ALGORITMA PARTITIONING AROUND MEDOIDS DALAM MENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN INDEKS KINERJA DAVIES BOULDIN PADA KASUS PENYAKIT HIV

Nisfu Rohmatul Aeni[1], Agung Nilogiri[2], Reni Umilasari[3]

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember

Jln. Karimata No. 49 Jember Kode Pos 68121

email: nisfurohmatul14@gmail.com

ABSTRAK

Clustering adalah proses mengelompokkan sejumlah data kedalam cluster, sehingga dalam setiap cluster akan berisi data yang semirip mungkin. Salah satu algoritma clustering adalah Partitioning Around Medoids (PAM). Algoritma PAM menggunakan objek pada sekumpulan objek untuk mewakili sebuah cluster, dan objek yang terpilih untuk mewakili sebuah cluster. Pada penelitian ini membahas tentang penerapan algoritma PAM untuk mengelompokkan provinsi di Indonesia berdasarkan kasus HIV pada tahun 2015 sampai dengan 2018. Data yang dijadikan sumber penelitian bersumber dari situs website resmi Kementerian Kesehatan Republik Indonesia. Hasil dari penerapan algoritma PAM dalam mengelompokkan provinsi di Indonesia pada kasus penyakit HIV diperoleh cluster optimum dengan hasil 3 cluster berdasarkan indeks nilai Davies Bouldin sebesar 0,453 dengan skenario 2 cluster sampai dengan 10 cluster. Sedangkan jumlah anggota pada masing cluster yaitu cluster 1 terdapat 4 provinsi, cluster 2 terdapat 28 provinsi, dan cluster 3 terdapat 2 provinsi.

Kata Kunci: HIV, Clustering, PAM, Davies Bouldin.

ABSTRACT

Clustering is the process of grouping a number of data into clusters, so that each cluster will contain the same data. One of the grouping algorithms is Partitioning Around Medoids (PAM). The PAM algorithm uses a collection of objects to represent a cluster, and the object is chosen to represent a cluster. In this study discusses the application of the Partitioning Around Medoids algorithm to classify provinces in Indonesia based on HIV cases in 2015 to 2018. The data is used as a source of research sourced from the official website of the Ministry of Health of the Republic of Indonesia. The results of applying the PAM method classify that the province in Indonesia in the case of HIV disease is obtained by an optimal cluster with the results of 3 clusters based on the Davies Bouldin index value from 0.453 with 2 cluster scenarios to 10 clusters. While the number of members in each cluster is cluster 1, there are 4 provinces, cluster 2 has 28 provinces, and cluster 3 has 2 provinces.

Keywords: HIV, Clustering, PAM, Davies Bouldin.

1. PENDAHULUAN

HIV adalah kepanjangan dari *Human Immunodeficiency Virus*. Virus HIV menyerang sistem kekebalan tubuh manusia dan melemahkan kemampuan tubuh untuk melawan penyakit. HIV adalah penyakit mematikan yang belum bisa disembuhkan, tetapi terdapat pengobatan yang dapat dilakukan untuk memperlambat perkembangan penyakit HIV. Sehingga pengobatan HIV dapat membuat penderitanya hidup lebih lama dan cukup efektif untuk mencegah berkembangnya sel baru HIV. Penyakit HIV berasal dari negara Afrika yang ditemukan pertama kali pada tubuh kera dan pada akhirnya virus tersebut menular ke dalam tubuh manusia, sehingga virus HIV menyebar ke berbagai negara. Faktor utama dari virus HIV menyebar yaitu kurangnya pengetahuan masyarakat tentang virus HIV, sehingga masyarakat berasumsi bahwa penyakit HIV merupakan penyakit yang tidak serius.

Virus HIV di Indonesia ditemukan pada tahun 1987 yang tersebar di 368 kabupaten/kota di seluruh provinsi. Menurut UNAIDS (*Joint United Nations Programme on HIV/AIDS*), di Indonesia terdapat 690.000 orang terinfeksi HIV sampai tahun 2015. Pada tahun 2014 negara Indonesia mendapatkan peringatan dari UNAIDS, karena Indonesia dianggap kurang memperhatikan pasien penderita penyakit HIV dan angka kematian penderita penyakit HIV semakin tinggi. Sedangkan menurut data Kementerian Kesehatan Republik Indonesia jumlah kasus HIV pada tahun 2005 sampai dengan tahun 2018 mengalami kenaikan di setiap tahunnya. Jumlah kasus HIV yang dilaporkan sampai dengan bulan Desember 2018 sebanyak 327.282 kasus.

Salah satu penyebab perkembangan infeksi HIV adalah kurangnya pengetahuan masyarakat tentang cara penularan dari virus HIV (Kemenkes, 2010). Informasi yang kurang kepada masyarakat akan menimbulkan masalah yang serius. Masyarakat yang penasaran dengan penyakit HIV akan mencari tahu tentang penyakit HIV terutama pada cara penularannya dengan melalui berbagai media, diantaranya internet, majalah atau media lainnya yang belum tentu memberikan informasi yang benar. Hal tersebut menyebabkan masyarakat terjerumus pada pengetahuan HIV yang salah. Maka berdampak pada jumlah kasus penderita HIV yang semakin meningkat setiap tahunnya. Upaya yang dapat dilakukan untuk mengantisipasi penularan dari HIV adalah dengan cara pencegahan, karena infeksi virus HIV pada saat ini belum ditemukan obat yang efektif untuk pengobatannya. Program pemerintah dalam melakukan pencegahan yang dilakukan melalui pendidikan kesehatan dan peningkatan pengetahuan yang benar tentang virus HIV dan cara penularannya. Pemberian penyuluhan dan sosialisasi tentang penyakit HIV pada masyarakat telah diwujudkan dalam program tahunan Kementrian Kesehatan pada masing-masing provinsi di Indonesia.

Pada penelitian sebelumnya yang dilakukan oleh (Riveranda, 2016) dengan studi kasus Analisis Klasterisasi kasus HIV/AIDS di Indonesia menggunakan metode *K-means*. Tujuan utama dari penelitian tersebut untuk menganalisis kasus HIV/AIDS di Indonesia dengan penyebaran di 33 provinsi. Data yang digunakan adalah HIV/AIDS dari tahun 2009 sampai dengan 2014 yang diperoleh dari Ditjen PP & PL Kemenkes RI. Penelitian tersebut menghasilkan jumlah *cluster* yang ditentukan adalah 3 *cluster* dan tidak menggunakan indeks pengukuran dalam menentukan *cluster* optimal. Setiap *cluster* menunjukan label kerawanan terjadinya HIV di setiap provinsi di Indonesia. *Cluster* 1 memiliki karakteristik data dengan anggota *cluster* terendah dan *cluster* 2 memiliki karakteristik data dengan anggota *cluster* sedang. Sedangkan *cluster* 3 memiliki karakteristik data dengan anggota *cluster* tertinggi. Berdasarkan latar belakang tersebut, penulis melakukan suatu penelitian dengan mengelompokkan provinsi di Indonesia berdasarkan kasus HIV dari tahun 2015 sampai dengan 2018. Pada penelitian ini, penulis menggunakan algoritma *Partitioning Around Medoids* berdasarkan pengukuran indeks kinerja *Davies Bouldin* dalam menentukan *cluster* optimum.

2. TINJAUAN PUSTAKA

2.1 Data *Mining*

Data *mining* merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar (Turban dkk, 2005). Data *mining* adalah analisis dari peninjauan kumpulan data untuk menemukan hubungan tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dengan bermanfaat bagi pemilik data (Larose, 2005).

Sehingga, pengertian dari data *mining* adalah proses menganalisa data dengan menggunakan pencarian pola data yang tidak diketahui atau tidak dapat di perkerikan sebelumnya.

2.2 Clustering

Clustering (pengelompokan data) mempertimbangkan sebuah pendekatan penting untuk mencari kesamaan dalam data dan menempatkan data yang sama ke dalam kelompok-kelompok. *Clustering* membagi kumpulan data ke dalam beberapa kelompok dimana kesamaan dalam sebuah kelompok adalah lebih besar dari pada kelompok-kelompok yang lain (RuiXu dan Donald, 2009). Tujuan dari *clustering* yaitu untuk meminimalkan jarak di dalam *cluster* dan memaksimalkan jarak antar *cluster*.

Clustering merupakan suatu proses pengelompokan *record*, observasi, atau mengelompokkan kelas yang memiliki kesamaan objek (Larose, 2006). Sehingga definisi dari *clustering* yaitu suatu kumpulan dari keseluruhan pengelompokan. Perbedaan *clustering* dengan klasifikasi yaitu tidak adanya variabel target dalam melakukan pengelompokan data pada proses *clustering*.

2.3 Algoritma *Partitioning Around Medoids*

Algoritma *Partitioning Around Medoids* menggunakan metode partisi *clustering* untuk mengelompokkan sekumpulan *n* objek menjadi sejumlah *k cluster*. Algoritma *Partitioning Around Medoids* (PAM) atau disebut dengan *K-Medoids* merupakan algoritma pengelompokan yang berkaitan dengan algoritma *K-Means* dan algoritma *medoidshift*. Algoritma *Partitioning Around Medoids* (PAM) dikembangkan oleh Leonard Kaufman dan Peter J. Rousseeuw. Algoritma ini sangat mirip dengan algoritma *K-Means*. Perbedaan dari kedua algoritma ini adalah algortima PAM menggunakan objek sebagai perwakilan (*medoid*) dalam pusat *cluster* untuk setiap *cluster,* sedangkan *K-Means* menggunakan nilai rata-rata (*mean*) sebagai pusat *cluster* (Kaur dkk, 2014).

Algoritma *Partitioning Around Medoids* memiliki kelebihan untuk mengatasi kelemahan dari algoritma *K-Means* yang sensitif terhadap *noise* atau *outlier*, dan objek dengan nilai yang besar yang memungkinkan menyimpang dari distribusi data, sedangkan kelebihan lainnya yaitu hasil proses *clustering* tidak bergantung pada urutan masuk dataset (Furqon dkk, 2015). Langkah-langkah dari algoritma PAM (Han and Kamber, 2006), yaitu:

1) Menentukan *k* objek pada sekumpulan *n* objek sebagai *medoid* dengan memilih secara acak.
2) Kemudian secara acak pilih o_random : sebuah objek *non-medoids.*
3) Menempatkan objek *non-medoids* ke dalam *cluster* yang paling dekat dengan *medoids*.
4) Menghitung total *cost*, *S* (selisih), dari pertukaran *medoids* o_j dengan o_{random}.

Nilai total biaya atau cost dinyatakan dengan persamaan:

$$Total\ Cost = \sum \sqrt{\sum_{k=1}^{n}(x_k - y_k)^2} \quad (1)$$

Dengan:

n = jumlah suatu data

k = indeks data

x_k = nilai atribut ke-*k* dari *x*

y_k = nilai atribut ke-*k* dari *y*

Nilai S dinyatakan dalam persamaan:

$$S = Total\ Cost\ baru - Total\ Cost\ Lama \quad (2)$$

Dengan:

S = Selisih

$Total\ Cost\ Baru$ = jumlah c*ost non-medoids*

$Total\ Cost\ Lama$ = jumlah *cost medoid*s

5) Jika $S < 0$ maka tukar o_j dengan o_{random} untuk membentuk sekumpulan k objek baru sebagai *medoids*.

6) Ulangi langkah 2 sehingga hasil dari langkah 5 tidak ada perubahan.

2.4 Davies Bouldin Indeks

Davies Bouldin Index (DBI) merupakan cara validasi *cluster* yang ditemukan oleh D.L.Davies. DBI adalah fungsi rasio dari jumlah distribusi didalam *cluster* untuk pemisahan antar *cluster*. Dalam penelitian ini, DBI digunakan untuk melakukan validasi data pada setiap *cluster*. Adapun langkah-langkah untuk menghitung *Davies Bouldin Index* (Retno, 2019), yaitu:

1. Menghitung *Sum of Square Within cluster* (SSW).

   Sum of Square Within cluster (SSW) merupakan persamaan yang digunakan untuk mengetahui matriks kohesi dalam sebuah *cluster* ke-*i* yang dirumuskan sebagai berikut:

   $$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j C_i) \quad (3)$$

   Dengan:

   m_i = jumlah data dalam *cluster* ke-*i*

   C_i = *centroid cluster* ke-*i*

   $d(x_j C_j)$ = merupakan jarak setiap data terhadap *centroid*

2. Menghitung *Sum of Square Between cluster* (SSB).

   Sum of Square Between cluster (SSB) merupakan persamaan yang digunakan untuk mengetahui separasi antar *cluster* yang dihitung menggunakan persamaan sebagai berikut:

   $$SSB_{i,j} = d(c_i, c_j) \quad (4)$$

3. Menghitung jumlah *ratio* (rasio).

   Setelah nilai SSW dan SSB diperoleh, kemudian melakukan pengukuran rasio (Rij) untuk mengetahui nilai perbandingan antara *cluster* ke-*i* dan *cluster* ke-*j.* nilai rasio dihitung menggunakan persamaan sebagai berikut:

   $$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \quad (5)$$

4. Menghitung nilai *Davies Bouldin Index* (DBI).

   Nilai rasio yang diperoleh tersebut digunakan untuk mencari nilai *Davies*

Bouldin Indeks (DBI) dengan menggunakan persamaan berikut:

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max_{i \neq j}(R_{ij}) \qquad (6)$$

Dengan:

k = jumlah *cluster* yang ditentukan

Rij = rasio *cluster* ke-*i* dan *cluster* ke-*j*

Skema *clustering* yang optimal menurut indeks pengukuran *Davies Bouldin* adalah *cluster* yang memiliki nilai indeks terkecil atau minimal (Salazar dkk, 2002).

2.5 Rapidminer

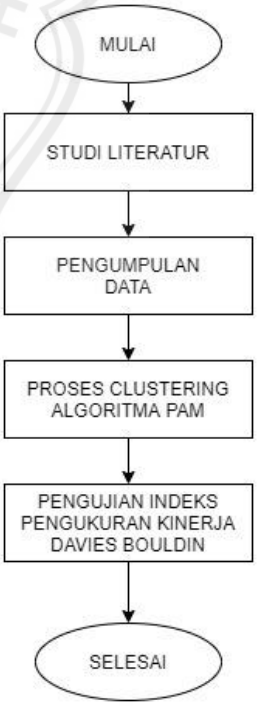
Rapidminer merupakan perangkat lunak yang bersifat terbuka (*open source*). *Rapidminer* adalah sebuah solusi untuk melakukan analisis terhadap data *mining*, *text mining* dan analisis prediksi (Dennis dkk, 2013). *Rapidminer* menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. *Rapidminer* memiliki beberapa sifat sebagai berikut.

1. Ditulis dengan bahasa pemrograman *java* sehingga dapat dijalankan di berbagai sistem operasi.
2. Proses penemuan pengetahuan dimodelkan sebagai operator *trees*.
3. Representasi XML internal untuk memastikan format standar pertukaran data.
4. Bahasa *scripting* memungkinkan untuk eksperimen skala besar dan otomatisasi eksperimen.
5. Konsep *multilayer* untuk menjamin tampilan data yang efisien dan menjamin penanganan data.
6. Memiliki GUI, *command line mode*, dan *java* API yang dapat dipanggil dari program lain.

3. METODOLOGI

3.1 Tahapan Penelitian

Adapun tahapan penelitian yang dilakukan dalam pemanfaatan algoritma *Partitioning Around Medoids* (PAM) untuk mengelompokkan provinsi di Indonesia berdasarkan indeks kinerja *Davies Bouldin* pada kasus penyakit HIV, memiliki tahapan sebagai berikut.



Gambar 1. Diagram Alur Penelitian

3.2 Studi *Literatur*

Tahapan awal dari penelitian ini adalah dengan mencari dan mempelajari masalah yang akan diteliti, kemudian menentukan ruang lingkup masalah, latar belakang, dan mempelajari beberapa literatur yang berkaitan dengan permasalahan dan bagaimana mencari solusi dari masalah tersebut. Untuk mencapai tujuan yang ditentukan, maka penulis perlu mempelajari beberapa *literetur* yang digunakan, kemudian *literatur* tersebut diseleksi untuk ditentukan sebagai *literatur* yang akan digunakan dalam penelitian.

3.3 Metode Analisis Data

Pada penelitian ini, metode analisa data yang digunakan adalah *Partitioning Around Medoids*. *Partitioning Around Medoids* merupakan teknik *cluter* atau mengelompokkan sekumpulan objek sebagai perwakilan (*medoid*) dalam pusat *cluster* untuk setiap *cluster*. Kelebihan dari PAM yaitu untuk mengatasi kelemahan dari algoritma *K-Means* yang sensitif terhadap *noise* atau *outlier*, dan objek dengan nilai yang besar yang memungkinkan menyimpang dari distribusi data.

Gambar 2. *Flowchart* Algoritma PAM

3.4 Pengumpulan Data

Penelitian ini menggunakan data jumlah kasus HIV dari 34 provinsi di Indonesia, data tersebut memilki 10 atribut yang berupa data dari tahun 2009 sampai dengan 2018 yang diunduh dari situs https://siha.depkes.go.id/portal/files upload/Laporan_Triwulan_IV_2018.pdf. Dalam penelitian ini, penulis menggunakan data dari 34 provinsi dengan 4 atribut yang berupa data dari tahun 2015 sampai dengan 2018.

3.5 Dataset

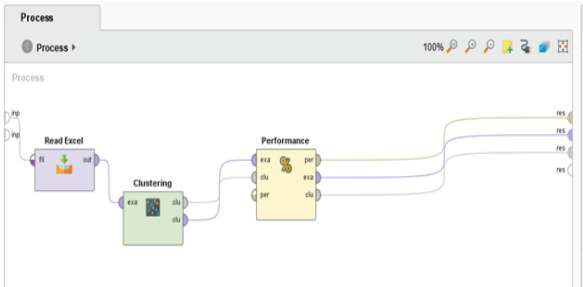
Data yang digunakan adalah data jumlah kasus HIV dari 34 provinsi di Indonesia oleh Kementerian Kesehatan Republik Indonesia. Dalam penelitian ini, penulis menggunakan data dari 34 provinsi dengan 4 atribut yang berupa data dari tahun 2015 sampai dengan 2018.

4. HASIL DAN PEMBAHASAN

4.1 Data Pengujian

Penelitian ini menggunakan data jumlah kasus HIV dari 34 provinsi di Indonesia, data tersebut memilki 10 atribut yang berupa data dari tahun 2009 sampai dengan 2018 yang diunduh dari situs https://siha.depkes.go.id/portal/files upload/Laporan_Triwulan_IV_2018.pdf. Dalam penelitian ini, penulis menggunakan data dari 34 provinsi dengan 4 atribut yang berupa data dari tahun 2015 sampai dengan 2018.

4.2 *Rapidminer* Studio



Gambar 3. Proses *Rapidminer*

Berdasarkan gambar 3 terdapat beberapa operator yang digunakan. Berikut fungsi dari operator-operator tersebut:

a. *Read Excel:* operator tersebut digunakan untuk memuat data dari *spreadsheet Microsoft Excel*.
b. *Clustering*: operator tersebut digunakan untuk melakukan pengelompokkan menggunakan metode *clustering Partitioning Around Medoids* (PAM). Pada penelitian ini jumlah *cluster* yang akan digunakan sebagai pengujian yaitu 2 *cluster* sampai 10 *cluster*.
c. *Performance*: operator tersebut digunakan untuk evaluasi kinerja metode PAM. Pada proses *performance* akan menampilkan hasil evaluasi kinerja berdasarkan nilai *Davies Bouldin.*

4.3 Penentuan Jumlah *Cluster* Optimum

Setelah melakukan proses *cluster* menggunakan *Partitioning Around Medoids*, kemudian melakukan proses penentuan jumlah *cluster* optimum dengan menggunakan teknik *Davies Bouldin*.

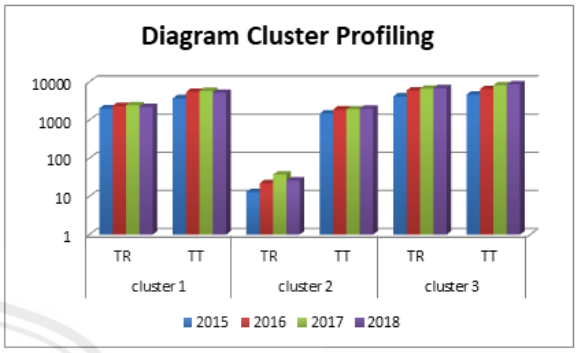
Tabel 1. Nilai *Davies Bouldin Indeks*

cluster	Nilai DBI	*Cluster* yang dihasilkan	Anggota *Cluster*
2	0,613	2	*Cluster* 1 : 6 dan *Cluster* 2 : 28
3	**0,453**	**3**	***Cluster* 1 : 4, *Cluster* 2 : 28, dan *Cluster* 3 : 2**
4	0,737	4	*Cluster* 1 : 2, *Cluster* 2 : 28, *Cluster* 3 : 2, dan *Cluster* 4 : 2
5	0,700	5	*Cluster* 1 : 2, *Cluster* 2 : 3, *Cluster* 3 : 12, *Cluster* 4 : 11, dan *Cluster* 5 : 6
6	0,686	6	*Cluster* 1 : 10, *Cluster* 2 : 6, *Cluster* 3 : 2, *Cluster* 4 : 3, *Cluster* 5 : 2, dan *Cluster* 6 : 11
7	1,189	7	*Cluster* 1 : 3, *Cluster* 2 : 12, *Cluster* 3 : 1, *Cluster* 4 : 2, *Cluster* 5 : 10, *Cluster* 6 : 1, dan *Cluster* 7 : 5
8	1,132	8	*Cluster* 1 : 3, *Cluster* 2 : 1, *Cluster* 3 : 10,

			Cluster 4 : 1, *Cluster* 5 : 2, *Cluster* 6 : 2, *Cluster* 7 : 10, dan *Cluster* 8 : 5
9	0,498	9	*Cluster* 1 : 2, *Cluster* 2 : 1, *Cluster* 3 : 2, *Cluster* 4 : 1, *Cluster* 5 : 2, *Cluster* 6 : 11, *Cluster* 7 : 2, *Cluster* 8 : 3, dan *Cluster* 9 : 10
10	1,091	10	*Cluster* 1 : 1, *Cluster* 2 : 10, *Cluster* 3 : 2, *Cluster* 4 : 4, *Cluster* 5 : 1, *Cluster* 6 : 2, *Cluster* 7 : 1, *Cluster* 8 : 2, *Cluster* 9 : 5, dan *Cluster* 10 : 6

Berdasarkan Tabel 1 menunjukkan nilai *Davies Bouldin* dengan pengujian skenario *cluster* 2 sampai dengan 10 *cluster*, maka jumlah *cluster* yang paling optimal berdasarkan nilai *Davies Bouldin* indeks terkecil yaitu 3 *cluster* dengan nilai *Davies Bouldin* yaitu 0,453.

4.4 Hasil *Cluster Profiling*



Dari hasil karakteristik kasus penyakit HIV didapatkan hasil fitur di tahun 2015 sampai dengan tahun 2018 pada *cluster* 2 memiliki kasus HIV terendah dibandingkan dengan *cluster* 1 dan *cluster* 3, sedangkan fitur di tahun 2015 sampai dengan tahun 2018 pada *cluster* 3 memiliki kasus HIV tertinggi dibandingkan dengan *cluster* 1 dan *cluster* 2.

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

1) Hasil dari penerapan metode *Partitioning Around Medoids* dalam mengelompokkan provinsi di Indonesia pada kasus penyakit HIV diperoleh *cluster* optimum dengan hasil 3 *cluster* berdasarkan indeks nilai *Davies Bouldin* sebesar 0,453 dengan skenario *cluster* 2 sampai dengan 10 *cluster*.
2) Hasil dari pengelompokkan dari 3 *cluster* pada *cluster* 1 terdapat 4 provinsi terdiri dari Jawa Barat, Jawa Tengah, Bali, dan Papua. Pada *cluster* 2 terdapat 28 provinsi terdiri dari NAD, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bangka Belitung, Kepulauan Riau, DI Yogyakarta, Banten, NTB, NTT, Kalimantan Barat, Kalimantan

Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, dan Papua Barat. Sedangkan pada *cluster* 3 terdapat 2 provinsi terdiri dari DKI Jakarta, dan Jawa Timur.

3) Hasil dari analisis tingkat *cluster* berdasarkan hasil *cluster profiling* menunjukkan label tingkat kerawanan kasus terjadinya HIV dalam setiap provinsi di Indonesia, pada *cluster* 1 memiliki karakteristik data dengan anggota-anggota data yang jumlah kasus HIV-nya sedang. Pada *cluster* 2 memiliki karakteristik data dengan anggota-anggota data yang jumlah kasus HIV-nya terendah. Sedangkan pada *cluster* 3 memiliki karakteristik data dengan anggota-anggota data yang jumlah kasus HIV-nya tertinggi.

5.2 Saran

1) Penelitian ini hanya sebagai penerapan dari algoritma *Partitioning Around Medoids,* agar lebih bermanfaat maka disarankan membangun Sistem Informasi sebagai media implementasi algoritma *Partitioning Around Medoids*.
2) Penelitian ini dapat dikembangkan menggunakan metode *clustering* yang berbeda, untuk uji validasi *cluster* dapat menggunakan teknik yang berbeda seperti metode *elbow*, *silhoute coefisien*. Dan dapat meningkatkan jumlah atribut dan variabel pada data yang digunakan.

6. DAFTAR PUSTAKA

[1] Dennis, A., Baskoro, D.A., Ambarwati L., Wicaksana I.W.S. 2012. Belajar Data Mining dengan RapidMiner. Jakarta: Gramdia Pustaka Utama.

[2] Han, J., & Kamber, M. 2006. Data Mining : Concept and Techniques, 2nd ed. San Fransisco: Morgan Kauffman.

[3] Kaur, N. U., & Singh, D. 2014. K-Medoids Clustering Algorithm – A Review. [pdf] International Journal of Computer Application and Technology (IJCAT). ISSN. 2349-1841 Vol. 1, Issue 1. April 2014.

[4] Kemenkes Kesehatan RI, 2010. Penyebab perkembangan virus HIV. Jakarta.

[5] Larose, D.T. 2005. Discovering Knowledge in Data: An Introduction to Data Mining. John Willey & Sons, Inc.

[6] Larose, D.T. 2006. DATA MINING METHODS AND MODELS, New Jersey: John Wiley & Sons, Inc.

[7] RuiXu and Donald, C.W. 2009. Clustering, A John Wiley & Sons, Inc., Publication

[8] Retno, S. 2019. Peningkatan Akurasi Algortima K-Means dengan Clustering Purity sebagai Titik Pusat Cluster Awal (Centroid).[pdf] repository.usu.ac.id. [diakses 9 Desember 2019].

[9] Turban, E., Ramesh, S., & Dursun, D. 2005. Decision Support Systems and Intelligent Systems. Yogyakarta: Andi Offset.