

Pengelompokan Jumlah Kasus Penyakit Aids Berdasarkan Provinsi Menggunakan Metode K-Means

Rut Indra Lita Sinaga¹, Widodo Saputra², Hendry Qurniawan³

^{1,3} STIKOM Tunas Bangsa, Pematangsiantar, Sumatera Utara, Indonesia

² AMIK Tunas Bangsa, Pematangsiantar, Sumatera Utara, Indonesia

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

¹Rutindralita@gmail.com, ²widodosaputra@amiktunasbangsa.ac.id,

³hndryQ@gmail.com

Abstract

Acquired Immune Deficiency Syndrome (AIDS) is a collection of symptoms due to a gradual decline in the immune system caused by infection with the Human Immunodeficiency Virus (HIV). This disease is a dangerous disease and should be watched out for where it spreads very quickly. AIDS is one of the top infectious diseases that can cause death. K-Means is an algorithm in data mining that can be used to group / cluster data. There are many approaches to creating clusters, one of which is to create rules that dictate membership in the same group based on the level of equality among its members. The purpose of this study was to classify the number of AIDS cases by province. To solve the existing problems, the authors will use the K-Means cluster method using 2 clusters to determine the province which has the highest cases of AIDS and the province with the lowest cases of AIDS by calculating the centroid / average of the data in the cluster. It is especially recommended that the government take advantage of the results of this research to pay more attention and make efforts in overcoming AIDS in provinces with high AIDS disease.

Keywords: Cluster, K-Means, Province, AIDS

Abstrak

Acquired Immune Deficiency Syndrome (AIDS) adalah kumpulan gejala penyakit akibat menurunnya sistem kekebalan tubuh secara bertahap yang disebabkan oleh infeksi Human Immunodeficiency Virus (HIV). Penyakit ini merupakan penyakit berbahaya dan harus diwaspadai dimana penyebarannya sangat cepat. AIDS merupakan salah satu penyakit infeksi peringkat atas yang dapat menyebabkan kematian. K-Means merupakan salah satu algoritma dalam data mining yang bisa digunakan untuk melakukan pengelompokan/clustering suatu data. Ada banyak pendekatan untuk membuat cluster, diantaranya adalah membuat aturan yang mendikte keanggotaan dalam kelompok yang sama berdasarkan tingkat persamaan diantara anggota-anggotanya. Tujuan penelitian ini untuk mengelompokkan Jumlah Kasus Penyakit AIDS berdasarkan Provinsi. Untuk penyelesaian permasalahan yang ada maka penulis akan menggunakan metode cluster K-Means dengan menggunakan 2 cluster untuk mengetahui kelompok Provinsi yang memiliki kasus Penyakit AIDS tertinggi dan Provinsi yang memiliki kasus Penyakit AIDS terendah dengan menghitung centroid/ rata-rata dari data-data pada cluster. Disarankan khususnya kepada Pemerintah untuk memanfaatkan hasil penelitian ini agar lebih memperhatikan lagi dan mengupayakan dalam penanggulangan penyakit AIDS di provinsi yang tinggi akan penyakit AIDS.

Keywords: Cluster, K-Means, Provinsi, AIDS.

1. Pendahuluan

Acquired Immune Deficiency Syndrome (AIDS) adalah kumpulan gejala penyakit akibat menurunnya sistem kekebalan tubuh secara bertahap yang disebabkan oleh

infeksi *Human Immunodeficiency Virus* (HIV). Penyakit ini merupakan penyakit berbahaya dan harus diwaspadai dimana penyebarannya sangat cepat. HIV/AIDS merupakan salah satu penyakit infeksi peringkat atas yang dapat menyebabkan kematian.

Berdasarkan data dari Pusat Data dan Informasi Kementerian Kesehatan RI, kasus HIV/AIDS yang ditemukan di Indonesia sampai Maret tahun 2011 diketahui bahwa jumlah penderita HIV di Indonesia sebanyak 242.699 orang dan AIDS sebanyak 87.453 orang. Jumlah ini mengalami peningkatan tajam dari tahun 2008, dimana jumlah kumulatif penderita HIV sebanyak 150.296 orang dan AIDS sebanyak 55.799 orang.

Infeksi HIV cenderung meningkat dan paling banyak terjadi pada kelompok usia produktif, yaitu kelompok umur 25-49 tahun dan kelompok umur 20-24 tahun. Usia remaja 15-19 tahun menduduki posisi keempat dan usia remaja merupakan usia yang sangat rentan untuk terinfeksi HIV. Ada lebih dari setengah infeksi baru HIV di dunia ditemukan pada usia 15-19 tahun dan mayoritas remaja terinfeksi karena hubungan seksual (*sumber: <https://pusdatin.kemkes.go.id>*) kasus penderita yang meninggal akibat HIV/AIDS juga menjadi faktor penentu kenaikan potensi tersebut. Apabila pemerintah tidak menanggulangi kejadian ini, dikhawatirkan ke depannya Indonesia akan menjadi salah satu negara penyumbang terbesar penderita HIV/AIDS.

Algoritma *K-Means* merupakan suatu algoritma yang digunakan dalam pengelompokan secara partisi yang memisahkan data ke dalam kelompok yang berbeda – beda. Algoritma ini mampu meminimalkan jarak antara data ke *clusternya*. Pada dasarnya penggunaan algoritma ini dalam proses *clustering* tergantung pada data yang didapatkan dan konklusi yang ingin dicapai di akhir proses. Penelitian sebelumnya yang telah dilakukan terkait dengan algoritma *K-Means* mengenai Analisis Klasterisasi Kasus HIV/AIDS di Indonesia dengan hasil analisis *cluster* pada data kasus HIV dan AIDS yang ada di Indonesia, dapat ditarik kesimpulan bahwa provinsi DKI Jakarta, Jawa Timur dan Papua berada dalam kondisi yang sangat kritis dan sangat rentan akan penyebaran virus HIV/AIDS. Hal itu dikarenakan jumlah kasus HIV dan AIDS terbanyak ada pada ketiga provinsi tersebut. Untuk itu, kampanye dan kebijakan pemerintah dapat difokuskan secara langsung kepada ketiga provinsi itu.

Berdasarkan latar belakang masalah diatas penelitian ini bertujuan untuk mengetahui hasil akhir perhitungan *K-Means Clustering* dalam mengelompokkan Jumlah Kasus Penyakit AIDS berdasarkan Provinsi. Dengan adanya hasil penelitian ini diharapkan kepada pemerintah pusat lebih memperhatikan provinsi yang tinggi dalam kasus penderita AIDS untuk mendapatkan tindakan lanjut.

2. Metodologi Penelitian

2.1. Data mining

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari database yang besar. Tujuan utama *data mining* adalah untuk menemukan, menggali atau menambang pengetahuan dari data atau informasi yang kita miliki [1].

Data mining adalah pengolahan data dengan skala besar, sehingga *data mining* memiliki peranan penting dalam bidang industri, keuangan, cuaca, ilmu dan teknologi. “*Data mining* juga dapat dilakukan pada berbagai jenis database dan penyimpanan informasi, namun jenis pola yang akan ditemukan ditentukan oleh berbagai fungsi *data mining* seperti deskripsi *class* / konsep, asosiasi, analisa korelasi, klasifikasi, prediksi, analisa *cluster* dan lain-lain” [2].

2.2. Algoritma K-Means

K-means merupakan suatu algoritma yang digunakan dalam pengelompokan secara partisi yang memisahkan data ke dalam kelompok yang berbeda-beda. Algoritma ini mampu meminimalkan jarak antara data ke clusternya. Pada dasarnya penggunaan algoritma ini dalam proses clustering tergantung pada data yang didapatkan dan konklusi yang ingin dicapai di akhir proses. Pada penelitian ini dipilih digunakan metode K-Means daripada metode K-Medoids karena K-Means lebih cepat dan mudah dibandingkan dengan K-Medoids yang prosesnya lama[3][4].

Langkah-langkah dalam Algoritma *K-means* :

- a) Tentukan jumlah *cluster* (*k*) pada data set
- b) Tentukan nilai pusat (*centroid*)

Penentuan nilai *centroid* pada tahap awal dilakukan secara *random*, sedangkan pada tahap *iterasi* digunakan rumus seperti pada persamaan (1) berikut ini :

$$V_{ij} = \frac{1}{N^i} \sum_{k=0}^{N^i} X_{kj} \quad (1)$$

Keterangan :

V_{ij} = *centroid* rata-rata *cluster* ke-I untuk variabel ke-j

N^i = Jumlah anggota *cluster* ke-i

i, k = indeks dari *cluster*

j = indeks dari variabel

X_{kj} = nilai data ke-k variabel ke-j untuk *cluster* tersebut

- c) Pada masing-masing *record*, hitung jarak terdekat dengan *centroid*. Jarak *centroid* yang digunakan adalah *Euclidean Distance*, dengan rumus seperti dibawah ini:

$$D(x_2, x_1) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j=1}^p |x_{2j} - x_{1j}|^2} \quad (2)$$

Keterangan :

D= *Euclidean Distance*

x = Banyaknya objek $\sum p$ = Jumlah data *record*

$$De = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (3)$$

Keterangan :

De = *Euclidean Distance*

i = Banyaknya objek ²

(*x, y*) = Koordinat objek

(*s, t*) = Koordinat *centroid*

- d) Kelompokkan objek berdasarkan jarak ke *centroid* terdekat
- e) Ulangi langkah ke-2, lakukan *iterasi* hingga *centroid* bernilai optimal.

3. Hasil dan Pembahasan

Sumber data penelitian yang digunakan pada penelitian ini diperoleh dari situs resmi <https://pusdatin.kemkes.go.id> (data dan informasi profil kesehatan Indonesia 2018) yaitu data Jumlah kasus HIV/ AIDS Menurut Provinsi di Indonesia tahun 2016-2018 yang terdiri dari 34 Provinsi. Data akan diolah dengan melakukan *clustering* dengan jumlah cluster ada 2 *cluster* yaitu tertinggi dan rendah. Pada tahapan ini dilakukan analisis data Kasus Penyakit HIV/AIDS berdasarkan Provinsi Menggunakan Metode *K-Means*. Data yang diperoleh diolah dengan menggunakan perhitungan secara manual dengan menggunakan *Microsoft Excel*. Kemudian hasil dari perhitungan manual tersebut akan diterapkan dengan menggunakan *RapidMiner*.

Tabel 1. Data Kasus Penyakit HIV/AIDS berdasarkan Provinsi


No.	Provinsi	2016	2017	2018
1	Aceh	70	111	155
2	Sumatera Utara	1891	1914	1999
3	Sumatera Barat	396	563	624
4	Riau	822	711	539
5	Jambi	215	278	246
6	Sumatera Selatan	346	486	508
7	Bengkulu	115	95	107
8	Lampung	381	580	524
9	Kep. Bangka Belitung	135	268	205
10	Kep. Riau	1037	1105	1033
11	Dki Jakarta	6019	6626	6896
12	Jawa Barat	5466	5819	5185
13	Jawa Tengah	4032	5425	5400
14	Di Yogyakarta	736	723	833
15	Jawa Timur	6513	8204	8608
16	Banten	1092	1315	1334
17	Bali	2367	2441	2211
18	Nusa Tenggara Barat	175	222	192
19	Nusa Tenggara Timur	487	837	842
20	Kalimantan Barat	525	608	692
21	Kalimantan Tengah	141	119	122
22	Kalimantan Selatan	454	572	297
23	Kalimantan Timur	813	1202	1126
24	Kalimantan Utara	163	172	166
25	Sulawesi Utara	409	516	555
26	Sulawesi Tengah	157	200	292
27	Sulawesi Selatan	993	1366	1174
28	Sulawesi Tenggara	134	134	106
29	Gorontalo	7	51	74
30	Sulawesi Barat	22	37	26
31	Maluku	621	688	462
32	Maluku Utara	120	145	200
33	Papua Barat	530	409	380
34	Papua	3866	4358	3546

a) Menentukan Nilai k Jumlah *Cluster*

Jumlah *cluster* sebanyak 2 *cluster*. *Cluster* yang dibentuk yaitu *cluster* tinggi (C1) dan rendah (C2).

b) Menentukan Nilai *Centroid* (Pusat *Cluster*)

Penentuan pusat *cluster* awal ditentukan secara *random* yang diambil dari data yang ada dalam *range*. Adapun nilai untuk *cluster* tinggi (*cluster* 1) diambil dari nilai tertinggi yang terdapat pada tabel 4.1 dan nilai untuk *cluster* rendah (*cluster* 2) diambil dari nilai terendah yang terdapat pada tabel 1. Berikut daftar tabel *centroid* data dapat dilihat pada tabel 2.:

Tabel 2. *Centroid* Data Awal

	2016	2017	2018
Cluster 1	6513	8204	8608
Cluster 2	7	37	26

c) Menghitung Jarak Setiap Data Terhadap *Centroid* (Pusat *Cluster*)

Setelah data nilai pusat *cluster* awal ditentukan, maka langkah selanjutnya adalah menghitung jarak masing-masing data terhadap pusat *cluster*. Proses pencarian jarak terpendek pada iterasi 1 dapat dilihat pada perhitungan dan tabel dibawah ini :

$$D_{Aceh,C1} = \sqrt{(70 - 6513)^2 + (111 - 8204)^2 + (155 - 8608)^2} = 13358.97$$

$$D_{Aceh,C2} = \sqrt{(70 - 7)^2 + (111 - 37)^2 + (155 - 26)^2} = 161.51$$

Hasil dari keseluruhan perhitungan dapat dilihat pada tabel 3 sebagai berikut :

Tabel 3. Hasil Perhitungan Jarak Pusat Iterasi 1

No.	Provinsi	2016	2017	2018	C1	C2	Jarak Terpendek
1	Aceh	70	111	155	13358.97	161.51	161.51
2	Sumatera Utara	1891	1914	1999	10227.70	3311.39	3311.39
3	Sumatera Barat	396	563	624	12631.18	886.34	886.34
4	Riau	822	711	539	12395.21	1175.44	1175.44
5	Jambi	215	278	246	13130.47	386.97	386.97
6	Sumatera Selatan	346	486	508	12775.34	740.84	740.84
7	Bengkulu	115	95	107	13377.49	146.93	146.93
8	Lampung	381	580	524	12691.65	826.27	826.27
9	Kep. Bangka Belitung	135	268	205	13201.11	319.04	319.04
10	Kep. Riau	1037	1105	1033	11737.25	1793.20	1793.20
11	Dki Jakarta	6019	6626	6896	2380.14	11258.60	2380.14
12	Jawa Barat	5466	5819	5185	4301.32	9478.79	4301.32
13	Jawa Tengah	4032	5425	5400	4916.25	8608.78	4916.25
14	Di Yogyakarta	736	723	833	12238.86	1285.80	1285.80
15	Jawa Timur	6513	8204	8608	0.00	13515.87	0.00
16	Banten	1092	1315	1334	11391.08	2126.35	2126.35
17	Bali	2367	2441	2211	9556.31	4015.35	4015.35
...
...
30	Sulawesi Barat	22	37	26	13508.65	15.00	15.00
31	Maluku	621	688	462	12552.42	995.44	995.44
32	Maluku Utara	120	145	200	13285.80	233.90	233.90
33	Papua Barat	530	409	380	12816.33	732.96	732.96
34	Papua	3866	4358	3546	6886.38	6778.89	6778.89

d) Menentukan Posisi *Cluster* atau Pengelompokan

Dalam menentukan posisi *cluster* masing-masing data Kasus Penyakit HIV/AIDS berdasarkan Provinsi berdasarkan jarak minimum data terhadap pusat *cluster*.

Tabel 4. Hasil *Cluster*

No.	Provinsi	Cluster
1	Aceh	2
2	Sumatera Utara	2
3	Sumatera Barat	2
4	Riau	2
5	Jambi	2
6	Sumatera Selatan	2
7	Bengkulu	2
8	Lampung	2
9	Kep. Bangka Belitung	2
10	Kep. Riau	2
11	Dki Jakarta	1
12	Jawa Barat	1
...
...
30	Sulawesi Barat	2
31	Maluku	2
32	Maluku Utara	2
33	Papua Barat	2
34	Papua	2

Proses *K-Means* akan terus beriterasi sampai pengelompokan data sama dengan pengelompokan data iterasi sebelumnya. Dengan kata lain, proses akan terus melakukan iterasi sampai data pada iterasi terakhir sama dengan iterasi sebelumnya.

e) Menghitung *centroid* baru menggunakan hasil dari setiap anggota pada masing-masing *cluster*. Setelah didapatkan hasil jarak dari setiap objek pada iterasi ke-1 maka lanjut ke iterasi ke-2 pada perhitungan dan tabel dibawah ini :

$$D_{2016,c1} = \frac{6019+5466+4032+6513}{4} = 5507.50$$

$$D_{2017,c1} = \frac{6626+5819+5425+8204}{4} = 6618.50$$

$$D_{2018,c1} = \frac{6896+5185+5400+8608}{4} = 6522.25$$

$$D_{2016,c2} = \frac{70+1891+396+822+215+346+115+381+135+1037+736+1092+2367+175+487+525+141+454+813+163+409+157+993+134+7+22+621+120+530+3866}{30} = 640.67$$

$$D_{2017,c2} = \frac{111+1914+563+711+278+486+95+580+268+1105+723+1315+2441+222+837+608+119+572+1202+172+516+200+1366+134+51+37+688+145+409+4358}{29} = 740.87$$

$$D_{2018,c2} = \frac{155+1999+624+539+246+508+107+524+205+1033+6896+5185+5400+833+8608+1334+2211+192+842+692+122+297+1126+166+555+292+1174+106+74+26+462+200+380+3546}{29} = 685.67$$

Tabel 5. *Centroid* baru iterasi 1

	2016	2017	2018
Cluster 1	5507.5	6518.5	6522.25
Cluster 2	640.67	740.87	685.67

Selanjutnya dilakukan kembali langkah ke 4 sampai 6. Jika nilai *centroid* hasil iterasi dengan nilai *centroid* sebelumnya bernilai sama serta posisi *cluster* data hasil Kasus Penyakit HIV/AIDS berdasarkan Provinsi tidak mengalami perubahan maka proses iterasi berhenti. Namun jika nilai *centroid* tidak sama serta posisi data masih berubah maka proses iterasi berlanjut pada iterasi berikutnya.

Berikut tabel hasil *cluster* iterasi 2 :

Tabel 6. *Centroid* baru untuk iterasi 2

	2016	2017	2018
Cluster 1	5507.5	6518.5	6522.25
Cluster 2	640.67	740.87	685.67

Langkah selanjutnya adalah melakukan perhitungan jarak terhadap data data Kasus Penyakit HIV/AIDS berdasarkan Provinsi dengan menggunakan data *centroid* baru

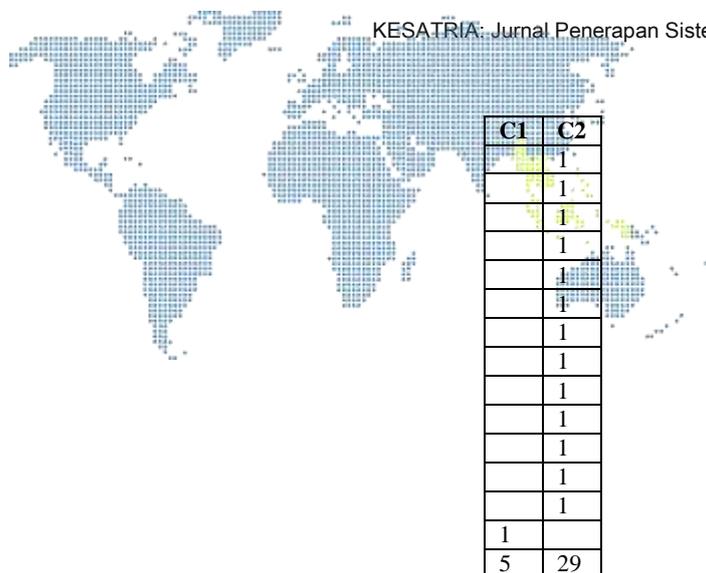
$$D_{Aceh,c1} = \sqrt{(70 - 5507.5)^2 + (111 - 6518)^2 + (155 - 6522.25)^2} = 10543.45$$

$$D_{Aceh,c2} = \sqrt{(70 - 640.67)^2 + (111 - 740.87)^2 + (155 - 685.67)^2} = 1002.00$$

Hasil dari keseluruhan perhitungan dapat dilihat pada tabel 7. sebagai berikut :

Tabel 7. Hasil Perhitungan Jarak Pusat Iterasi 2

No.	Provinsi	2016	2017	2018	C1	C2	Jarak Terpendek
1	Aceh	70	111	155	10543.45	970.34	970.34
2	Sumatera Utara	1891	1914	1999	7398.67	2198.30	2198.30
3	Sumatera Barat	396	563	624	9817.57	302.49	302.49
4	Riau	822	711	539	9564.53	202.12	202.12
5	Jambi	215	278	246	10312.41	733.27	733.27
6	Sumatera Selatan	346	486	508	9960.09	406.26	406.26



Perhitungan manual pada data diatas didapatkan hasil akhir yang dimana pada iterasi 2 dan iterasi 3 pengelompokan data yang dilakukan terhadap 2 cluster didapatkan hasil yang sama. Hasil dari kedua iterasi tersebut bernilai C1 =5 dan C2 =29 pada posisi data tiap cluster. Sehingga posisi cluster pada data tersebut tidak mengalami perubahan lagi maka proses iterasi berhenti sampai iterasi 3.

Berdasarkan dari penjelasan diatas mengenai tahap-tahap penggunaan serta hasil yang telah ditampilkan maka akan dibahas mengenai keterkaitan dari hasil yang didapat dari *RapidMiner* dan *Microsoft excel* menjelaskan bahwa hasil dari perhitungan manual algoritma *k-means* dan *Microsoft excel* data memiliki nilai yang sama yaitu antara beberapa cluster yaitu cluster tinggi 4 dan rendah 29. serta memasukan perhitungan *Microsoft excel* ke dalam *RapidMiner* memiliki nilai yang sama juga. Dapat kita lihat pada gambar 1 berikut.

Row No.	Provinsi	id	cluster	2010	2012	2015
1	ACEH	1	cluster_1	111	111	118
2	SUMATERA UTARA	2	cluster_1	3841	3819	3888
3	SUMATERA BARAT	3	cluster_1	386	383	324
4	RIAU	4	cluster_1	322	711	538
5	JAWA	5	cluster_1	212	270	246
6	SUMATERA SELATAN	6	cluster_1	246	486	588
7	BENGKALU	7	cluster_1	112	85	187
8	LAMPUNG	8	cluster_1	301	583	524
9	KEP. BANGKA BELITUNG	9	cluster_1	125	288	285
10	KEP. RIAU	10	cluster_1	637	1405	1033
11	DIY JAGATTA	11	cluster_0	0018	628	600
12	JAWA BARAT	12	cluster_0	5488	5918	5188
13	JAWA TENGAH	13	cluster_0	4032	5420	5430
14	DI YOGYAKARTA	14	cluster_0	756	725	835
15	JAWA TIMUR	15	cluster_0	5513	6204	6620
16	BAHARA	16	cluster_1	2082	1315	1354
17	BALI	17	cluster_1	2387	2441	2211
18	SULAWESI BARAT	18	cluster_1	179	322	182
19	SULAWESI TENGAH	19	cluster_1	487	837	842
20	KALIMANTAN BARAT	20	cluster_1	626	888	882
21	KALIMANTAN TENGAH	21	cluster_1	741	119	122

Gambar 1. Tampilan Data Perhitungan *RapidMiner*

4. Kesimpulan

- Berdasarkan hasil dan pembahasan sebelumnya menghasilkan beberapa kesimpulan dari skripsi ini sebagai berikut :
- Hasil penelitian yang diperoleh dapat disimpulkan bahwa perhitungan manual excel yang di kombinasikan dengan *RapidMiner*, bahwa metode tersebut mampu untuk menyelesaikan sebuah pengelompokan berdasarkan provinsi.

c) Hasil yang diperoleh dari metode K-Means Clustering yang di implementasikan ke dalam RapidMiner memiliki nilai validasi yang sama yaitu menghasilkan beberapa cluster yaitu cluster tinggi ada 4 diantaranya DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur dan Papua, sedangkan untuk cluster rendah ada 30 yaitu Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara dan Papua Barat. Hasil yang didapat dari penelitian dapat menjadi masukan kepada pemerintah, yang menjadi perhatian lebih pada Provinsi yang memiliki sarana kesehatan berdasarkan cluster yang telah dilakukan.

Daftar Pustaka

- [1] R. Setiawan And N. Tes, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus : Politeknik Lp3i Jakarta)," Vol. 3, No. 1, Pp. 76-92, 2016.
- [2] Y. Darmi And A. Setiawan, "Penerapan Metode Clustering K-Means Dalam," *J. Media Infotama*, Vol. 12, No. 2, Pp. 148-157, 2016.
- [3] I. Parlina, A. P. Windarto, A. Wanto, And M. R. Lubis, "Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Asessment Center Untuk Clustering Program Sdp," *Cess (Journal Comput. Eng. Syst. Sci.*, Vol. 3, No. 1, Pp. 87-93, 2018.
- [4] L. Teori, "(K-Means Algorithm Implementation For Clustering Of Patients Disease In Kajein Clinic Of Pekalongan) Anindya Khrisna Wardhani Magister Sistem Informasi Universitas Diponegoro," Vol. 14, Pp. 30-37, 2016.