1,10	0,99	1,03	0,92
Kalimantan Selatan	4,10	3,66	3,87	3,18	2,66	1,65	1,61	1,72	1,48
Kalimantan Timur	2,82	2,64	2,74	2,19	2,26	1,33	1,18	1,18	0,96
Kalimantan Utara	0,00	0,00	0,00	0,00	0,00	0,00	4,31	4,95	4,37
Sulawesi Utara	0,73	0,65	1,09	1,10	0,89	0,41	0,34	0,21	0,27
Sulawesi Tengah	3,75	3,50	5,03	4,58	3,71	2,76	2,40	2,49	2,17
Sulawesi Selatan	11,33	10,84	10,52	9,90	8,68	7,73	7,67	8,48	7,47
Sulawesi Tenggara	7,34	7,10	7,59	7,38	6,44	5,24	5,09	5,75	4,95
Gorontalo	3,82	3,61	4,75	4,33	3,07	2,08	1,62	1,56	1,53
Sulawesi Barat	10,81	10,09	11,01	9,69	8,07	6,83	6,25	7,25	6,29
Maluku	2,23	2,21	3,06	2,62	1,98	1,12	1,01	1,06	0,81
Maluku Utara	3,78	3,48	3,65	3,12	2,34	1,54	1,39	1,33	1,22
Papua Barat	6,40	4,41	6,99	4,80	4,31	3,37	2,83	2,95	2,62
Papua	27,77	29,59	33,82	31,76	30,87	27,35	26,81	28,98	24,20

2.2. Metode Penelitian

2.2.1. Data Mining

Data Mining merupakan proses analitik yang dirancang untuk memeriksa sejumlah data yang besar dalam mencari suatu pengetahuan tersembunyi yang berharga dan konsisten.” [3]. Tujuan utama Data mining adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau

informasi yang kita miliki. *Data Mining* sering juga disebut Knowledge discovery in database (KDD) merupakan meliputi kegiatan pengumpulan, pemakaian data historis menemukan pola atau hubungan dalam set data berukuran besar. "Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan." [4].

2.2.2. *Clustering*

Clustering merupakan proses partisi sekumpulan objek data dari satu set menjadi beberapa kelas." [4] Hal ini dapat dilakukan dengan menerapkan berbagai persamaan dan langkah-langkah mengenai jarak algoritma yaitu dengan Euclidean Distance. Analisis kluster ialah metode yang dipakai untuk membagi rangkaian data menjadi beberapa grup berdasarkan kesamaan-kesamaan yang telah ditentukan sebelumnya. Dalam menentukan cluster berdasarkan data yang telah tersedia dibutuhkan sebuah flowchart untuk memudahkan dalam menentukan alur perhitungan sebagai alur untuk menemukan hasil dari penerapan cluster terhadap data yang akan diproses.

2.2.3. Algoritma *K-Medoids*

Menurut Listiyanti, Youanda Alim Syahbana & Henim (2016) "Algoritma k-medoids adalah algoritma pengelompokan yang berkaitan dengan algoritma k-medoids dan algoritma medoidshift". Algoritma k-medoids dikembangkan oleh Leonard Kaufman dan Peter J. Rousseeuw pada tahun 1987. Algoritma k-medoids sering disebut juga algoritma Partitioning Around Medoids (PAM)[6]. Algoritma k-medoids memiliki kesamaan dengan algoritma k-medoids yaitu sama-sama termasuk algoritma partitioning. Algoritma partitioning merupakan algoritma pengelompokan data ke dalam sejumlah cluster tanpa adanya struktur hirarki antara satu dengan yang lainnya. Algoritma k-medoids memiliki keunggulan dibandingkan dengan algoritma k-medoids. "K-medoids memiliki kinerja yang lebih optimal jika jumlah data yang digunakan berjumlah sedikit. Algoritma ini menggunakan objek pada kumpulan objek untuk mewakili sebuah cluster, Objek yang terpilih untuk mewakili sebuah cluster disebut medoids" [7]. Beberapa langkah-langkah dalam perhitungan algoritma *k-medoids* adalah:

- a) Inisialisasi pusat *cluster* sebanyak k (jumlah *cluster*)
- b) Alokasikan setiap data (objek) ke *cluster* terdekat menggunakan persamaan ukuran jarak *Euclidian Distance* dengan persamaan:
$$d_{ij} = \sqrt{\sum_{a=1}^p (x_{ia} - x_{ja})^2} = \sqrt{(x_i - x_j)'(x_i - x_j)} \quad (1)$$
dimana $i=1, \dots, n$; $j=1, \dots, n$ dan p adalah banyak variable, serta V adalah matrik varian kovarian.
- c) Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat *medoids* baru.
- d) Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat *medoids* baru.

- e) Hitung total simpangan (S) dengan menghitung nilai total $distance$ baru – total $distance$ lama. Jika $S < 0$, maka tukar objek dengan data $cluster$ untuk membentuk sekumpulan k objek baru sebagai $medoids$.
- f) Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan $medoids$, sehingga didapatkan $cluster$ beserta anggota $cluster$ masing-masing

3. HASIL DAN PEMBAHASAN

3.1. Perhitungan Menggunakan Algoritma *K-Medoids*

Langkah-langkah metode *k-medoids clustering* diantaranya:

a) Inisialisasi pusat $cluster$ sebanyak k (jumlah $cluster$)

Untuk Tabel Medoid awal asumsikan *centroid* awal yang telah ditentukan seperti pada tabel 2. berikut

Tabel 2. Tabel Medoid Awal

Nama	Provinsi	2011	2011	2013	2014	2015	2016	2017	2018	2019
C1	Sulawesi Tengah	3,75	3,50	5,03	4,58	3,71	2,76	2,40	2,49	2,17
C2	Sulawesi Selatan	11,33	10,84	10,52	9,90	8,68	7,73	7,67	8,48	7,47
C3	Sulawesi Tenggara	7,34	7,10	7,59	7,38	6,44	5,24	5,09	5,75	4,95

b) Alokasikan setiap data (objek) ke $cluster$ terdekat menggunakan persamaan ukuran jarak *Euclidian Distance* dengan persamaan:

dimana $i=1, \dots, n$; $j=1, \dots, n$ dan p adalah banyak variable, serta V adalah matrik varian kovarian.

c) Menghitung nilai jarak terdekat (*cost*) dengan persamaan *Euclidian Distance*.

$$d_{Aceh,C1} = \sqrt{(3,13 - 3,75)^2 + (2,74 - 3,50)^2 + (3,66 - 5,03)^2 + (3,38 - 4,58)^2 + (2,98 - 3,71)^2 + (2,28 - 2,76)^2 + (2,09 - 2,40)^2 + (2,26 - 2,49)^2 + (1,85 - 2,17)^2} = 2,30$$

$$d_{Aceh,C2} = \sqrt{(3,13 - 11,33)^2 + (2,74 - 10,84)^2 + (3,66 - 10,52)^2 + (3,38 - 9,90)^2 + (2,98 - 8,68)^2 + (2,28 - 7,73)^2 + (2,09 - 7,67)^2 + (2,26 - 8,48)^2 + (1,85 - 7,47)^2} = 19,65$$

$$d_{Aceh,C3} = \sqrt{(3,13 - 7,34)^2 + (2,74 - 7,10)^2 + (3,66 - 7,59)^2 + (3,38 - 7,38)^2 + (2,98 - 6,44)^2 + (2,28 - 5,24)^2 + (2,09 - 5,09)^2 + (2,26 - 5,75)^2 + (1,85 - 4,95)^2} = 10,94$$

Hasil dari keseluruhan dapat dilihat pada tabel 3 sebagai berikut :

Tabel 3. Hasil Perhitungan Algoritma *K-Medoids* Iterasi ke-1

No	Data Ke I	Jarak ke medoids			Terdekat	Klaster yang diikuti
		C1	C2	C3		
1	Aceh	2,30	19,65	10,94	2,30	1
2	Sumatera Utara	4,73	22,13	13,46	4,73	1
3	Sumatera Barat	3,81	21,29	12,61	3,81	1
4	Riau	5,90	23,37	14,67	5,90	1
5	Jambi	2,02	19,06	10,42	2,02	1
6	Sumatera Selatan	3,88	21,32	12,62	3,88	1
7	Bengkulu	1,66	18,11	9,49	1,66	1

No	Data Ke I	Jarak ke medoids			Terdekat	Klaster yang diikuti
		C1	C2	C3		
8	Lampung	2,19	16,20	7,58	2,19	1
9	Kep. Bangka Belitung	1,74	18,34	9,70	1,74	1
10	Kep. Riau	5,11	22,09	13,50	5,11	1
:	:	:	:	:	:	:
:	:	:	:	:	:	:
34	Papua	77,14	59,89	68,36	59,89	2
Jumlah		327,51	581,98	416,70		
Total Cost		1326,18				

Setelah didapatkan hasil jarak dari setiap objek (*cost*) pada iterasi ke-1 maka lanjut ke iterasi ke-2. Kandidat *medoid* baru(*non-medoid*) pada iterasi ke-2 dapat dilihat pada tabel sebagai berikut:

Tabel 4. Tabel Medoids Baru

Nama	Provinsi	2009	2010	2011	2012	2013	2014	2015	2016	2017
D1	Gorontalo	3,82	3,61	4,75	4,33	3,07	2,08	1,62	1,56	1,53
D2	Sulawesi Barat	10,81	10,09	11,01	9,69	8,07	6,83	6,25	7,25	6,29
D3	Maluku	2,23	2,21	3,06	2,62	1,98	1,12	1,01	1,06	0,81

Hitung kembali jarak dari setiap objek pada iterasi ke-2 dengan menggunakan *medoid* baru. Maka didapatkan hasil keseluruhannya dari iterasi ke-2 dapat dilihat pada tabel 3.4. sebagai berikut:

Tabel 5. Hasil Perhitungan Algoritma *K-Medoids* Iterasi ke-2

No	Data Ke I	Jarak ke medoids			Terdekat	Klaster yang diikuti
		C1	C2	C3		
1	Aceh	2,05	17,76	2,84	2,05	1
2	Sumatera Utara	3,64	20,19	0,57	0,57	3
3	Sumatera Barat	2,65	19,32	1,07	1,07	3
4	Riau	4,94	21,47	1,40	1,40	3
5	Jambi	1,23	17,11	3,31	1,23	1
6	Sumatera Selatan	3,05	19,42	1,26	1,26	3
7	Bengkulu	1,45	16,15	4,29	1,45	1
8	Lampung	3,23	14,35	6,20	3,23	1
9	Kep. Bangka Belitung	1,57	16,41	4,05	1,57	1
10	Kep. Riau	4,01	20,19	1,77	1,77	3
:	:	:	:	:	:	:
34	Papua	78,47	61,90	81,93	61,90	2
Jumlah		327,74	537,90	359,68		
Total Cost		1225,31				

d) Hitung Total Simpangan (S)

Setelah didapatkan nilai jarak antara iterasi ke-1 dan iterasi ke-2, hitung total simpangan (S) dengan mencari selisih dari nilai total *cost* baru- nilai total *cost* lama. Dengan ketentuan jika $S < 0$, maka tukar nilai objek dengan menentukan *medoid* baru.

$$S = \text{Total cost baru} - \text{Total cost lama}$$

$$\begin{aligned}
 &= 1225,31 - 1326,18 \\
 &= -100,88
 \end{aligned}$$

- e) Ulangi langkah sebelumnya hingga $S > 0$, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing. Dengan mengasumsikan objek yang telah ditentukan sebagai *non medoids* sebagai berikut:

Tabel 6. Non Medoids Baru

Nama	Keterangan	2009	2010	2011	2012	2013	2014	2015	2016	2017
E1	Maluku Utara	3,78	3,48	3,65	3,12	2,34	1,54	1,39	1,33	1,22
E2	Papua Barat	6,40	4,41	6,99	4,80	4,31	3,37	2,83	2,95	2,62
E3	Papua	27,77	29,59	33,82	31,76	30,87	27,35	26,81	28,98	24,20

Dengan langkah yang sama seperti sebelumnya untuk menentukan jarak dari setiap objek pada iterasi ke-3. Hasil keseluruhan iterasi ke-3 dapat dilihat pada tabel 3.5. sebagai berikut:

Tabel 7. Data Perhitungan Algoritma K-Medoids Hasil Iterasi ke-3

No	Data Ke I	Jarak ke medoids			Terdekat	Klaster yang diikuti
		C1	C2	C3		
1	Aceh	1,94	5,58	79,21	1,94	1
2	Sumatera Utara	2,10	7,81	81,77	2,10	1
3	Sumatera Barat	1,38	6,91	80,91	1,38	1
4	Riau	3,53	9,09	82,93	3,53	1
5	Jambi	1,37	4,86	78,76	1,37	1
6	Sumatera Selatan	1,92	7,13	80,89	1,92	1
7	Bengkulu	2,23	3,99	77,83	2,23	1
8	Lampung	4,27	3,06	75,91	3,06	2
9	Kep. Bangka Belitung	2,11	4,38	78,05	2,11	1
10	Kep. Riau	2,21	7,86	81,84	2,21	1
:	:	:	:	:	:	:
:	:	:	:	:	:	:
34	Papua	80,04	74,41	0,00	0,00	3
	Jumlah	329,77	350,94	2382,86		
	Total Cost	3063,56				

Setelah didapatkan nilai jarak iterasi ke-3, hitung total simpangan (S) dengan mencari selisih dari nilai total *cost* baru- nilai total *cost* lama. Dengan ketentuan jika $S < 0$, maka tukar nilai objek dengan menentukan *medoid* baru.

$$S = \text{Total cost baru} - \text{Total cost lama}$$

$$= 3063,56 - 1225,31$$

$$= 1838,25$$

Dengan nilai $S > 0$ proses *cluster* dihentikan, sehingga diperoleh anggota tiap *cluster* dapat dilihat pada tabel 4.9 sebagai berikut.

Tabel 8. Hasil Pengklasteran Dengan K-Medoids Clustering

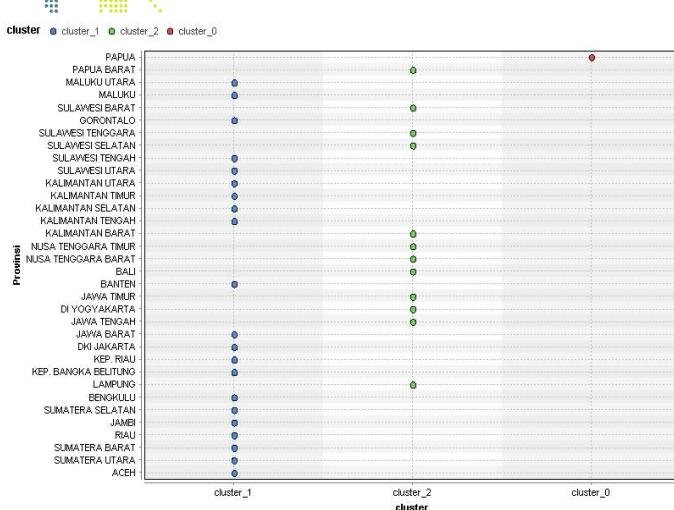
No	Data Ke I	Jarak ke medoids			Terdekat	Klaster yang diikuti
		C1	C2	C3		
1	Aceh	1,94	5,58	79,21	1,94	1

No	Data Ke I	Jarak ke medoids			Terdekat	Klaster yang diikuti
		C1	C2	C3		
2	Sumatera Utara	2,10	7,81	81,77	2,10	1
3	Sumatera Barat	1,38	6,91	80,91	1,38	1
4	Riau	3,53	9,09	82,93	3,53	1
5	Jambi	1,37	4,86	78,76	1,37	1
6	Sumatera Selatan	1,92	7,13	80,89	1,92	1
7	Bengkulu	2,23	3,99	77,83	2,23	1
8	Lampung	4,27	3,06	75,91	3,06	2
9	Kep. Bangka Belitung	2,11	4,38	78,05	2,11	1
10	Kep. Riau	2,21	7,86	81,84	2,21	1
11	Dki Jakarta	5,73	11,46	85,36	5,73	1
12	Jawa Barat	1,05	5,49	79,35	1,05	1
13	Jawa Tengah	15,53	10,15	64,60	10,15	2
14	Di Yogyakarta	12,67	7,35	67,54	7,35	2
15	Jawa Timur	19,99	14,51	60,18	14,51	2
16	Banten	2,17	5,32	78,75	2,17	1
17	Bali	18,11	12,64	62,25	12,64	2
18	Nusa Tenggara Barat	33,86	28,35	46,75	28,35	2
19	Nusa Tenggara Timur	19,32	13,78	60,79	13,78	2
20	Kalimantan Barat	16,01	10,65	64,18	10,65	2
21	Kalimantan Tengah	2,43	8,14	82,14	2,43	1
22	Kalimantan Selatan	0,75	5,31	79,35	0,75	1
23	Kalimantan Timur	1,87	7,60	81,58	1,87	1
24	Kalimantan Utara	9,43	13,10	83,53	9,43	1
25	Sulawesi Utara	5,88	11,60	85,47	5,88	1
26	Sulawesi Tengah	3,27	3,62	77,14	3,27	1
27	Sulawesi Selatan	20,31	14,82	59,89	14,82	2
28	Sulawesi Tenggara	11,71	6,45	68,36	6,45	2
29	Gorontalo	1,93	4,51	78,47	1,93	1
30	Sulawesi Barat	18,32	12,71	61,90	12,71	2
31	Maluku	2,30	7,94	81,93	2,30	1
32	Maluku Utara	0,00	5,97	80,04	0,00	1
33	Papua Barat	5,97	0,00	74,41	0,00	2
34	Papua	80,04	74,41	0,00	0,00	3

Dari hasil perhitungan *k-medoids* diperoleh sebanyak 1 provinsi sebagai *cluster* tinggi, 12 provinsi sebagai *cluster* menengah dan 21 provinsi sebagai *cluster* rendah.

3.2. Pengujian Data Menggunakan Tools Rapidminer

Untuk mendapatkan hasil berupa pengelompokan maka pada tahap selanjutnya dilakukan pengelompokan data menggunakan *Tools Rapidminer*. Hasil akhir yang akan ditampilkan adalah berupa pengelompokan dimana hasil dari pengujian data akan tampak *cluster* dengan masing-masing anggota. Tampak pada gambar 1 sebagai berikut:

**Gambar 1.** Hasil Pengelompokan

Pada gambar 1 dapat dilihat bahwa terdapat 3 warna yang berbeda dimana *cluster* 0 berwarna hijau, *cluster* 1 berwarna merah, dan *cluster* 2 berwarna biru. Dari gambar diatas dapat dilihat bahwa diperoleh 1 provinsi (papua) pada *cluster* tertinggi, 12 provinsi (Lampung, Jawa Tengah, Yogyakarta, Jawa Timur, Bali, NTB, NTT, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Barat, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Barat, dan papua barat) pada *cluster* menengah. 21 provinsi (Aceh, Sumatera Utara, Sumatera Utara, Riau, Jambi, Sumatera Selatan, Bengkulu, Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Barat, Banten, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi tengah, Gorontalo, Maluku, Maluku Utara) pada *cluster* rendah.

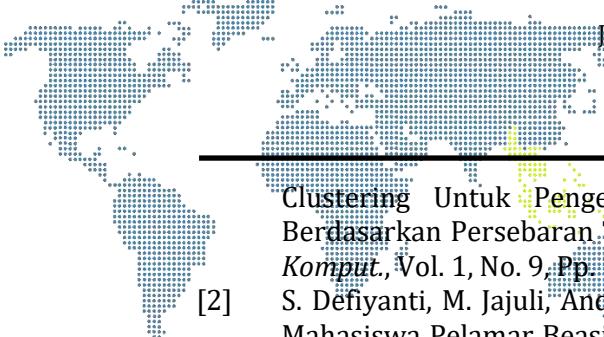
4. SIMPULAN

Berdasarkan pembahasan pada bab sebelumnya dapat disimpulkan bahwa:

- Penerapan data mining dengan metode *k-medoids clustering* dapat diterapkan. Sumber data yang digunakan pada penelitian ini adalah www.bps.go.id dengan subjek data Penyakit Alergi Pada Anak berdasarkan provinsi (2011-2019). Jumlah *record* yang digunakan sebanyak 34 provinsi dengan menghasilkan 21 *cluster* yakni *cluster* rendah sebanyak 12 provinsi *cluster* sedang sebanyak dan *cluster* tinggi sebanyak 1 provinsi.
- Berdasarkan hasil pengujian *k-medoids* untuk kasus persentase data Penyakit Alergi Pada Anak menggunakan *tools Rapidminer* diperoleh hasil yang sama dengan analisis perhitungan metode dimana diperoleh 21 provinsi dengan *cluster* rendah yang menjadi pusat perhatian bagi pemerintah terutama Bagi Dinas Kesehatan dalam melakukan sosialisasi dan pemerataan dalam pemberian imunisasi Untuk Anak yang terkena Alergi pada provinsi tersebut.

DAFTAR PUSTAKA

- [1] D. F. Pramesti, M. T. Furqon, And C. Dewi, "Implementasi Metode K-Medoids

- 
- Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan / Lahan Berdasarkan Persebaran Titik Panas (Hotspot)," *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, Vol. 1, No. 9, Pp. 723-732, 2017.
- [2] S. Defiyanti, M. Jajuli, And N. R. W, "Optimalisasi K - Medoid Dalam Pengklasteran Mahasiswa Pelamar Beasiswa Dengan Cubic Clustering Criterion," *Teknosi*, Vol. 03, No. 01, Pp. 211-218, 2017.
- [3] N. Pulungan, S. Suhada, And D. Suhendro, "Penerapan Algoritma K-Medoids Untuk Mengelompokkan Penduduk 15 Tahun Keatas Menurut Lapangan Pekerjaan Utama," *Komik (Konferensi Nas. Teknol. Inf. Dan Komputer)*, Vol. 3, No. 1, Pp. 329-334, 2019, Doi: 10.30865/Komik.V3i1.1609.
- [4] N. Sirait, "Implementasi K-Means Clustering Pada Pengelompokan Mutu Biji Sawit," *J. Pelita Inform.*, Vol. 16, No. 4, Pp. 368-372, 2017.
- [5] D. Listiyanti, Y. A. Syahbana, And S. R. Henim, "Perancangan Dan Implementasi Aplikasi Android Penentu Salient Area Pada Video Dengan Algoritma," In *Annual Research Seminar 2016*, 2016, Vol. 2, No. 1, Pp. 96-101.
- [6] S. Sundari, I. S. Damanik, A. P. Windarto, H. S. Tambunan, J. Jalaluddin, And A. Wanto, "Analisis K-Medoids Clustering Dalam Pengelompokan Data Imunisasi Campak Balita Di Indonesia," *Pros. Semin. Nas. Ris. Inf. Sci.*, Vol. 1, No. September, P. 687, 2019, Doi: 10.30645/Senaris.V1i0.75.
- [7] A. Y. Rofiqi, "Clustering Berita Olahraga Berbahasa Indonesia Menggunakan Metode K-Medoid," *J. Simantec*, Vol. 6, No. 1, 2017.