

ALGORITMA FUZZY C-MEANS DENGAN METODE ELBOW UNTUK MENGELOMPOKKAN PROVINSI DI INDONESIA BERDASARKAN JUMLAH KASUS TERINFEKSI HIV

¹Ade Yugo Dwinovit (1610651072), ²Agung Nilogiri, ³Habibatul Azizah Al Faruq

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
e-mail: ¹adeyugodwinovit@gmail.com

ABSTRAK

HIV (*Human Immunodeficiency Virus*) adalah virus yang menyerang sistem kekebalan tubuh. Jika seseorang dengan HIV positif tidak menjalani pengobatan tertentu, akan mengalami kumpulan gejala infeksi yang disebabkan oleh penurunan kekebalan tubuh akibat tertular virus HIV, yang disebut AIDS (*Acquired Immunodeficiency Syndrome*). Menurut data Kementerian Kesehatan Republik Indonesia, jumlah kasus HIV di Indonesia terus meningkat, sementara jumlah kasus AIDS mengalami peningkatan yang tidak signifikan dan relatif stabil. Berdasarkan penelitian sebelumnya, maka penelitian ini dilakukan untuk mengelompokkan provinsi di Indonesia berdasarkan jumlah kasus HIV dengan menggunakan metode *clustering* yaitu algoritma *Fuzzy C-Means*. Untuk pengukuran *cluster* optimum dalam menentukan *cluster* terbaik, metode yang digunakan adalah metode *Elbow*. Data yang digunakan yaitu jumlah kasus HIV di 34 provinsi di Indonesia periode tahun 2015-2018. Dari serangkaian pengujian mulai dari 2 *cluster* sampai 10 *cluster*, dihasilkan *cluster* optimum berada pada 2 *cluster* berdasarkan jarak SSE (*Sum of Squares Error*) pada metode *Elbow*. Pada *cluster* 1 terdiri dari 29 anggota provinsi dan *cluster* 2 terdiri dari 5 anggota provinsi. Berdasarkan hasil karakteristik data tahun 2015 sampai dengan tahun 2018, *cluster* 1 memiliki jumlah kasus HIV lebih rendah dibandingkan dengan *cluster* 2.

Kata Kunci: HIV, clustering, fuzzy c-means, elbow.

ABSTRACT

HIV (*Human Immunodeficiency Virus*) is a virus that attacks the body's immune system. If someone with HIV is not undergoing certain treatments, will have some of symptoms of infection caused by decreased immunity due to contracting the HIV virus, called AIDS (*Acquired Immunodeficiency Syndrome*). According to data from the Ministry of Health of the Republic of Indonesia, the number of HIV cases in Indonesia continues to increase, while the number of AIDS cases has increased insignificantly and relatively stable. Based on previous research, this study was conducted to classify provinces in Indonesia based on the number of HIV cases using the clustering method, namely the *Fuzzy C-Means* algorithm. For optimum cluster measurement in determining the best cluster, the method used is the *Elbow* method. The data used is the number of HIV cases in 34 provinces in Indonesia in the period 2015-2018. From a test series ranging from 2 clusters to 10 clusters, optimum clusters are produced in 2 clusters based on the distance of SSE (*Sum of Squares Error*) on the *Elbow* method. In cluster 1 consists of 29 members of the province and cluster 2 consists of 5 members of the province. Based on the results of data characteristics from 2015 to 2018, cluster 1 has a lower number of HIV cases than cluster 2.

Keywords: HIV, clustering, fuzzy c-means, elbow.

1. PENDAHULUAN

Tubuh manusia memiliki sel darah putih yang berguna sebagai pertahanan tubuh dari serangan virus maupun bakteri. Virus HIV (*Human Immunodeficiency Virus*) yang masuk ke tubuh manusia dapat melemahkan bahkan mematikan sel darah

putih dan memperbanyak diri, sehingga dapat melemahkan sistem kekebalan tubuh. Dalam kurun waktu tertentu jika seseorang dengan HIV positif tidak menjalani pengobatan tertentu, akan mengalami kumpulan gejala infeksi yang disebabkan oleh penurunan kekebalan

tubuh akibat tertular virus HIV, yang disebut AIDS (*Acquired Immunodeficiency Syndrome*). Pengobatan tersebut digunakan untuk memperlambat perkembangan penyakit dan membuat masa hidup lebih lama, sehingga bisa menjalani hidup dengan normal. Penderita virus HIV ini dapat dengan mudah menularkannya ke individu lain melalui perilaku seks bebas, penggunaan jarum suntik bekas, transfusi darah, dan melalui air susu ibu atau ASI.

Permasalahan infeksi virus HIV ini menjadi tantangan kesehatan hampir di seluruh dunia, termasuk di Indonesia. Jumlah kasus HIV di Indonesia terus meningkat, sementara jumlah kasus AIDS mengalami peningkatan yang tidak signifikan dan relatif stabil. Menurut data Kementerian Kesehatan Republik Indonesia, jumlah kasus infeksi HIV di Indonesia pada tahun 2015 terdapat 30.935 kasus, pada tahun 2016 meningkat menjadi 41.250 kasus, pada tahun 2017 meningkat menjadi 48.300 kasus, dan pada tahun 2018 sedikit menurun menjadi 46.659 kasus. Untuk kasus AIDS, pada tahun 2015 terdapat 9.215 kasus, pada tahun 2016 terdapat 10.146 kasus, pada tahun 2017 terdapat 10.488 kasus, dan pada tahun 2018 terdapat 10.190 kasus.

Pada penelitian sebelumnya yang dilakukan oleh (Riveranda & Nengsih, 2016) dengan studi kasus “K-Means Analysis Klasterisasi Kasus HIV/AIDS di Indonesia” menggunakan data HIV/AIDS sampai dengan tahun 2014. Pada penelitian tersebut, peneliti tidak menggunakan pengukuran *cluster* optimum dalam penentuan *cluster* terbaik. Dari penelitian tersebut menghasilkan data analisa jumlah kasus HIV/AIDS dengan 33 provinsi di Indonesia (tidak menggunakan Provinsi Kalimantan Utara) dibagi menjadi 3 *cluster*. Penelitian lainnya yang dilakukan oleh (Saputra & Riksakomara, 2018) dengan studi kasus “Implementasi *Fuzzy C-Means* dan Model RFM untuk

Segmentasi Pelanggan (Studi Kasus : PT. XYZ)” menggunakan algoritma *clustering Fuzzy C-Means* dan metode *Elbow* untuk mencari *cluster* terbaik. Karena jumlah kasus HIV yang terus meningkat, penulis ingin melakukan penelitian terhadap data jumlah kasus HIV dari tahun 2015 sampai dengan tahun 2018 di 34 provinsi di Indonesia (termasuk provinsi baru yang baru terbentuk tahun 2012 yaitu Kalimantan Utara). Algoritma yang digunakan oleh penulis adalah algoritma *Fuzzy C-Means* dan menggunakan metode *Elbow* untuk menentukan *cluster* optimum dalam penentuan *cluster* terbaik. Maka dari itu, penelitian ini dibuat dengan judul “Algoritma *Fuzzy C-Means* dengan Metode *Elbow* untuk Mengelompokkan Provinsi di Indonesia Berdasarkan Jumlah Kasus Terinfeksi HIV”.

2. TINJAUAN PUSTAKA

2.1 Provinsi di Indonesia

Indonesia adalah negara kepulauan yang terletak di kawasan Asia Tenggara. Wilayah Indonesia terbagi menjadi beberapa provinsi. Menurut Peraturan Menteri Dalam Negeri Nomor 137 Tahun 2017, jumlah provinsi di Indonesia saat ini ada 34 provinsi. Sebanyak 34 provinsi di Indonesia dibagi dalam beberapa kelompok pulau. Ada 10 provinsi di pulau Sumatera, 6 provinsi di pulau Jawa, 3 provinsi di pulau Bali dan Kepulauan Nusa Tenggara, 5 provinsi di pulau Kalimantan, 6 provinsi di pulau Sulawesi serta 4 provinsi di Kepulauan Maluku dan Papua.

Provinsi dengan daerah terluas adalah Papua, yakni mencapai 319.036 km², diikuti Kalimantan Tengah dengan luas 153.564 km² di urutan kedua. Provinsi dengan luas daerah terkecil adalah DKI Jakarta, yakni 664 km² (Badan Pusat Statistika, 2017).

2.2 HIV

HIV (*Human Immunodeficiency Virus*) adalah virus yang menyerang sistem

kekebalan tubuh. Virus ini dapat melemahkan kemampuan tubuh melawan infeksi dan penyakit. Sampai saat ini belum ada obat yang dapat menyembuhkan HIV, namun ada jenis obat yang dapat memperlambat perkembangan virus. Jenis obat ini adalah antiretroviral (ARV). ARV bekerja dengan menghilangkan unsur yang dibutuhkan virus HIV untuk menggandakan diri dan mencegah virus HIV menghancurkan sel CD4 (jenis sel darah putih atau limfosit). Pengidap HIV bisa menjalani pengobatan ARV untuk memperlambat perkembangan virus ini sehingga bisa menjalani hidup dengan normal. Jika tidak langsung menjalani pengobatan, pengidap HIV bisa masuk pada tahap infeksi akhir, yang disebut AIDS (*Acquired Immune Deficiency Syndrome*). Ketika seseorang sudah mengalami AIDS, maka tubuh tidak lagi memiliki kemampuan untuk melawan infeksi yang ditimbulkan.

HIV berada dalam cairan tubuh manusia. Cairan yang berpotensi mengandung HIV adalah darah, cairan sperma, cairan vagina dan air susu ibu (KPA, 2007). Pria yang sudah disunat memiliki resiko HIV yang lebih kecil dibandingkan dengan pria yang tidak disunat (Widoyono, 2008). Penularan HIV dapat terjadi melalui berbagai cara diantaranya melalui cairan tubuh seperti darah, cairan genitalia, dan ASI. Namun, virus HIV tidak terdapat dalam air mata dan keringat. Virus HIV menular melalui beberapa cara penularan (Nursalam dan Kurniawati, 2007), yaitu:

- a. Berhubungan badan dengan pengidap HIV/AIDS.
- b. Ibu yang mengidap HIV yang dapat ditularkan kepada bayinya.
- c. Darah dan produk darah yang tercemar HIV/AIDS.
- d. Pemakaian alat kesehatan yang tidak steril.
- e. Penggunaan jarum suntik secara bergantian.

2.3 Data Mining

Data mining adalah sekumpulan data dalam jumlah besar yang diolah untuk mendapatkan sebuah informasi, pola data yang menarik dan beberapa pengetahuan. Data dalam jumlah besar tersebut tersimpan dalam media penyimpanan data seperti *database* atau basis data, *data warehouse*, dan media penyimpanan data lainnya. *Data mining* merupakan irisan dari beberapa disiplin ilmu yang lebih dulu ada seperti visualisasi data, pengolahan *database* matematika, statistika, *artificial intelligence*, dan *machine learning* (Han dan Kamber, 2006).

Data mining memiliki peranan penting dalam bidang ilmu dan teknologi, industri, cuaca, serta keuangan. Hal tersebut dikarenakan *data mining* dapat mengolah data dalam skala besar. *Data mining* memiliki beberapa metode yaitu klasifikasi, *clustering*, regresi, seleksi variabel, dan *market basket analysis* (Santoso, 2007).

2.4 Clustering

Clustering merupakan salah satu teknik dalam *data mining* yang digunakan untuk mengolah data yang tidak diketahui label atau kelasnya. *Clustering* juga dikenal dengan istilah *cluster analysis*. Cara kerja *clustering* adalah dengan mengelompokkan sekumpulan data ke dalam beberapa kelas atau *cluster*. Objek-objek yang ada pada kelas tersebut mempunyai kemiripan yang tinggi jika dibandingkan dengan objek lain yang ada di kelas tersebut, tetapi memiliki kemiripan yang rendah jika dibandingkan dengan objek yang ada di kelas/*cluster* lain (Han dan Kamber, 2006).

Dalam sebuah kelompok data, ada beberapa objek data yang ditempatkan ke dalam *cluster* yang sama berdasarkan sifat dan karakteristik data tersebut. Akan tetapi, dalam kasus tertentu akan ada suatu objek data yang memiliki karakteristik dan sifat yang sangat berlainan dengan data lain. Biasanya data tersebut ditempatkan ke

dalam *cluster* tersendiri. Jumlah anggota *cluster* yang menempati *cluster* tersebut biasanya sedikit sekali jika dibandingkan dengan rasio jumlah data yang ada. Data tersebut dikenal dengan istilah data *outlier* atau anomali. Salah satu keunggulan teknik *clustering* adalah mampu mendeteksi data *outlier* (*outlier detection*) karena tidak ada label atau kelas ketika data ini pertama kali diolah (Han dan Kamber, 2006).

2.5 Fuzzy C-Means

Jim Bezdek pada tahun 1981 (Jain, Murthy, & Flynn 1999) pertama kali memperkenalkan Metode *Fuzzy C-Means Clustering* (FCM). *Fuzzy C-Means* adalah salah satu teknik pengelompokan data yang mana keberadaan tiap titik data dalam suatu kelompok (*cluster*) ditentukan oleh derajat keanggotaan.

Fuzzy C-Means ini dikenal juga sebagai *Fuzzy ISODATA* yang merupakan bagian dari *Hard K-Means*. *Fuzzy C-Means* menggunakan model pengelompokan *fuzzy* sehingga data yang didapat menjadi anggota dari semua *cluster* yang terbentuk dengan derajat atau tingkat keanggotaan yang berada antara 0 hingga 1. Tingkat data dalam suatu kelas *cluster* ditentukan oleh derajat keanggotannya.

Menurut Kusumadewi dan Purnomo (2004), langkah-langkah algoritma *Fuzzy C-Means* adalah sebagai berikut:

1. *Input* data yang akan di *cluster* X , berupa matriks berukuran $n \times m$ (n =jumlah sampel data, m = atribut setiap data). X_{ij} = data sampel ke- i ($i=1,2,\dots,n$), atribut ke- j ($j = 1,2,\dots,m$).
2. Menentukan :
 - Jumlah *cluster* (c)
 - Pangkat (w)
 - Maksimum iterasi (MaxIter)
 - Error terkecil yang diharapkan (ξ)
 - Fungsi objektif awal ($P_0 = 0$)
 - Iterasi awal ($t=1$)

3. Bangkitkan bilangan random μ_{ik} , $i = 1, 2, 3, \dots, n$; $k = 1, 2, 3, \dots, n$; sebagai elemen-elemen matriks partisi awal U (derajat keanggotaan dalam *cluster*). μ_{ik} adalah derajat keanggotaan yang merujuk pada seberapa besar kemungkinan sesuatu data bisa menjadi anggota kedalam suatu *cluster*. Posisi dan nilai matriks dibangun secara random. Dimana nilai keanggotaan terletak pada interval 0 sampai dengan 1. Pada posisi awal matriks partisi U masih belum akurat begitu juga pusat *clusternya*. Sehingga kecenderungan data untuk masuk suatu *cluster* juga belum akurat.

Hitung setiap kolom atribut:

$$Q_j = \sum_{k=1}^c \mu_{ik} \quad (1)$$

Q_j adalah jumlah nilai dengan derajat keanggotaan per kolom = 1
Dengan $j = 1, 2, \dots, m$.

4. Menghitung pusat *cluster* ke- k : V_{kj} dengan $k=1, 2, 3, \dots, c$; dan $j=1, 2, 3, \dots, m$

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w \times X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \quad (2)$$

V_{kj} = Pusat cluster

i = iterasi

μ_{ik} = perubahan matrix partisi

X_{ij} = atribut

5. Menghitung fungsi objektif pada iterasi ke- t . Fungsi objektif digunakan sebagai syarat perulangan untuk mendapatkan pusat *cluster* yang tepat. Sehingga diperoleh kecenderungan data untuk masuk ke *cluster* mana pada langkah akhir.

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left(\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right) \quad (3)$$

P_t = Fungsi objektif

$\sum_{i=1}^n$ = jumlah data yang di *cluster*

$\sum_{i=1}^n$ = jumlah perhitungan cluster awal

6. Menghitung perubahan matrix partisi:

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{w-1}}{\sum_{k=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{w-1}} \quad (4)$$

Dengan $i=1,2,\dots,n$; dan $k=1,2,\dots,c$.

7. Memeriksa kondisi berhenti:
- Jika $(|Pt - Pt-1| < \xi)$ atau $(t > \text{MaxIter})$ maka berhenti.
 - Jika tidak : $t = t + 1$, mengulang ke langkah empat.

2.6 Metode Elbow

Menurut Merliana, Ernawati, & Santoso (2015), metode *Elbow* merupakan suatu metode penentuan jumlah *cluster* optimum atau terbaik untuk menghasilkan suatu informasi dengan cara melihat persentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik. Dengan metode ini, ide atau gagasan akan diberikan dengan cara memilih nilai *cluster*. Selanjutnya nilai *cluster* tersebut akan ditambah sebagai model data untuk penentuan *cluster* terbaik. Selain itu, hasil persentase perhitungan akan menjadi pembanding antara jumlah *cluster* yang ditambah. Grafik merupakan sebagai sumber informasi dari hasil persentase yang berbeda dari setiap nilai *cluster*. Jika nilai *cluster* pertama dengan nilai *cluster* kedua memberikan sudut dalam grafik atau nilainya mengalami penurunan paling signifikan atau paling besar maka nilai *cluster* tersebut adalah yang terbaik. Secara tabel, jarak antara 2 titik *cluster* dapat dihitung dengan cara mengurangi nilai SSE (*Sum of Squares Error*) antara 2 titik *cluster*. Berikut ini merupakan rumus SSE:

$$SSE = \sum_{k=1}^k \sum_{x_i \in S_k} ||x_i - C_k||^2 \quad (5)$$

X_i = nilai atribut dari data ke- i

C_k = nilai atribut titik pusat *Cluster* ke- i

2.7 RStudio

RStudio adalah aplikasi yang didirikan oleh JJ Allaire yang digunakan untuk mempermudah pengolahan bahasa R. Bahasa R yaitu bahasa pemrograman untuk komputasi statistika dan grafik. RStudio memiliki dua versi, yaitu *open source* (gratis) dan *commercial edition* (berbayar). RStudio juga tidak hanya terbatas dalam bentuk aplikasi desktop (*Windows, MacOS, dan Linux*), melainkan terdapat versi RStudio *Server*, yaitu RStudio yang dapat diakses melalui *browser* yang terhubung dengan suatu jaringan komputer.

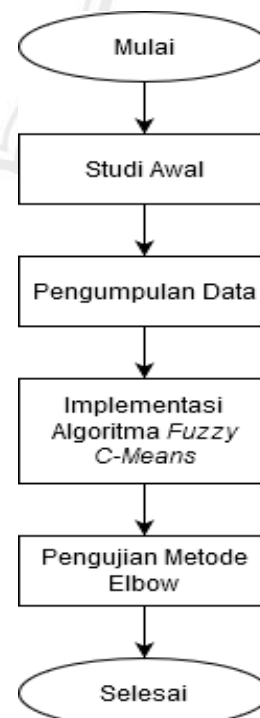
3. METODOLOGI PENELITIAN

3.1 Metode Penelitian

Penelitian ini menggunakan metode eksperimental. Metode eksperimental merupakan suatu metode yang bertujuan untuk menyelidiki adanya kemungkinan hubungan sebab akibat dengan cara melakukan pengujian pada suatu variabel.

3.2 Diagram Alur Penelitian

Pada penelitian ini dilakukan beberapa tahapan, antara lain:



Gambar 1. Tahapan Penelitian

3.3 Studi Awal

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 literatur yang digunakan, kemudian literatur tersebut diseleksi untuk ditentukan sebagai literatur yang akan digunakan dalam penelitian.

3.4 Pengumpulan Data

Jenis data yang digunakan dalam penelitian ini adalah data sekunder. Data sekunder adalah data yang diperoleh dari sumber yang sudah ada yaitu dari Sistem Informasi HIV AIDS & IMS (SIHA) Kementerian Kesehatan Republik Indonesia. Pengumpulan data diambil dari Laporan Perkembangan HIV AIDS & Infeksi Menular Seksual (IMS) Triwulan IV Tahun 2018 yang diakses pada link https://siha.depkes.go.id/portal/files_upload/Laporan_Triwulan_IV_2018.pdf

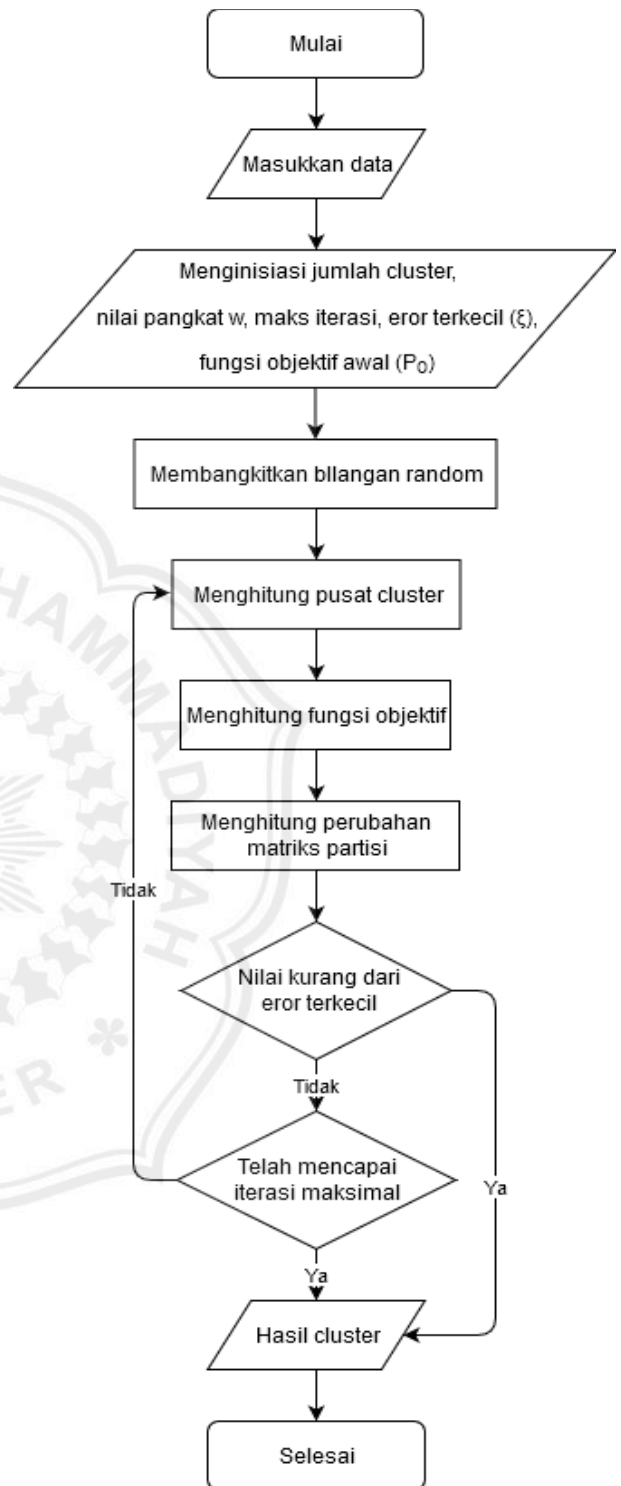
3.5 Dataset

Dataset untuk penelitian ini adalah jumlah kasus terinfeksi HIV mulai tahun 2015 sampai dengan tahun 2018. Terdiri dari 136 data dari 34 Provinsi di Indonesia. Dimana data tersebut akan di *cluster* menggunakan algoritma *Fuzzy C-Means* dan menggunakan metode *Elbow* sebagai metode optimasi *cluster* guna mengetahui jumlah *cluster* optimum yang akan digunakan. Pada bab ini, data yang digunakan hanya 15 data sampel untuk mewakili data yang dihitung.

3.6 Implementasi Algoritma *Fuzzy C-Means*

Diagram alur dari algoritma *Fuzzy C-Means* yang digunakan berdasarkan jumlah kasus terinfeksi HIV di 34 provinsi di

Indonesia, pada umumnya kinerja *Fuzzy C-Means* adalah sebagai berikut:



Gambar 2. Flowchart *Fuzzy C-Means*

3.7 Metode *Elbow*

Metode *Elbow* untuk menentukan jumlah *cluster* yang paling optimum atau yang terbaik. Langkah-langkah metode *Elbow*:

1. Inisialisasi awal nilai *cluster* (2 *cluster*).
2. Menaikkan nilai *cluster* sampai jumlah *cluster* yang ditentukan (10 *cluster*).
3. Menghitung nilai SSE (*Sum of Squares Error*) dari setiap *cluster*.

$$SSE = \sum_{k=1}^k \sum_{xi \in S_k} \|X - C_k\|_2^2 \quad (5)$$

X_i = nilai atribut dari data ke- i

C_k = nilai atribut titik pusat *Cluster* ke- i

4. Melakukan perhitungan SSE (*Sum of Squares Error*) sampai *cluster* yang ditentukan.
5. Melihat hasil SSE (*Sum of Squares Error*) dari nilai *cluster* yang turun secara drastis atau nilai yang berubah signifikan.
6. Menetapkan nilai *cluster* yang berbentuk siku.

4. HASIL DAN PEMBAHASAN

4.1 Data Pengujian

Pada bab ini membahas mengenai hasil yang diperoleh dari perhitungan yang telah dilakukan. Data yang telah didapat akan di *cluster* menggunakan algoritma *Fuzzy C-Means* kemudian diolah untuk mendapatkan hasil *cluster* optimum atau *cluster* terbaik dengan menggunakan metode *Elbow*. Data yang digunakan adalah data jumlah kasus terinfeksi HIV di 34 Provinsi di Indonesia periode 2015 sampai dengan 2018.

4.2 *Fuzzy C-Means* pada RStudio

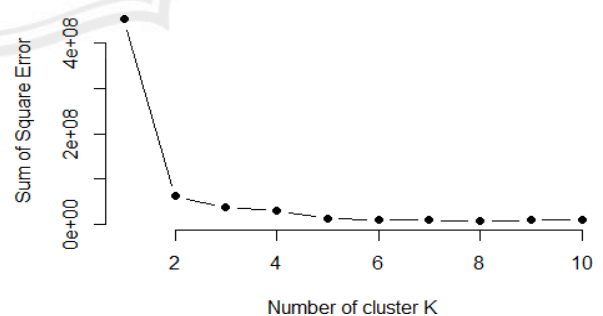
Data diolah menggunakan RStudio yang di *cluster* menggunakan algoritma *Fuzzy C-Means* dari 2 *cluster* sampai dengan 10 *cluster*. *Output* yang dapat dihasilkan dari eksekusi perintah pada RStudio adalah jumlah iterasi, pusat *cluster*, fungsi objektif, dan derajat keanggotaan setiap objek terhadap tiap *cluster*. *Cluster* yang dihasilkan dari perintah *Fuzzy C-Means* pada RStudio ditampilkan ke dalam *plot* dari setiap *cluster* yang terbentuk.

4.3 Penentuan Jumlah *Cluster* Optimum

Setelah melakukan proses *cluster* menggunakan *Fuzzy C-Means*, kemudian dilakukan proses metode *Elbow* untuk menentukan *cluster* optimum dalam penentuan *cluster* terbaik. Hasil perhitungan metode *Elbow* pada RStudio ditampilkan dalam nilai SSE (*Sum of Squares Error*) dan grafik yang terdiri dari sumbu x dan sumbu y. Nilai pada sumbu x adalah jumlah *cluster* yaitu 2 *cluster* hingga 10 *cluster*, sedangkan sumbu y adalah nilai SSE (*Sum of Squares Error*) yang dihasilkan dari pusat setiap *cluster*. Berikut ini adalah hasil metode *Elbow*:

Tabel 1. Hasil Nilai *Elbow*

C	SSE	Jarak	Keterangan
1	453117693	-	-
2	62152429	390965264	Jarak C1 ke C2
3	38019234	24133195	Jarak C2 ke C3
4	30379069	7640165	Jarak C3 ke C4
5	12764132	17614937	Jarak C4 ke C5
6	9738375	3025757	Jarak C5 ke C6
7	9382624	355751	Jarak C6 ke C7
8	6433009	2949615	Jarak C7 ke C8
9	9564712	-3131703	Jarak C8 ke C9
10	9029334	535378	Jarak C9 ke C10



Gambar 3. Hasil Nilai *Elbow*

Pada metode *Elbow*, nilai *cluster* yang diambil sebagai *cluster* optimum atau terbaik adalah titik yang membentuk siku. Penjelasan pada titik yang membentuk siku

adalah pada titik yang terjadi penurunan yang signifikan antara 2 titik *cluster* dan kemudian diikuti oleh nilai yang relatif konstan. Pada tabel 1 menunjukkan nilai SSE (*Sum of Squares Error*) dengan jarak paling signifikan atau paling besar terdapat pada 2 *cluster* dengan jarak 1 *cluster* ke 2 *cluster* yaitu 390965264. Nilai jarak 1 *cluster* ke 2 *cluster* tersebut merupakan nilai jarak yang mengalami penurunan paling signifikan atau paling besar dan kemudian diikuti oleh nilai jarak yang relatif konstan, sehingga 2 *cluster* merupakan *cluster* optimum atau terbaik. Dapat dilihat juga pada gambar 3, bahwa titik yang membentuk siku terdapat pada titik 2 *cluster*, dimana dari 1 *cluster* ke 2 *cluster* mengalami penurunan yang signifikan dibandingkan dengan yang lainnya. Kemudian dari titik 2 *cluster* ke titik selanjutnya diikuti nilai yang relatif konstan. Jadi, *cluster* optimum terdapat pada 2 *cluster*.

4.4 Profiling Cluster Optimum

Dari proses *clustering* menggunakan algoritma *Fuzzy C-Means* dan mencari *cluster* optimum menggunakan metode *Elbow*, diketahui *cluster* optimum berada di 2 *cluster*. Untuk mengetahui karakteristik dari masing-masing *cluster*, dilakukan *profiling cluster* data jumlah kasus terinfeksi HIV setiap provinsi yang masuk ke dalam kelompok *cluster* 1 dan *cluster* 2 berdasarkan perhitungan yang telah dilakukan sebelumnya.

Pada *cluster* 1 terdiri dari 29 anggota provinsi dengan rentang jumlah kasus HIV tahun 2015 terdapat 13 s/d 2028 kasus, tahun 2016 terdapat 7 s/d 2367 kasus, tahun 2017 terdapat 37 s/d 2441 kasus, dan tahun 2018 terdapat 26 s/d 2211 kasus. Sedangkan pada *cluster* 2 terdiri dari 5 anggota provinsi dengan rentang jumlah kasus HIV pada tahun 2015 terdapat 3005 s/d 4695 kasus, tahun 2016 terdapat 3866 s/d 6513 kasus, tahun 2017 terdapat 4358 s/d 8204 kasus, dan tahun

2018 terdapat 3546 s/d 8608 kasus. Dari hasil karakteristik data tahun 2015 sampai dengan tahun 2018, *cluster* 1 memiliki jumlah kasus HIV lebih rendah dibandingkan dengan *cluster* 2.

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan uraian permasalahan dan pembahasan pada bab sebelumnya, maka diperoleh kesimpulan sebagai berikut:

1. Penerapan algoritma *Fuzzy C-Means* dari jumlah kasus terinfeksi HIV untuk mengelompokkan provinsi di Indonesia menghasilkan 2 *cluster* optimum dengan nilai jarak SSE (*Sum of Squares Error*) 1 *cluster* ke 2 *cluster* pada *Elbow* yaitu 390965264. Nilai jarak 1 *cluster* ke 2 *cluster* tersebut merupakan nilai jarak yang mengalami penurunan paling signifikan atau paling besar dan kemudian diikuti oleh nilai jarak yang relatif konstan, sehingga 2 *cluster* merupakan *cluster* optimum atau terbaik.
2. Hasil pengelompokkan pada *cluster* 1 terdapat 29 provinsi yaitu NAD, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bangka Belitung, Kepulauan Riau, DI Yogyakarta, Banten, Bali, 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. Pada *cluster* 2 terdapat 5 provinsi yaitu DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, dan Papua. Berdasarkan hasil karakteristik data tahun 2015 sampai dengan tahun 2018, *cluster* 1 memiliki jumlah kasus HIV lebih rendah dibandingkan dengan *cluster* 2.

5.2 Saran

Saran dari penelitian ini adalah:

1. Validitas *cluster* untuk mencari *cluster* terbaik, bisa menggunakan alternatif lain selain metode *Elbow*, seperti metode *Silhouette*, *Gap Statistic*, *Davies Bouldin Index*, dll.
2. Perhitungan manual *cluster Fuzzy C-Means* di *Microsoft Excel* digunakan sebagai pembandingan hasil *cluster* pada RStudio. Perhitungan manual di *Microsoft Excel* dapat menggunakan bilangan acak/random untuk menentukan *cluster* awal yang diambil dari RStudio agar menghasilkan pengelompokan yang serupa.

DAFTAR PUSTAKA

- Anonim. Badan Pusat Statistika. 2017. *Luas Daerah dan Jumlah Pulau Menurut Provinsi*. Jakarta: Badan Pusat Statistika.
- Anonim. Departemen Kesehatan. 2019. *Laporan Perkembangan HIV AIDS & Infeksi Menular Seksual (IMS) Triwulan IV Tahun 2018*. Jakarta: Kementerian Kesehatan Republik Indonesia.
- Anonim, "Hari AIDS Sedunia, Momen STOP Penularan HIV: Saya Berani, Saya Sehat!", 01 Desember 2018, <<https://www.depkes.go.id/article/view/18120300001/hari-aids-sedunia-momen-stop-penularan-hiv-saya-berani-saya-sehat-.html>>. Diakses pada 22 Januari 2020.
- Anonim, "HIV dan AIDS", 12 November 2019, <<https://www.halodoc.com/kesehatan/hiv-dan-aids>>. Diakses pada 22 Januari 2020.
- Anonim, "Permendagri 137 Tahun 2017, Kode Dan Data Wilayah Administrasi Pemerintahan.", 20 Juli 2019, <<https://www.jogloabang.com/pustaka/permendagri-137-2017-kode-data-wilayah-administrasi-pemerintahan>>. Diakses pada 22 Januari 2020.
- Han, J., & Kamber, M. 2006. *Data Mining: Concept and Techniques, Second Edition*. Waltham: Morgan Kaufmann Publishers.
- Jain, A. K., Murty, M. N., & Flynn, P. J. 1999. *Data Clustering: A Review*. Columbus: ACM Computing Surveys, Vol. 31, No. 3.
- KPA. 2007. *Strategi Nasional Penanggulangan HIV dan AIDS 2007-2010. Draft Final 040107*. Jakarta: Kemenkes.
- Kusumadewi, S., & Purnomo, H. 2004. *Aplikasi Logika Fuzzy untuk Pendukung Keputusan*. Yogyakarta: Graha Ilmu.
- Larose, Daniel, T. 2005. *Discovering Knowledge In Data: An Introduce to Data Mining*. Canada: John Willey and Sons, Inc.
- Merliana, N. P. E., Ernawati, & Santoso, A. J. 2015. *Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means Clustering*. Yogyakarta: Prosiding Seminar Nasional Multi Disiplin Ilmu & Call for Papers Unisbank (Sendi_U), 978-979.
- Nursalam & Kurniawati, N.D. 2007. *Asuhan Keperawatan Pasien Terinfeksi HIV/AIDS*. Jakarta: Salemba Medika.
- Riveranda, O., & Nengsih, W. 2016. *K-Means Analysis Klasterisasi Kasus HIV/AIDS di Indonesia*. Riau: ResearchGate.
- Santoso, B. 2007. *Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Saputra, D. B., & Riksakomara, E. 2018. *Implementasi Fuzzy C-Means dan Model RFM untuk Segmentasi Pelanggan (Studi Kasus : PT. XYZ)*. Surabaya: Jurnal Teknik ITS Vol. 7, No. 1 (2018) 2337-3520 (2301-928X Print).
- Widiyono. 2008. *Penyakit Tropis: Epidemiologi, Penularan, Pencegahan dan*

Pemberantasannya.
Erlangga.

Jakarta:

