

Penerapan *Data Mining* Untuk Klastering Balita Penerima Imunisasi Campak Menggunakan Metode K-Means

Muhammad Faisal*¹, Wiranti Sri Utami²

^{1,2} Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Raharja
E-mail: *muhammad.faisal@raharja.info, ²wiranti.utami@raharja.info

Abstrak

Imunisasi Campak adalah imunisasi yang dilakukan untuk mencegah penyakit campak terhadap balita yang disebabkan oleh virus rubeola dan rubella, imunisasi ini diberikan mulai dari rentang usia 9 bulan hingga dewasa. Proses imunisasi ini diberikan untuk merangsang kekebalan tubuh agar kebal terhadap penyakit tersebut. Di Indonesia penyakit campak merupakan masalah yang harus dihadapi sebab dapat meningkatkan kematian pada balita. Oleh sebab itu penting bagi balita melakukan imunisasi campak untuk membentengi sistem kekebalan tubuh dari penyakit campak. Tujuan dari penelitian ini adalah untuk mengetahui balita yang sudah menerima imunisasi campak berdasarkan Provinsi dengan menggunakan *data mining* dan algoritma K-Means dan menggunakan dan proses perhitungan menggunakan Microsoft Excel. Penelitian ini menggunakan 34 Provinsi dengan tingkat pemberian Imunisasi campak yang sudah dilakukan oleh pemerintah.

Kata Kunci : *Data Mining*, Imunisasi Campak, K-Means

Abstract

Measles immunization is an immunization carried out to prevent measles in infants caused by the rubeola and rubella viruses, this immunization is given starting from the age range of 9 months to adults. This immunization process is given to stimulate the body's immunity to be immune to the disease. In Indonesia, measles is a problem that must be faced because it can increase mortality in children under five. Therefore, it is important for toddlers to immunize against measles to fortify the immune system from measles. The purpose of this study was to determine children under five who had received measles immunization by province by using data mining and the K-Means algorithm and using and calculating processes using Microsoft Excel. This study used 34 provinces with the level of measles immunization that had been carried out by the government.

Keyword : *Data Mining*, *Measles immunization*, *K-Means*

1. PENDAHULUAN

Imunisasi campak adalah jenis vaksin yang digunakan untuk mencegah penyakit campak. Vaksin ini termasuk kedalam program kategori rutin lengkap yang dianjurkan oleh pemerintah. Terdapat Dua jenis Vaksin campak yang digunakan untuk mencegah penyakit campak, yaitu vaksin MR dan Vaksin MMR. Vaksin MR mencegah penyakit campak dan rubella, sedangkan vaksin MMR mencegah penyakit campak, rubella, dan gondongan. Proses penelitian ini menggunakan *data mining* sebagai tahapan untuk melakukan analisa data yang memiliki *record* yang banyak berdasarkan observasi. Untuk menemukan suatu hubungan yang saling berkaitan dan dapat merangkum data yang baru sehingga data yang telah melalui proses *data mining* dapat di mengerti oleh pengguna. Pemanfaatan perkembangan sistem informasi memberikan dampak

terhadap perkembangan data dengan kapasitas yang besar di berbagai bidang baik industri maupun instansi. Perencanaan pemberian imunisasi campak terhadap balita adalah salah satu program pemerintah yang sangat penting sebagai upaya pemernitah dalam menanggulangi resiko penyakit campak terhadap balita. Oleh karena itu hal tersebut akan mempengaruhi proses pengadaan, pendistrbusian, dan pemberian imunisasi campak untuk balita.

Proses penerapan menggunakan *data mining* akan pemberian imunisasi campak terhadap balita akan menjadi lebih efisien dan efektif sehingga imunisasi campak dapat tersedia dengan cukup dan diberikan kepada wilayah dengan tingkat imunisasi campak yang rendah. Klustering adalah bidang penelitian untuk proses analisa dan *data mining*. Pada tahapan klustering bagian dari data yang memiliki karakteristik yang sama akan dikelompokkan pada kelompok lain. *Data mining* memiliki banyak algoritma untuk proses pengelompokan klustering salah satunya adalah K-Means. Peneliti menggunakan metode K-Means karena memiliki algoritma yang sederhana dan lebih efisien.

K-Means merupakan metode pengelompokan data yang dapat membagi data kedalam dua kelompok ataupun lebih. Metode ini akan membagi data kesebuah kelompok dimana data yang memiliki karakteristik yang sama akan berada dalam satu kelompok yang sama dan data yang memiliki karaktersiti berbeda akan berada di kelompok yang berbeda. Tujuan dari pengelompokkan data yaitu untuk meminimalkan fungsi dari data yang telah disesuaikan dalam proses pengelompokkan data.

2. METODE PENELITIAN

Metode Penelitian data yang diambil dari Badan Pusat Statistik (BPS) pada tahun 2017, 2019, dan 2020.

Tabel 1. Data Balita Penerima Imunisasi Campak

PROVINSI	Persentase Balita yang Pernah Mendapat Imunisasi Campak (Persen)		
	2017	2019	2020
Aceh	53,22	40,59	39,21
Sumatera Utara	65,39	60,65	62,72
Sumatera Barat	61,59	58,84	56,40
Riau	65,04	56,45	60,96
Jambi	67,08	66,92	63,47
Sumatera Selatan	72,51	66,45	69,00
Bengkulu	72,94	71,53	75,38
Lampung	74,44	70,47	73,18
Kep. Bangka Belitung	73,67	66,07	65,47
Kep. Riau	71,28	74,26	68,47
Dki Jakarta	74,40	72,28	72,82
Jawa Barat	69,44	65,77	67,47
Jawa Tengah	75,49	74,72	73,36
Di Yogyakarta	85,45	79,14	78,67
Jawa Timur	73,26	71,94	70,67
Banten	66,86	60,17	58,65
Bali	82,03	78,35	80,18
Nusa Tenggara Barat	76,57	75,07	73,07
Nusa Tenggara Timur	75,82	72,74	74,39

Kalimantan Barat	65,42	63,49	65,88
Kalimantan Tengah	67,05	64,79	65,52
Kalimantan Selatan	71,14	68,20	67,89
Kalimantan Timur	72,94	70,46	68,86
Kalimantan Utara	72,51	71,47	71,63
Sulawesi Utara	75,03	73,83	74,39
Sulawesi Tengah	68,58	66,49	69,85
Sulawesi Selatan	71,38	68,50	73,39
Sulawesi Tenggara	71,87	71,43	69,82
Gorontalo	76,67	73,68	75,13
Sulawesi Barat	69,29	67,61	66,76
Maluku	67,97	65,91	68,78
Maluku Utara	68,46	61,46	65,74
Papua Barat	63,47	64,25	69,08
Papua	58,77	55,43	51,84

Sumber : Badan Pusat Statistik

2.1. Literatur Review

Untuk mendukung penulis jurnal ini maka penulis melakukan studi pustaka dari beberapa penelitian yang berkaitan dengan proses penelitian ini:

1. Penelitian yang dilakukan oleh Gustientiedina, M.Hasmil Adiya, dan Yenny Desnelita (2019) dengan judul “Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru”. Tujuan dari hasil penelitian ini adalah untuk mengetahui perencanaan kebutuhan obat-obatan baik pengadaan maupun ketersediaan pada RSUD Pekanbaru lebih efektif dan efisien dengan menggunakan metode klustering K-Means. [1]
2. Penelitian yang dilakukan oleh Asroni, Hidayatul Fitri, Eko Prasetyo (2018) dengan judul “Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokkan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)”. Tujuan dari penelitian ini adalah untuk mengetahui kesenjangan jumlah mahasiswa baru pada Universitas Muhammadiyah Yogyakarta dengan menggunakan klastering algoritma K-Means. [2]
3. Penelitian yang dilakukan oleh Benri Melpa Metisen, Herlina Latipa Sari (2015) dengan judul “Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokkan Penjualan Produk Pada Swalayan Fadhila” Tujuan dari penelitian ini adalah untuk mengelompokkan data penjualan dengan tingkat penjualan tinggi, sedang, dan rendah dengan menggunakan klustering dan metode K-Means. [3]
4. Penelitian yang dilakukan oleh Randi Rian Putra, dan Cendra Wadisman (2018) dengan judul “Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K-Means” Tujuan penelitian ini adalah untuk implementasi pemilihan pelanggan potensial pada MC Laundry menggunakan *Software Tanagra*.

2.1.1. Proses Algoritma K-Means

1. Melakukan proses perhitungan dari setiap data ke masing-masing klustering dan masing-masing *centeroid* menggunakan Euclidean Distance, dengan menggunakan rumus berikut ini:

$$d(x,y) = \|x - y\| = \sqrt{\sum_i^n (x_i - y_i)^2} : i = 1,2,3, \dots n$$

2. Meletakkan data ke masing-masing obyek ke dalam *centeroid* yang lebih dekat.
3. Melakukan proses literasi dan selanjutnya menentukan *centeroid* baru dengan menggunakan persamaan.
4. Bila proses *centeroid* belum sama ulangi langkah ketiga.

2.1.2. Hasil dan Pembahasan

Penelitian ini menggunakan data yang diambil dari Badan Pusat Statistik (BPS). Parameter yang digunakan adalah nama provinsi, tahun pemberian imunisasi campak 2019 dan 2020. Proses klustering dilakukan untuk mengetahui pemberian imunisasi dengan tingkat paling rendah di setiap provinsi.

1. Literasi 1

Dalam proses penghitungan nilai literasi 1 maka ditentukan nilai K data pusat klustering secara acak. Proses uji coba pertama ini ditentukan 3 data dengan mengambil sampel secara acak sebagai titik awal untuk melakukan proses perhitungan jarak dari seluruh kelompok kluster yang ada.

Jumlah klustering = 2
 Jumlah *centeroid* = 6
 Jumlah atribut = 3

Tabel 2. Data *Centeroid* klustering Iterasi 1

Provinsi	<i>Centeroid</i>		Cluster
DKI Jakarta	72,28	72,82	C1
Yogyakarta	79,14	78,67	C2
Banten	60,17	58,65	C3

2. Proses selanjutnya menghitung jarak setiap masing-masing klustering *centeroid* menggunakan persamaan yaitu dengan Euclidean Distance.

A. Klustering C1

$$\begin{aligned} \text{C1 (Aceh)} &= \sqrt{(40,59 - 72,28)^2 + (39,21 - 72,82)^2} = 46,19 \\ \text{C1 (SUMUT)} &= \sqrt{(60,65 - 72,28)^2 + (62,72 - 72,82)^2} = 15,40 \\ \text{C1 (SUMBAT)} &= \sqrt{(58,84 - 72,28)^2 + (56,40 - 72,82)^2} = 21,22 \\ \text{C1 (Riau)} &= \sqrt{(56,45 - 72,28)^2 + (60,96 - 72,82)^2} = 19,78 \\ \text{C1 (Jambi)} &= \sqrt{(66,92 - 72,28)^2 + (63,47 - 72,82)^2} = 10,78 \\ \text{C1 (SUMSEL)} &= \sqrt{(66,45 - 72,28)^2 + (69,00 - 72,82)^2} = 6,97 \\ \text{C1 (Bengkulu)} &= \sqrt{(71,53 - 72,28)^2 + (75,38 - 72,82)^2} = 2,67 \\ \text{C1 (Lampung)} &= \sqrt{(70,47 - 72,28)^2 + (73,18 - 72,82)^2} = 1,85 \\ \text{C1 (Kep. Bangka)} &= \sqrt{(66,07 - 72,28)^2 + (65,47 - 72,82)^2} = 9,62 \end{aligned}$$

$$\begin{aligned} \text{C1 (Kep. Riau)} &= \sqrt{(74,26 - 72,28)^2 + (68,47 - 72,82)^2} = 4,78 \\ &\dots\dots\dots \\ &\dots\dots\dots \\ \text{C1 (Papua)} &= \sqrt{(55,43 - 72,28)^2 + (51,84 - 72,82)^2} = 26,91 \end{aligned}$$

B. Klustering C2

$$\begin{aligned} \text{C2 (Aceh)} &= \sqrt{(40,59 - 79,14)^2 + (39,21 - 78,67)^2} = 55,17 \\ \text{C2 (SUMUT)} &= \sqrt{(60,65 - 79,14)^2 + (62,72 - 78,67)^2} = 24,42 \\ \text{C2 (SUMBAT)} &= \sqrt{(58,84 - 79,14)^2 + (56,40 - 78,67)^2} = 30,13 \\ \text{C2 (Riau)} &= \sqrt{(56,45 - 79,14)^2 + (60,96 - 78,67)^2} = 28,78 \\ \text{C2 (Jambi)} &= \sqrt{(66,92 - 79,14)^2 + (63,47 - 78,67)^2} = 19,50 \\ \text{C2 (SUMSEL)} &= \sqrt{(66,45 - 79,14)^2 + (69,00 - 78,67)^2} = 15,95 \\ \text{C2 (Bengkulu)} &= \sqrt{(71,53 - 79,14)^2 + (75,38 - 78,67)^2} = 8,29 \\ \text{C2 (Lampung)} &= \sqrt{(70,47 - 79,14)^2 + (73,18 - 78,67)^2} = 10,26 \\ \text{C2 (Kep. Bangka)} &= \sqrt{(66,07 - 79,14)^2 + (65,47 - 78,67)^2} = 18,58 \\ \text{C2 (Kep. Riau)} &= \sqrt{(74,26 - 79,14)^2 + (68,47 - 78,67)^2} = 11,31 \\ &\dots\dots\dots \\ &\dots\dots\dots \\ \text{C2 (Papua)} &= \sqrt{(55,43 - 55,43)^2 + (51,84 - 78,67)^2} = 35,81 \end{aligned}$$

C. Klustering C3

$$\begin{aligned} \text{C3 (Aceh)} &= \sqrt{(40,59 - 60,17)^2 + (39,21 - 58,65)^2} = 27,59 \\ \text{C3 (SUMUT)} &= \sqrt{(60,65 - 60,17)^2 + (62,72 - 58,65)^2} = 4,10 \\ \text{C3 (SUMBAT)} &= \sqrt{(58,84 - 60,17)^2 + (56,40 - 58,65)^2} = 2,61 \\ \text{C3 (Riau)} &= \sqrt{(56,45 - 60,17)^2 + (60,96 - 58,65)^2} = 4,38 \\ \text{C3 (Jambi)} &= \sqrt{(66,92 - 60,17)^2 + (63,47 - 58,65)^2} = 8,29 \\ \text{C3 (SUMSEL)} &= \sqrt{(66,45 - 60,17)^2 + (69,00 - 58,65)^2} = 12,11 \\ \text{C3 (Bengkulu)} &= \sqrt{(71,53 - 60,17)^2 + (75,38 - 58,65)^2} = 20,22 \\ \text{C3 (Lampung)} &= \sqrt{(70,47 - 60,17)^2 + (73,18 - 58,65)^2} = 17,81 \\ \text{C3 (Kep. Bangka)} &= \sqrt{(66,07 - 60,17)^2 + (65,47 - 58,65)^2} = 9,02 \\ \text{C3 (Kep. Riau)} &= \sqrt{(74,26 - 60,17)^2 + (68,47 - 58,65)^2} = 17,17 \\ &\dots\dots\dots \\ &\dots\dots\dots \\ \text{C3 (Papua)} &= \sqrt{(55,43 - 60,17)^2 + (51,84 - 58,65)^2} = 8,30 \end{aligned}$$

Berdasarkan hasil dari perhitungan klustering diatas maka didapatkan literasi pertama yang dapat dilihat pada tabel dibawah ini:

Tabel 3. Perhitungan Iterasi 1

Data 1	C1	C2	C3	Cluster
1	46,19	55,17	27,59	3
2	15,40	24,42	4,10	3
3	21,22	30,13	2,61	3
4	19,78	28,78	4,38	3
5	10,78	19,50	8,29	3
6	6,97	15,95	12,11	1
7	2,67	8,29	20,22	1
8	1,85	10,26	17,81	1
9	9,62	18,58	9,02	3
10	4,78	11,31	17,17	1
11	0,00	9,02	18,64	1
12	8,43	17,44	10,45	1
13	2,50	6,91	20,69	1
14	9,02	0,00	27,58	2
15	2,18	10,76	16,82	1
16	18,64	27,58	0,00	3
17	9,54	1,70	28,18	2
18	2,80	6,92	20,74	1
19	1,64	7,70	20,14	1
20	11,20	20,21	7,96	3
21	10,46	19,46	8,28	3
22	6,40	15,36	12,24	1
23	4,36	13,10	14,50	1
24	1,44	10,41	17,21	1
25	2,21	6,82	20,84	1
26	6,51	15,42	12,86	1
27	3,82	11,88	16,93	1
28	3,12	11,74	15,86	1
29	2,70	6,51	21,31	1
30	7,65	16,58	11,01	1
31	7,54	16,52	11,64	1
32	12,93	21,90	7,21	3
33	8,86	17,71	11,20	1
34	26,91	35,81	8,30	3

Setelah mendapatkan nilai literasi pertama selanjutnya menghitung nilai pusat kluster, pengelompokan data pada literasi pertama dan cluster awal dapat dilihat pada tabel berikut ini:

Tabel 4. Nilai *Centeroid* cluster 1

X	Y	X	Y	X	Y
0	0	0	0	40,59	39,21
0	0	0	0	60,65	62,72
0	0	0	0	58,84	56,4
0	0	0	0	56,45	60,96
0	0	0	0	66,92	63,47
66,45	69	0	0	0	0
71,53	75,38	0	0	0	0
70,47	73,18	0	0	0	0
0	0	0	0	66,07	65,47
74,26	68,47	0	0	0	0
72,28	72,82	0	0	0	0
65,77	67,47	0	0	0	0
74,72	73,36	0	0	0	0
0	0	79,14	78,67	0	0
71,94	70,67	0	0	0	0
0	0	0	0	60,17	58,65
0	0	78,35	80,18	0	0
75,07	73,07	0	0	0	0
72,74	74,39	0	0	0	0
0	0	0	0	63,49	65,88
0	0	0	0	64,79	65,52
68,2	67,89	0	0	0	0
70,46	68,86	0	0	0	0
71,47	71,63	0	0	0	0
73,83	74,39	0	0	0	0
66,49	69,85	0	0	0	0
68,5	73,39	0	0	0	0
71,43	69,82	0	0	0	0
73,68	75,13	0	0	0	0
67,61	66,76	0	0	0	0
65,91	68,78	0	0	0	0
0	0	0	0	61,46	65,74
64,25	69,08	0	0	0	0
0	0	0	0	55,43	51,84

Dari hasil perhitungan nilai *centroid* diatas maka akan ditentukan nilai *centeroid* untuk menghitung literasi kedua, nilai *centeroid* pertama dapat dilihat pada tabel dibawah ini:

Tabel 5. Nilai Centeroid 1

Klustering	Centeroid	
C1	70,3362	71,1138
C2	78,745	79,425
C3	59,5327	59,6236

3. Literasi 2

Setelah nilai *centeroid* dari hasil perhitungan literasi pertama telah didapatkan maka selanjutnya melakukan perhitungan literasi kedua. Hasil dari nilai literasi kedua dapat dilihat pada tabel dibawah ini:

Tabel 6. Perhitungan Iterasi 2

Data 2	C1	C2	C3	Cluster
1	43,62	55,44	27,85	3
2	12,82	24,63	3,29	3
3	18,67	30,44	3,30	3
4	17,20	28,95	3,36	3
5	8,37	19,86	8,33	3
6	4,42	16,12	11,65	1
7	4,43	8,27	19,80	1
8	2,07	10,37	17,42	1
9	7,07	18,85	8,77	1
10	4,73	11,84	17,18	1
11	2,59	9,24	18,35	1
12	5,84	17,64	10,02	1
13	4,93	7,28	20,48	1
14	11,60	0,85	27,34	2
15	1,66	11,09	16,61	1
16	16,08	27,87	1,16	3
17	12,10	0,85	27,87	2
18	5,12	7,34	20,55	1
19	4,06	7,84	19,81	1
20	8,62	20,40	7,40	3
21	7,88	19,70	7,90	1
22	3,87	15,63	11,98	1
23	2,26	13,43	14,31	1
24	1,25	10,66	16,93	1
25	4,79	7,04	20,55	1
26	4,05	15,55	12,37	1
27	2,92	11,89	16,43	1
28	1,69	12,07	15,67	1

29	5,23	6,64	20,99	1
30	5,14	16,86	10,78	1
31	5,00	16,67	11,16	1
32	10,38	22,05	6,41	3
33	6,42	17,81	10,57	1
34	24,37	36,12	8,80	3

Nilai literasi pertama dan kedua sudah sama, maka proses perhitungan berhenti pada literasi kedua dan tidak perlu melanjutkan perhitungan ke literasi 3.

Proses klustering data balita penerima imunisasi campak dengan menggunakan metode K-Means diperoleh hasil klustering pada literasi kedua yaitu terdapat 3 kelompok dengan penerima imunisasi tinggi yaitu C3 terdapat 9 provinsi, dengan penerima imunisasi sedang yaitu C2 terdapat 2 provinsi dan imunisasi Sedikit C1 terdapat 23 provinsi.

Tabel 7. Tingkat Klusteri masing-masing Provinsi

Provinsi	Cluster
Aceh	C3
Sumatera Utara	C3
Sumatera Barat	C3
Riau	C3
Jambi	C3
Sumatera Selatan	C1
Bengkulu	C1
Lampung	C1
Kep. Bangka Belitung	C1
Kep. Riau	C1
Dki Jakarta	C1
Jawa Barat	C1
Jawa Tengah	C1
Di Yogyakarta	C2
Jawa Timur	C1
Banten	C3
Bali	C2
Nusa Tenggara Barat	C1
Nusa Tenggara Timur	C1
Kalimantan Barat	C3
Kalimantan Tengah	C1
Kalimantan Selatan	C1
Kalimantan Timur	C1
Kalimantan Utara	C1
Sulawesi Utara	C1

Sulawesi Tengah	C1
Sulawesi Selatan	C1
Sulawesi Tenggara	C1
Gorontalo	C1
Sulawesi Barat	C1
Maluku	C1
Maluku Utara	C3
Papua Barat	C1
Papua	C3

4. KESIMPULAN

Berdasarkan hasil dari proses klustering diatas dapat diperoleh informasi mengenai tingkat imunisasi dengan kategori rendah, sedang, dan tinggi dengan memanfaatkan algoritma K-Means.

5. SARAN

Dari hasil penelitian klustering menggunakan algoritma K-Means diatas mungkin bisa dilakukan penelitian lebih lanjut agar proses analisa imunisasi campak diatas dapat diperoleh hasil analisa yang lebih tepat dengan menetapkan nilai *centeroid*.

DAFTAR PUSTAKA

- [1] Gustientiedina, M.Hasmil Adiya, dan Yenny Desnelita, 2019, *Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru*, Vol. 05, Jurnal Naskah Teknologi dan Informasi, Universitas Andalas, Kota Padang.
- [2] Asroni, Hidayatul Fitri, Eko Prasetyo, 2018, *Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)*, Vol. 21, Jurnal Semesta Teknik, Universitas Muhammadiyah Yogyakarta, Yogyakarta.
- [3] Benri Melpa Metisen, Herlina Latipa Sari, 2015, *Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila*, Vol. 5, Jurnal Media Infotama, Universitas Dehasen, Bengkulu.
- [4] Benri Melpa Metisen, Herlina Latipa Sari, 2018, *Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K-Means*, Vol. 1, Intecom: Journal of Information Technology and Computer Science, IPM2KPE, Sumatera Selatan.