Analisis K-Medoids Clustering Dalam Pengelompokkan Data Imunisasi Campak Balita di Indonesia

Siti Sundari¹, Irfan Sudahri Damanik², Agus Perdana Windarto³, Heru Satria Tambunan⁴, Jalaluddin⁵, Anjar Wanto⁶

¹Mahasiswi Sistem Informasi STIKOM Tunas Bangsa Pematangsiantar, Indonesia ^{2,3,4,5,6}Dosen STIKOM Tunas Bangsa Pematangsiantar, Indonesia Jl. Jend Sudirman Blok A No 1,2,3 Pematangsiantar, Indonesia sitisundari751@gmail.com

Abstract - Measles is a contagious infections disease that attacks children caused by a virus. Transmission of measles from people through coughing and sneezing. Measles causes disability and death, so further threatment is needed. Measles immunization program that can inhibit the development of measles is one of the efforts in eradicating the disease. In this study the data used were sourced from the Central Statistics Agency National in 2013-2017. This study uses datamining techniques in data processing with K-Medoids algorithm. The K-Medoids method is a clustering method that functions to break datasets into groups. The advantages of this method are the ability to overcome the weaknesses of the K-Means method which is sensitive to outliers. Another advantage of this algorithm is that the results of the clustering process do not depend on the entry sequence of the dataset. The k-medoids clustering method can be applied to the data on the percentage of measles immunization can be identified based on province, so that the grouping of provinces based on these data. From the data grouping three clusters are obtained: low cluster (2 provinces), medium cluster (30 provinces) and high cluster (2 provinces) with the percentage of measles immunization in each of these provinces from data grouping in percentage. It is expected this research can provide information to the governant about the data on grouping measles immunization for toddlers in Indonesia which has an impact on the distribution of immunization against measles toddlers in Indonesia.

Keywords: Data Mining, K-Medoids Clustering, Measles Immunization

Abstrak - Penyakit campak merupakan penyakit infeksi menular yang menyerang anakanak karena disebabkan oleh virus. Penularan penyakit campak dari orang melalui batuk dan bersin. Penyakit campak perlu ditangani karena menyebabkan kecacatan dan kematian. Program imunisasi campak merupakan salah satu upaya dalam pemberantasan penyakit campak yaitu dengan cara memberikan kekebalan tubuh yang dapat menghambat perkembangan penyakit campak tersebut. Pada penelitian ini data yang digunakan bersumber dari Badan Pusat Statistik Nasional tahun 2004-2017. 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 kmedoids clustering dapat diterapkan pada data persentase imunisasi campak balita berdasarkan provinsi, sehingga dapat diketahui pengelompokkan provinsi berdasarkan data tersebut. Dari data pengelompokan tersebut diperoleh 3 cluster yakni cluster rendah (2 provinsi), cluster sedang (30 provinsi) dan cluster tinggi (2 provinsi) dari persentase imunisasi campak balita pada setiap provinsi. Diharapkan penelitian ini dapat memberikan

informasi kepada pemerintah tentang pengelompokan data imunisasi campak balita di Indonesia yang berdampak pada pemerataan dalam pemberian imunisasi campak balita di Indonesia.

Kata kunci: Data Mining, K-Medoids Clustering, Imunisasi Campak

1. PENDAHULUAN

Penyakit campak merupakan penyakit infeksi menular yang menyerang anak-anak maupun dewasa yang disebabkan oleh virus. Penularan penyakit campak dapat melalui kontak langsung melalui batuk dan bersin. Berdasarkan data Organisasi Kesehatan Dunia[1], Indonesia termasuk 10 negara dengan kasus campak terbesar didunia. Kementerian kesehatan mencatat jumlah kasus campak sangat banyak dan cenderung meningkat dalam kurun waktu lima tahun terakhir. Beberapa komplikasi yang dialami penderita campak diantaranya radang paru, infeksi telinga, diare dan radang otak. Penyakit campak merupakan penyebab utama kematian anak di antara penyakit yang dapat dicegah dengan imunisasi.

Berdasarkan data Kemenkes RI (2016) [2] "Indonesia memiliki cakupan imunisasi campak sebesar 84% dan termasuk kedalam kategori imunisasi campak sedang dibandingkan dengan negara lain di Asia Tenggara". Indonesia ikut serta dalam program eliminasi campak pada tahun 2020 dengan cakupan campak minimal 95% di setiap wilayah secara merata sehingga imunisasi campak mendapatkan perhatian lebih dari pemerintah. Program imunisasi campak merupakan salah satu upaya dalam pemberantasan penyakit campak yaitu dengan memasukkan racun penyakit tertentu yang sudah dilemahkan ke dalam tubuh dengan cara disuntik atau diminum. Hal ini bertujuan untuk meningkatkan kekebalan tubuh terhadap penyakit tersebut. Di Indonesia tingkat penyakit campak pada balita masih sangat tinggi. Diperkirakan sekitar 30.000 anak Indonesia meninggal setiap tahunnya disebabkan komplikasi campak. Hal ini di buktikan dengan lebih dari 95% kematian akibat penyakit campak yang terjadi di negara-negara didunia seperti Amerika Serikat, Inggris, Yaman, Amerika Latin, Myanmar, Afrika, Chad, Afrika Tengah, Liberia, Guinea dan Indonesia. Dalam hal ini pemerintah harus berperan aktif dan berkomitmen untuk melenyapkan penyakit campak. Oleh karena itu, untuk dapat memutuskan mata rantai penularan penyakit campak khususnya balita diperlukan imunisasi minimal 95%. Untuk bisa mewujudkan hal ini pemerintah memerlukan data yang akurat untuk melakukan pengelompokkan data imunisasi campak balita. Data diperoleh dari BPS mengenai data persentasi imunisasi campak balita berdasarkan provinsi di Indonesia pada tahun 2004-2017.

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: [3] diperoleh penilaian berdasarkan indeks Desa/Kelurahan yang memiliki sarana kesehatan dengan 4 provinsi dengan tingkat sarana kesehatan tinggi yaitu Sumatera Utara, Jawa Barat, Jawa Tengah, Jawa Timur, 14 provinsi dengan tingkat sarana kesehatan sedang, dan 16 provinsi lainnya termasuk tingkat sarana kesehatan rendah.

Penelitian selanjutnya [4] 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 [5] dihasilkan dataset dengan kodifikasi keseluruhan data menempati predikat terbaik dalam keseragaman dalam pengelompokan dengan nilai 2,245, dataset kodifikasi keseluruhan menunjukkan nilai CCC berada diantara 2 sampai 3 ini menunjukkan bahwa dataset kodifikasi keseluruhan mempunyai keseragaman yang baik. Dari beberapa penelitian tersebut data mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang diperoleh dengan cara mengenali pola yang penting dari data yang terdapat pada basis data. Metode k-medoids merupakan salah satu metode pengelompokkan 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-means vang sensitive terhadap outlier dan hasil proses clustering tidak bergantung pada urutan masuk dataset" [6].

Metode *k-medoids clustering* dapat diterapkan pada data persentase imunisasi campak balita berdasarkan provinsi, sehingga dapat diketahui pengelompokkan provinsi berdasarkan data tersebut. Dari data pengelompokan tersebut dapat dilihat karakteristiknya sehingga diketahui *cluster* rendah, *cluster* sedang dan *cluster* tinggi dengan persentase imunisasi campak pada setiap provinsi tersebut. Berdasarkan latar belakang masalah di atas, penulis mengambil judul skripsi "Analisis *K-Medoids Clustering* Dalam Pengelompokkan Data Imunisasi Campak Balita di Indonesia". Diharapkan penelitian ini dapat memberikan informasi kepada pemerintah tentang data pengelompokan imunisasi campak balita di Indonesia yang berdampak pada pemerataan pemberian imunisasi campak balita di Indonesia.

2. METODOLOGI PENELITIAN

Penelitian ini dilakukan untuk mengaplikasikan data mining dalam pengelompokkan data imunisasi campak balita untuk menemukan tingkat imunisasi rendah supaya dapat dilakukan sosialisasi ke provinsi tersebut.

2.1. Analisa Permasalahan

Berdasarkan analisa yang dilakukan penulis menyimpulkan dasar permasalahan dalam penelitian ini ialah cakupan penyakit campak balita yang semakin meningkat setiap tahunnya. Indonesia termasuk 10 besar negara dengan cakupan penyakit campak terbesar didunia. Untuk dapat memutuskan mata rantai penularan penyakit campak khususnya balita diperlukan imunisasi minimal 95%. Mengingat di Indonesia memiliki cakupan imunisasi campak sebesar 84% dan termasuk kedalam kategori imunisasi campak sedang dibandingkan dengan negara lain di Asia Tenggara. Untuk itu diperlukan peningkatan lebih lanjut yaitu dengan dilakukan pemerataan ke provinsi di Indonesia.

2.2. Metode Pengumpulan Data

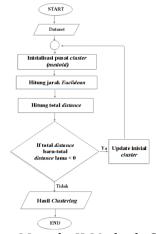
Dalam hal pengumpulan data penulis memanfaatkan perpustakaan, buku, prosiding atau jurnal sebagai media untuk bahan referensi dalam menentukan parameter yang digunakan dalam penelitian. Sumber data penelitian diperoleh dari Badan Pusat Statistik Nasional dengan situs https://www.bps.go.id. Data yang digunakan dalam penelitian ini adalah data persentase imunisasi campak balita pada setiap provinsi dari tahun 2004-2017 yang terdiri dari 34 provinsi. Variabel yang digunakan ialah persentase imunisasi campak balita yang sudah diakumulasikan.

2.3. Analisis Data

Proses analisis data dapat dilakukan setelah adanya pengumpulan data. Dalam melakukan penelitian ini, penulis melakukan analisis data statistik deskriptif. Jenis data yang digunakan dalam penelitian ini adalah data sekunder. Data sekunder adalah data yang diperoleh tidak dari sumbernya langsung, melainkan sudah dikumpulkan oleh pihak lain dan sudah diolah serta memiliki keterkaitan dengan permasalahan yang diteliti.

2.4. Algoritma K-Medoids Clustering

Adapun metode yang digunakan dalam penelitian ini adalah metode *k-medoids clustering*. Menurut [7] "Metode *k-medoids* adalah metode pengelompokkan yang berkaitan dengan metode *k-means* dan metode *medoidshift*". Penelitian ini dilakukan untuk mengetahui diketahui *cluster* rendah, *cluster* sedang dan *cluter* tinggi dari data imunisasi campak balita pada provinsi di Indonesia.



Gambar 1. Flowchart Metode K-Medoids Clustering

Berdasarkan gambar 2.1 dapat dijelaskan bahwa langkah-langkah metode k-medoids clustering diantaranya:

- 1. Inisialisasi pusat *cluster* sebanyak *k* (jumlah *cluster*)
- 2. Alokasikan setiap data (objek) ke *cluster* terdekat menggunakan persamaan ukuran jarak *Euclidian Distance* dengan persamaan:

$$d_{ij} = \sqrt{\sum_{a=1}^{p} (x_{ia} - x_{ja})^2} = \sqrt{(x_i - x_j)'(x_i - x_j)} \dots (2.1)$$

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

- 3. Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat *medoid* baru.
- 4. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat *medoid* baru.
- 5. 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 medoid.
- 6. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan *medoid*, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing.

3. HASIL DAN PEMBAHASAN

3.1. Penerapan Algoritma K-Medoids Clustering

Pada tahapan ini pengolahan data imunisasi campak balita menurut provinsi di Indonesia menggunakan algoritma *k-medoids*. Berikut adalah data imunisasi campak balita menurut provinsi di Indonesia yang diambil penulis melalui situs Badan Pusat Statistik Nasional. Langkah-langkah dalam menyelesaikan perhitungan manual data mining menggunakan *k-medoids clustering* adalah sebagai berikut:

a. Inisialisasi pusat *cluster* sebanyak 3 *cluster* dari data sampel. Untuk pemilihan setiap *medoid* dipilih secara acak (*random*). Diasumsikan aceh, bali dan papua sebagai *medoid* awal.

Tabel 1. Medoid Awal

Tabel 1. Medola Awai									
Inisial	2004	2005	2006	2007	2008	2009	2010		
Aceh(C1)	74,02	0	70,93	66,98	71,86	74,08	74,24		
Bali(C2)	83,77	80,38	85,08	80,26	83,05	82,96	82,5		
Papua(C3)	71,07	65,53	73,09	67,88	58,63	62,63	59,22		

Tabel 2. Lanjutan *Medoid* Awal

Inisial	2011	2012	2013	2014	2015	2016	2017
Aceh(C1)	74,13	75,1	74,09	74,11	60,73	59,09	53,22
Bali(C2)	82,71	80,22	83,24	84,56	81,27	80,8	82,03
Papua(C3)	58,68	62,91	64,54	63,53	57,61	55,85	58,77

b. Menghitung nilai jarak terdekat (cost) dengan persamaan Euclidian Distance. Untuk menghitung jarak antara titik centroid dengan titik tiap objek menggunakan Euclidian Distance. Rumus untuk menghitung jarak menggunakan persamaan (2.1). Maka perhitungan untuk menghitung jarak setiap objek dengan medoid awal adalah sebagai berikut:

$$D_{aceh,c1} = \begin{pmatrix} ((74,02 - 74,02)^2 + (0 - 0)^2 + (70,93 - 70.93)^2 \\ +(66,98 - 66,98)^2 + (71,86 - 71,86)^2 + (74,08 - 74,08)^2 \\ +(74,24 - 74,24)^2 + (74,13 - 74,13)^2 + (75,1 - 75,1)^2 \\ +(74,09 - 74,09)^2 + (74,11 - 74,11)^2 + (60,73 - 60,73)^2 \\ +(59,09 - 59,09)^2 + (53,22 - 53,22)^2) \end{pmatrix} = 0$$

Hasil dari keseluruhan dapat diihat pada tabel 4.3 sebagai berikut:

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

Tabel 3. Has		rak Ke <i>Med</i> o		Cluster	
Provinsi	C1 C2		С3	Terdekat	Yang Diikuti
Aceh	0	96,02538	74,0376	0	1
Sumatera Utara	63,11258	57,15747	33,0071	33,0071	3
Sumatera Barat	71,98269	47,38079	31,26423	31,26423	3
Riau	70,33624	38,63995	40,05805	38,63995	2
Jambi	75,34014	32,65599	44,78451	32,65599	2
Sumatera Selatan	80,66526	21,72912	55,28308	21,72912	2
Bengkulu	83,37242	17,39546	60,30339	17,39546	2
Lampung	84,51967	14,76612	62,18156	14,76612	1
Kep.Bangka Belitung	78,71582	26,65942	53,18272	26,65942	2
Kep. Riau	113,5591	85,2557	95,05084	85,2557	2
Dki Jakarta	87,36542	16,83969	61,94578	16,83969	2
Jawa Barat	78,36313	24,9206	53,62511	24,9206	2
Jawa Tengah	89,53621	10,83436	65,95398	10,83436	2
Di Yogyakarta	100,8915	7,921073	78,52987	7,921073	2
Jawa Timur	82,33637	18,63446	58,0792	18,63446	2
Banten	67,81212	49,10187	35,21937	35,21937	3
Bali	96,02538	0	74,98529	0	2
NTB	88,59683	12,1318	68,79151	12,1318	2
NTT	86,07404	15,30406	61,69807	15,30406	2
Kalimantan Barat	68,19714	47,00769	35,09256	35,09256	3
Kalimantan Tengah	82,25741	30,07454	49,17509	30,07454	2
Kalimantan Selatan	74,75191	36,92948	40,82986	36,92948	2
Kalimantan Timur	88,27127	13,74933	66,74689	13,74933	2
Kalimantan Utara	232,4592	274,4266	215,724	215,724	3
Sulawesi Utara	91,12718	10,85926	68,18873	10,85926	2
Sulawesi Tengah	71,91873	40,23259	37,64712	37,64712	3
Sulawesi Selatan	72,01142	32,2397	47,64244	32,2397	2
Sulawesi Tenggara	81,41584	22,8474	53,1835	22,8474	2
Gorontalo	83,54772	17,22122	59,32806	17,22122	2
Sulawesi Barat	78,64983	125,5459	102,1439	78,64983	1
Maluku	70,19747	50,13856	33,31849	33,31849	3
Maluku Utara	72,10136	38,3397	46,39503	38,3397	2
Papua Barat	76,86977	121,2894	106,4806	76,86977	1
Papua	74,0376	74,98529	0	0	3
Jumlah	251,5723	1529,24	2069,877		
Total Cost		3850,68971			

Setelah didapatkan hasil jarak dari setiap objek (*cost*) pada iterasi ke-1 maka lanjut ke iterasi ke-2. Inisialisasi kandidat *medoid* baru(*non-medoid*) pada iterasi ke-2 dapat dilihat pada tabel 3.4 berikut:

Tabel 4. Medoid Baru (Non-Medoid 1) Iterasi ke-2

Inisial	2004	2005	2006	2007	2008	2009	2010	
Sulawesi Tengah(C1)	72,95	68,63	73,74	70,63	72,04	72,77	74,78	
Sulawesi Selatan(C2)	71,97	66,55	75,79	72,65	73,89	75,99	76,65	
Sulawesi Tenggara(C3)	81,33	74,68	81,3	75,88	77,45	78,62	76,87	

Tabel 5. Lanjutan *Medoid* Baru (*Non-Medoid* 1) Iterasi ke-2

Inisial	2011	2012	2013	2014	2015	2016	2017
Sulawesi Tengah(C1)	70,54	71,59	74,31	76	67,33	70,47	68,58
Sulawesi Selatan(C2)	75,29	77,1	77,66	78,22	71,92	72,99	71,38
Sulawesi Tenggara(C3)	74,79	75,81	78,4	79,39	71,59	74,56	71,87

Hitung kembali jarak dari setiap objek pada iterasi ke-2 dengan menggunakan *medoid* baru pada tabel 3.4.

$$D_{aceh,c1} = \begin{pmatrix} ((74,02-72,95)^2 + (0-68,63)^2 + (70,93-73,74)^2 \\ +(66,98-70,63)^2 + (71,86-72,04)^2 + (74,08-72,77)^2 \\ +(74,24-74,78)^2 + (74,13-70,54)^2 + (75,1-71,59)^2 \\ +(74,09-74,31)^2 + (74,11-76)^2 + (60,73-67,33)^2 \\ +(59,09-70,47)^2 + (53,22-68,58)^2) \end{pmatrix} = 71,91873$$

Sehingga didapatkan hasil keseluruhannya dari iterasi ke-2 dapat dilihat pada tabel 3.6 sebagai berikut:

Tabel 6. Hasil Perhitungan Algoritma K-Medoids Iterasi ke-2

		rak Ke <i>Med</i> o		autus Iterasi	Cluster
Provinsi				Terdekat	Yang
	C1	C2	C3		Diikuti
Aceh	71,91873	72,01142	72,01142	71,91873	1
Sumatera Utara	19,91459	25,46553	25,46553	19,91459	1
Sumatera Barat	11,07023	19,58501	19,58501	11,07023	1
Riau	9,929995	13,89013	13,89013	9,929995	1
Jambi	12,76089	11,6799	11,6799	11,6799	1
Sumatera Selatan	21,5769	14,03715	14,03715	14,03715	3
Bengkulu	26,65353	19,69605	19,69605	19,69605	3
Lampung	27,34701	19,27787	19,27787	19,27787	3
Kep. Bangka Belitung	18,38122	10,28028	10,28028	10,28028	3
Kep. Riau	78,451	74,71066	74,71066	74,71066	3
Dki Jakarta	29,40755	23,11587	23,11587	23,11587	3
Jawa Barat	19,32088	13,21098	13,21098	13,21098	3
Jawa Tengah	31,61015	24,06678	24,06678	24,06678	3
Di Yogyakarta	45,10679	37,73624	37,73624	37,73624	3
Jawa Timur	23,01887	15,07444	15,07444	15,07444	3
Banten	13,42127	18,43516	18,43516	13,42127	1
Bali	40,23259	32,2397	32,2397	32,2397	3
NTB	33,28346	24,00823	24,00823	24,00823	3
NTT	26,89794	19,01886	19,01886	19,01886	3
Kalimantan Barat	10,05772	16,53914	16,53914	10,05772	1
Kalimantan Tengah	18,62203	16,41238	16,41238	16,41238	2
Kalimantan Selatan	7,851115	10,04155	10,04155	7,851115	1
Kalimantan Timur	31,88479	23,85058	23,85058	23,85058	2
Kalimantan Utara	240,8193	248,0118	248,0118	240,8193	1
Sulawesi Utara	33,87507	27,08849	27,08849	27,08849	3
Sulawesi Tengah	0	11,6271	11,6271	0	1
Sulawesi Selatan	11,6271	0	0	0	2
Sulawesi Tenggara	19,21439	14,85877	14,85877	14,85877	2
Gorontalo	24,91042	17,59364	17,59364	17,59364	3

	Jai	rak Ke <i>Med</i> o	oid		Cluster
Provinsi				Terdekat	Yang
	C1	C2	C3		Diikuti
Sulawesi Barat	101,4308	100,6836	100,6836	100,6836	2
Maluku	13,82004	19,81427	19,81427	13,82004	1
Maluku Utara	14,23434	11,8947	11,8947	11,8947	2
Papua Barat	100,9995	98,82697	98,82697	98,82697	2
Papua	37,64712	47,64244	47,64244	37,64712	1
Jumlah	1227,297	1152,426	1152,426		
Total Cost	· ·	3532,148633	,		

c. 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 =Total cost baru Total cost lama
 - =3532,148633 3850,689717
- = -318,541
- d. Ulangi langkah sebelumnya hingga S>0, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing.

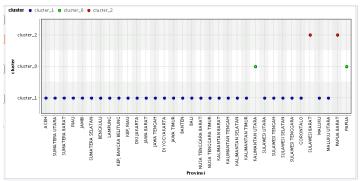
Setelah didapatkan nilai jarak iterasi ke-3, hitung total simpangan (S) dengan mencari selisih dari nilai total *cost* baru- nilai total *cost* lama. Dengan ketentuan jika S<0, maka tukar nilai objek dengan menentukan *medoid* baru.

- S =Total cost baru Total cost lama
 - =6857.06637-3532.148633
 - = 3324,918

Dengan nilai S>0 proses *cluster* dihentikan, sehingga diperoleh anggota tiap *cluster*.

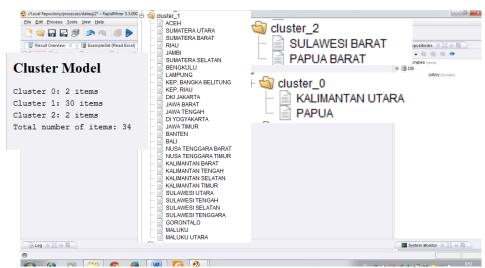
3.2. Implementasi Sistem Rapidminer 5.3

Pada tahap ini menampilkan hasil akhir serta langkah terakhir dalam penggunaan *tools Rapidminer* 5.3 ini. Hasil akhir yang akan ditampilkan adalah berupa pengelompokkan dimana hasil dari pengujian data akan tampak *cluster* dengan masing-masing anggota. Tampak pada gambar 3.1 berikut:



Gambar 2. Hasil Pengelompokkan

Pada gambar 3.1 dapat dilihat bahwa terdapat 3 warna yang berbeda dimana *cluster* 1 berwarna biru, *cluster* 0 berwarna hijau, dan *cluster* 2 berwarna merah.



Gambar 3. Cluster Model

Dari gambar 3.2 dapat dilihat bahwa diperoleh 2 provinsi (kalimantan utara dan papua) pada *cluster* rendah, 30 provinsi (Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep.Bangka Belitung, Kep.Riau, Dki Jakarta, Jawa Barat, Jawa Tengah, Yogyakarta, Jawa Timur, Banten, Bali, NTB, NTT, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Maluku, Maluku Utara) pada *cluster* sedang dan 2 provinsi (sulawesi barat dan papua barat) pada *cluster* tinggi.

3.3. Validasi Data

Dalam melakukan validasi data perhitungan algoritma dengan sistem pada rapidminer menghasilkan hasil akhir berupa pengelompokkan dengan 3 *cluster*, serta data yang digunakan merupakan data yang *valid* dan sama dengan yang dipakai pada *tools Rapidminer 5.3.*

4. KESIMPULAN

Berdasarkan pembahasan pada bab sebelumnya dapat disimpulkan bahwa:

- 1. 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 imunisasi campak balita berdasarkan provinsi (2004-2017). Jumlah *record* yang digunakan sebanyak 34 provinsi dengan menghasilkan 3 *cluster* yakni *cluster* rendah sebanyak 2 provinsi *cluster* sedang sebanyak 30 provinsi dan *cluster* tinggi sebanyak 2 provinsi.
- 2. Berdasarkan hasil pengujian *k-medoids* untuk kasus persentase data imunisasi campak balita menggunakan *tools Rapidminer* 5.3 diperoleh hasil yang sama dengan analisis perhitungan metode dimana diperoleh 2 provinsi dengan *cluster* rendah yang menjadi pusat perhatian bagi pemerintah dalam melakukan

sosialisasi dan pemerataan dalam pemberian imunisasi campak balita pada provinsi tersebut.

DAFTAR PUSTAKA

- [1] WHO, "Organisasi Kesehatan Dunia," www.depkes.go.id (25 Februari 2018), 2015.
- [2] Kemenkes RI, Pusat Data dan Informasi Kementerian Kesehatan RI. 2016.
- [3] M. G. Sadewo, A. P. Windarto, and S. R. Andani, "Pemanfaatan Algoritma Clustering Dalam Pengelompokkan Jumlah Desa/Kelurahan Yang Memiliki Sarana Kesehatan," *KOMIK(Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. I, no. 1, 2017.
- [4] 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, 2017.
- [5] S. Defiyanti, M. Jajuli, and N. rohmawati W, "Optimalisasi K Medoid Dalam Pengklasteran Mahasiswa Pelamar Beasiswa dengan Cubic Clustering Criterion," *TEKNOSI*, vol. 3, no. 1, 2017.
- [6] E. Setyowati, A. Rusgiyono, and M. A. Mukid, "Analisis Pengelompokan Daerah Menggunakan Metode Non- Hierarchical Partitioning K-Medoids dari Hasil Komoditas Pertanian Tanaman Pangan," *J. GAUSSIAN*, vol. 4, no. 4, 2015.
- [7] D. Listiyanti, Y. A. Syahbana, and S. R. Henim, "Perancangan dan Implementasi Aplikasi Android Penentu Salient Area pada Video dengan Algoritma K-Medoids," in *ANNUAL RESEARCH SEMINAR 2016*, 2016, vol. 2, no. 1.