

Implementasi Algoritma K-means untuk Clustering penyakit HIV/AIDS di Kabupaten Semarang

Ardila Rifani, Desi Purwanti K., M.Kom
Jurusan Teknik Informatika, Fakultas Ilmu Komputer,
Universitas Dian Nuswantoro Semarang,
Jln. Nakula 1 No 5-11 Semarang 50131 INDONESIA
111201206706@mhs.dinus.ac.id

Abstract- The increase of HIV / AIDS cases in Indonesia become the case that never escape from the government's attention. With the death rate continues to rise makes people worry about the spread of HIV / AIDS. HIV / AIDS can only be transmitted through sexual intercourse from an infected person, transfusions of blood contaminated with HIV / AIDS, and from an HIV-infected woman to her child during pregnancy, childbirth, and breastfeeding. In Semarang regency, the number of HIV / AIDS cases make the relevant agencies should act quickly and accurately. It required a medium that can help to conduct a HIV/AIDS counseling and testing accurately in various sub-districts in Semarang based on sub-districts with the highest patient. Furthermore, this web-based application will be integrated by clustering the number of people with HIV / AIDS using K-Means algorithm that will help related parties to take action against HIV / AIDS in Semarang Regency.

Keywords: *HIV / AIDS, clustering, K-Means algorithm, Semarang Regency, web based.*

I. PENDAHULUAN

Menurut WHO (World Health Organization), penderita HIV di seluruh dunia sampai dengan tahun 2013 tercatat mencapai 35.000.000 jiwa, 1.500.000 diantaranya meninggal karena AIDS yang terdiri dari anak-anak dan orang dewasa. Di Indonesia sendiri penderita HIV sampai dengan tahun 2014 tercatat 150.296 jiwa, 55.799 diantaranya menderita AIDS dan angka kematian yang disebabkan virus ini mencapai 9.796 jiwa [2].

Hal ini disebabkan oleh kurangnya akses terhadap pelayanan pengobatan dan pencegahan virus HIV/AIDS. Virus HIV/AIDS paling banyak ditemukan di dalam cairan sperma, vagina, darah, serta asi. Penyebarannya sendiri dapat terjadi akibat seks bebas, jarum suntik yang tercemar virus ini, transfusi darah, serta antara ibu yang terkena virus HIV/AIDS dan bayinya saat mengandung, melahirkan, dan menyusui.

Di provinsi Jawa Tengah sampai dengan 30 September 2014 tercatat penderita HIV mencapai 9.032 jiwa sedangkan penderita AIDS 3.767 jiwa [2]. Dengan penderita yang tercatat demikian banyaknya, salah satu kabupaten di provinsi Jawa Tengah yang menyumbang cukup banyak penderita HIV/AIDS yaitu Kabupaten Semarang. Berdasarkan banyaknya, penderita HIV/AIDS dikelompokkan berdasarkan tingkat epidemic yaitu epidemic meluas (epidemic tinggi) yaitu HIV sudah menyebar ke masyarakat umum dan prevalensi ibu hamil mencapai 1%, epidemic terkonsentrasi (epidemic sedang) yaitu apabila HIV sudah menyebar ke sub populasi tertentu dan mencapai 5% pada sub populasi tersebut, epidemic

rendah yaitu apabila HIV telah ada dan belum menyebar pada sub populasi tertentu. Dengan adanya kelompok epidemic tersebut maka dilakukan intervensi suatu program yang cepat tanggap dan tepat sasaran. Program tersebut adalah menempatkan pencegahan pada prioritas utama, memilih intervensi sesuai pola penularan, memfokuskan pada wilayah geografis dengan kasus yang tinggi, memilih pendekatan konseling dan tes HIV yang tepat, merencanakan penyediaan perawatan dan pengobatan, dan menajmin pelayanan konseling dan tes HIV [21].

Dengan demikian diperlukan sebuah media yang berbentuk aplikasi untuk membantu dinas-dinas tertentu dalam melakukan intervensi program dalam melakukan konseling dan tes HIV di Kabupaten Semarang. Penderita HIV/AIDS biasanya dibagi berdasarkan pola penyebaran seperti jumlah penderita HIV/AIDS, rata-rata umur, jumlah penderita hidup, dan jumlah penderita mati. Nantinya aplikasi ini akan diimplementasikan dengan metode clustering sehingga mampu mengelompokkan penderita yang terkena HIV dan AIDS serta pola penyebarannya.

II. STUDI PUSTAKA

A. Literature Review

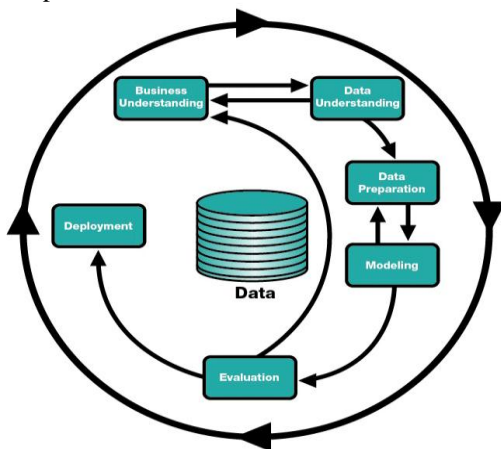
Ada beberapa referensi yang diambil penulis sebagai bahan pertimbangan untuk penelitian yang dilakukan, referensi tersebut diambil dari beberapa penulisan yang dilakukan sebelumnya yang membahas permasalahan yang hampir sama, antara lain :

1. Analisis Penerapan Metode Single Linkage dan K-Means untuk Pewilayahan Pola Hujan Provinsi Sumatera Selatan Berbasis Arc View.
2. Sistem Informasi Geografis Visualisasi Clustering Penyakit ISPA di Kecamatan Kaliwungu.

3. Sistem Informasi Geografis Berbasis Web untuk Produksi Kedelai di Indonesia menggunakan K-Means Clustering.
4. Pemetaan Daerah Rawan Kriminalitas di Wilayah Hukum Poltabes Semarang Tahun 2013 dengan menggunakan Metode Clustering.
5. Penggunaan metode Statistik K-means Clustering pada Analisis Peruntukan Lahan Usaha Tambang Berbasis Sistem Informasi Geografis.

B. CRISP-DM

CRISP-DM menyediakan standart proses data mining sebagai pemecahan masalah yang umum untuk bisnis atau penelitian. Dalam penelitian ini, tahapan penelitian menggunakan CRISP-DM sebagai langkah-langkah dalam melakukan penelitian.



Gambar 2.1 CRISP-DM

Berikut ini merupakan enam fase CRISP-DM:

- a. Business Understanding atau Pemahaman Bisnis
Pada fase ini dibutuhkan pemahaman tentang penentuan tujuan proyek dan kebutuhan yang detail.
- b. Data Understanding atau Pemahaman Data
Pada fase ini dilakukan pengumpulan data menggunakan analisis penyelidikan data untuk mengenali data.
- c. Data Preparation atau Persiapan Data
Fase ini merupakan fase persiapan data yang dilakukan dengan memilih tabel atau *field* yang akan ditransformasikan menjadi database baru.
- d. Modeling Phase atau Fase Pemodelan
Fase ini merupakan fase untuk memilih dan mengaplikasikan teknik pemodelan yang sesuai dengan proyek.
- e. Evaluation Phase atau Fase Evaluasi
Fase ini merupakan fase untuk mengevaluasi satu atau lebih model yang digunakan dalam fase pemodelan untuk mendapatkan kualitas dan efektivitas sebelum digunakan.
- f. Deployment Phase atau Fase Penyebaran
Fase ini merupakan fase penyusunan laporan atau presentasi pengetahuan yang didapat dari evaluasi proses data mining.

C. Algoritma K-means

Menurut [4], pencarian pusat cluster secara iterative merupakan konsep dasar dari algoritma k-means. Pusat cluster ditetapkan berdasarkan jarak setiap data dengan pusat cluster dimulai dengan mengidentifikasi data yang akan di cluster. Pusat awal cluster ditetapkan secara bebas pada tahap iterasi. Kemudian jarak setiap data dengan setiap pusat cluster akan dihitung. Perhitungan jarak data ke-*i* (*x_i*) pada pusat cluster ke-*k* (*c_k*) diberi nama (*d_{ik}*), dapat menggunakan rumus Euclidean, yaitu :

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2}$$

Keterangan :

d_{ik} : jarak data ke-*i*

m : jumlah variable

x_{ij} : data yang akan di cluster

c_{kj} : pusat dari cluster

Langkah – Langkah :

- a. Menentukan jumlah kelompok yang akan dibentuk
- b. Menentukan jumlah awal cluster secara acak
- c. Menghitung jumlah jarak setiap data ke masing-masing centroid/ pusat cluster
- d. Setiap data akan memilih berdasarkan centroid yang paling dekat
- e. Menentukan posisi centroid baru dengan cara menghitung nilai rata-rata dari beberapa data yang memilih centroid yang sama.
- f. Melakukan iterasi atau kembali ke langkah c jika posisi centroid baru dengan centroid lama belum sama.

III. METODE PENELITIAN

A. Objek Penelitian

Dalam tugas akhir ini penulis melakukan penelitian yang berkaitan dengan penyakit HIV/AIDS di Dinas Kesehatan Kabupaten Semarang yang beralamat di jalan MT Haryono No. 29, Ungaran Timur. Penelitian yang dilakukan adalah dengan mengumpulkan data-data dan informasi yang berkaitan dengan penyakit HIV/AIDS di Kabupaten Semarang.

B. Sumber Data

a. Data Primer

Data primer adalah data yang diperoleh langsung dari sumber data yang berkaitan dengan penelitian yang dilakukan, yaitu penderita penyakit HIV/AIDS di Kabupaten Semarang yang didapatkan dari pengumpulan data pada Dinas Kesehatan Kabupaten Semarang.

b. Data Sekunder

Data sekunder adalah data dan informasi yang didapatkan dari berbagai buku, publikasi, jurnal ilmiah, laporan penelitian yang bersumber dari instansi ataupun internet sebagai penunjang landasan teori yang berkaitan dengan *Clustering*, algoritma K-Means, dan data-data HIV/AIDS.

C. Teknik Pengumpulan Data

Teknik pengumpulan data yang dilakukan dalam penelitian ini adalah mengumpulkan data yang berkaitan dengan penderita HIV/AIDS di Kabupaten Semarang dengan mengumpulkan data secara langsung dari Dinas Kesehatan Kabupaten Semarang.

D. Metode Penelitian

Metode penelitian yang digunakan adalah CRISP-DM. langkah-langkahnya sebagai berikut :

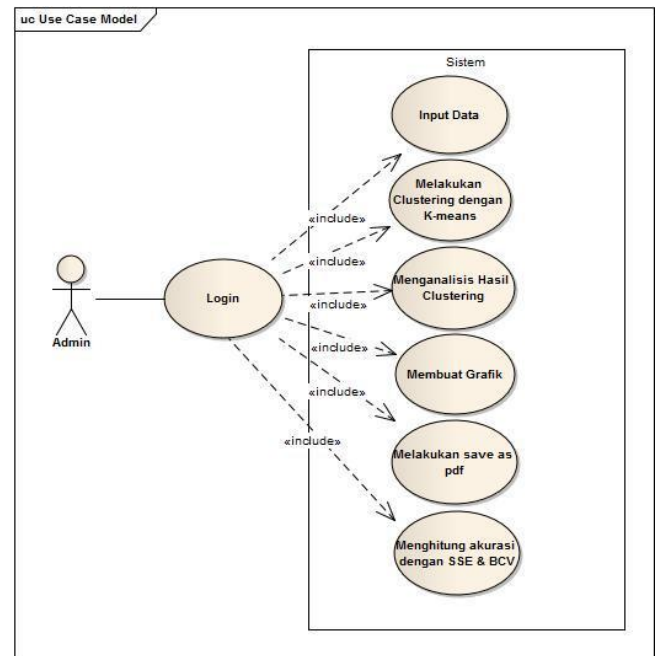
- a. Pemahaman Bisnis
Pada fase ini akan dilakukan pemahaman apa itu HIV/AIDS cara penularannya, serta pencegahannya.
- b. Pemahaman Data
Pada fase ini akan dilakukan pengumpulan data-data penderita HIV/AIDS di Kabupaten Semarang.
- c. Persiapan Data
Pada fase ini data-data yang telah dikumpulkan akan disiapkan untuk diolah menggunakan algoritma k-means.
- d. Fase Pemodelan
Tahap ini merupakan tahap implementasi dari data-data yang telah dikumpulkan yang selanjutnya akan dilakukan clustering dengan menggunakan algoritma k-means.
- e. Fase Evaluasi
Tahap ini merupakan tahap evaluasi dari hasil perhitungan algoritma k-means. Tahap evaluasi dilakukan dengan memeriksa kualitas suatu cluster dengan menggunakan SSE dan BCV.
- f. Tahap Penyebaran
Tahap ini merupakan implementasi algoritma k-means dalam bentuk web.

IV. HASIL & PEMBAHASAN

A. Usecase Diagram

Dalam membuat use case diagram diperlukan langkah-langkah sebagai berikut:

- a. Identifikasi Aktor
- b. Diagram model usecase

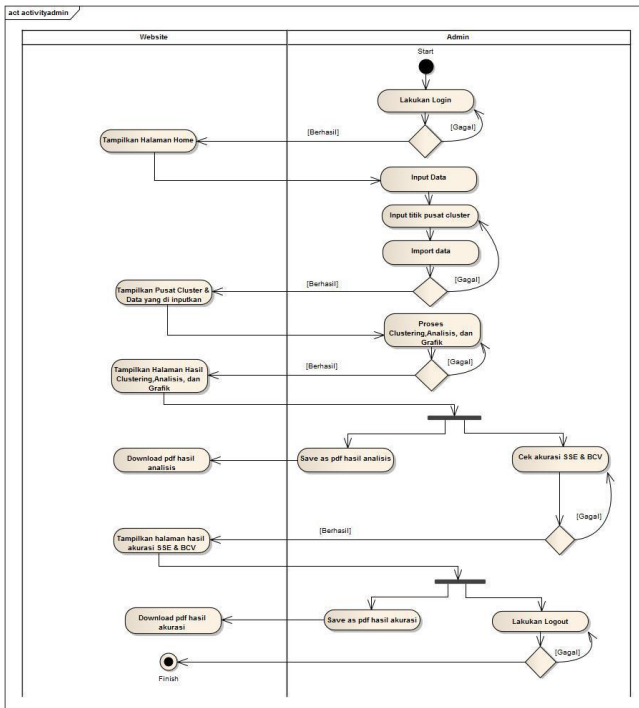


Gambar 4.1 Use Case Diagram

Skenario yang didapat berdasarkan use case diatas adalah sebagai berikut:

- a. Input Data
Admin akan menginputkan data ke dalam system. Data yang diinputkan merupakan file excel.
- b. Melakukan *Clustering* dengan K-Means
Data yang sudah diinputkan selanjutnya akan di *clustering* dengan menggunakan algoritma k-means.
- c. Menganalisis Hasil *Clustering*
Dari hasil *clustering* yang didapat, maka dapat dianalisa penderita penyakit HIV/AIDS di Kabupaten Semarang berdasarkan wilayah dengan penderita tertinggi, sedang, dan rendah.
- d. Membuat Grafik
Dari hasil analisis yang didapat, maka hasilnya dapat diubah menjadi bentuk grafik yang lebih memudahkan dalam pembacaan tiap cluster.
- e. Melakukan Save as Pdf
Dari hasil analisis dan hasil akurasi maka hasilnya dapat diubah ke dalam bentuk pdf dengan melakukan *download* maka sistem akan mengubah hasil analisis dan hasil akurasi dalam bentuk pdf.

B. Activity Diagram



Gambar 4.2 Activity Diagram

Gambar 4.2 adalah aktifitas yang dilakukan admin dalam mengelola data penyakit HIV/AIDS untuk mendapatkan hasil *clustering* dengan menggunakan algoritma k-means. Aktifitas yang dilakukan admin diantaranya adalah melakukan input data dan input pusat *cluster*, selanjutnya system akan melakukan *clustering* dan akan ditampilkan hasil *clustering* dan hasil analisis. Selanjutnya admin akan memeriksa akurasi *cluster* dengan menggunakan perbandingan BCV dan SSE. Lalu hasil analisis dan hasil akurasi tersebut bisa diubah dan disimpan menjadi bentuk pdf. Namun, semua aktifitas tersebut bisa dilakukan dengan melakukan login terlebih dahulu yang dilakukan oleh admin.

C. Hasil dan Pembahasan

Dalam melakukan input data dalam system, terdapat data yang akan diinputkan yang terdiri dari jumlah penderita HIV/AIDS, rata-rata umur penderita HIV/AIDS, jumlah penderita hidup, dan juga jumlah penderita mati yang meliputi semua kecamatan di Kabupaten Semarang.

Table 4.1 Data Penderita HIV/AIDS

Kecamatan	Jumlah Penderita HIV/AIDS	Rata-Rata Umur	Jumlah Penderita Hidup	Jumlah Penderita Mati
GETASAN	2	31	2	0
TENGARAN	8	28	8	0
SUSUKAN	3	52	0	3
KALIWUNGU	0	0	0	0
SURUH	6	32	5	1
PABELAN	0	0	0	0
TUNTANG	6	31	6	0
BANYUBIRU	2	27	2	0
JAMBU	15	32	10	5
SUMOWONO	5	32	4	1
AMBARAWA	27	32	23	4
BANDUNGAN	27	29	20	7
BAWEN	18	31	15	3
BRINGIN	10	33	8	2
BANCAK	4	27	4	0
PRINGAPUS	9	41	8	1
BERGAS	24	43	23	1
UNGARAN BARAT	23	34	23	0
UNGARAN TIMUR	3	34	3	0

Data diatas akan dinormalisasi dengan menggunakan perhitungan berikut ini :

$$\text{Nilai Baru} = \frac{\text{Nilai asal} - \text{Nilai min}}{\text{Nilai maks} - \text{Nilai min}}$$

Dalam melakukan percobaan, terdapat 3 kecamatan yang dijadikan sebagai titik pusat cluster. Kecamatan tersebut merupakan kecamatan Getasan, kecamatan Jambu, dan kecamatan Bergas.

Clustering dilakukan dengan menggunakan data yang telah diinputkan dan juga dengan titik pusat cluster yang telah ditentukan. Total iterasi yang terjadi sebanyak 3 iterasi yang ditunjukkan seperti gambar dibawah ini. Dan rumus Euclidean Distance yang digunakan dalam menghitung clustering adalah dengan menggunakan algoritma k-means.

ITERASI 1							
Data 1	Data 2	Data 3	Objek	kecamatan	Cluster 1	Cluster 2	Cluster 3
3.1178283476675E-10	0.92918314024292	1.2534888021155	Objek 1	GETASAN	X		
0.34751122046924	0.76869892766716	0.93814445385763	Objek 2	TENGARAN	X		
0.5964045307714	0.78493848737216	1.310162793714	Objek 3	SUSUKAN	X		
0.60699902245217	1.1775314002159	1.5793458758514	Objek 4	KALIWUNGU	X		
0.2444158178683	0.69634811887744	1.0496043711064	Objek 5	SURUH	X		
0.60699902245217	1.1775314002159	1.5793458758514	Objek 6	PABELAN	X		
0.22845923142392	0.80742229465134	1.0317077325609	Objek 7	TUNTANG	X		
0.076923076769231	0.93394702617306	1.2699032622149	Objek 8	BANYUBIRU	X		
0.9291831400702	7.2091617137227E-10	0.89546688573186	Objek 9	JAMBU		X	
0.20170548186114	0.72921721984013	1.1056071107046	Objek 10	SUMOWONO	X		
1.4205237231012	0.73308234603964	0.49068082191539	Objek 11	AMBARAWA			X
1.5720351397059	0.68667884480956	0.91462451090596	Objek 12	BANDUNGAN		X	
0.9242889669376	0.37630723266458	0.55249687216415	Objek 13	BAWEN		X	
0.48883163581881	0.47528757752064	0.8669386825561	Objek 14	BRINGIN		X	
0.13771552137704	0.86803419832906	1.1602582576563	Objek 15	BANCAK	X		
0.43892904360029	0.64298552575162	0.85758502682433	Objek 16	PRINGAPUS	X		
1.2534888018993	0.89546688531448	1.96649467136E-10	Objek 17	BERGAS			X
1.2007976793878	0.95861648495072	0.22745444984708	Objek 18	UNGARAN BARAT			X
0.081181917002143	0.89545621624768	1.1880428122375	Objek 19	UNGARAN TIMUR	X		

Gambar 4.3 Iterasi Pertama

ITERASI 2							
Data 1	Data 2	Data 3	Objek	kecamatan	Cluster 1	Cluster 2	Cluster 3
0.13549243231994	0.9682269527188	1.2671336728888	Objek 1	GETASAN	X		
0.25560175270263	0.74056157691703	0.94273100509956	Objek 2	TENGARAN	X		
0.6054579027178	0.90077988960291	1.3307925335974	Objek 3	SUSUKAN	X		
0.58174253134594	1.2172713021804	1.5425709009034	Objek 4	KALIWUNGU	X		
0.14494341548787	0.72515151168708	1.0518873647138	Objek 5	SURUH	X		
0.58174253134594	1.2172713021804	1.5425709009034	Objek 6	PABELAN	X		
0.16099576357451	0.80586627987619	1.0447480344396	Objek 7	TUNTANG	X		
0.12309781074757	0.97165425557308	1.275660995081	Objek 8	BANYUBIRU	X		
0.81564701488425	0.2005690035209	0.82544181691616	Objek 9	JAMBU		X	
0.11451047558786	0.76931645702264	1.1085999120128	Objek 10	SUMOWONO	X		
1.3040808746109	0.55225494966962	0.35429356369976	Objek 11	AMBARAWA			X
1.450165779073	0.60509352688642	0.79048637172158	Objek 12	BANDUNGAN		X	
0.80618134205399	0.19504622397839	0.47827803331424	Objek 13	BAWEN		X	
0.37833473578227	0.48343813078137	0.8525165696761	Objek 14	BRINGIN	X		
0.076716257888593	0.8871562567842	1.1649939688941	Objek 15	BANCAN	X		
0.37547199236199	0.63397860113047	0.88268992305254	Objek 16	PRINGAPUS	X		
1.1647940346387	0.71012549175276	0.1616060421017	Objek 17	BERGAS			X
1.1103134770465	0.76981601623867	0.25002642887411	Objek 18	UNGERAN BARAT			X
0.14363973296145	0.9265170644877	1.2078092571226	Objek 19	UNGERAN TIMUR	X		

Gambar 4.4 Iterasi Kedua

ITERASI 3							
Data 1	Data 2	Data 3	Objek	kecamatan	Cluster 1	Cluster 2	Cluster 3
0.15863071248494	1.1287870998985	1.2671336728888	Objek 1	GETASAN	X		
0.23990793664896	0.89609842866603	0.94273100509956	Objek 2	TENGARAN	X		
0.59542725924533	1.0352329489442	1.3307925335974	Objek 3	SUSUKAN	X		
0.5945439232463	1.3535247471413	1.5425709009034	Objek 4	KALIWUNGU	X		
0.11757668391303	0.88605047560038	1.0518873647138	Objek 5	SURUH	X		
0.5945439232463	1.3535247471413	1.5425709009034	Objek 6	PABELAN	X		
0.14975114319788	0.96551836838775	1.0447480344396	Objek 7	TUNTANG	X		
0.1520376773471	1.1309691787872	1.275660995081	Objek 8	BANYUBIRU	X		
0.7882719074951	0.28672285261912	0.82544181691616	Objek 9	JAMBU		X	
0.092213153117633	0.92981907331843	1.1085999120128	Objek 10	SUMOWONO	X		
1.2763880308431	0.45745379629027	0.35429356369976	Objek 11	AMBARAWA			X
1.4220356279502	0.44399806357693	0.79048637172158	Objek 12	BANDUNGAN		X	
0.7780310554597	0.29522993225349	0.47827803331424	Objek 13	BAWEN		X	
0.34923206379902	0.64458417437517	0.8525165696761	Objek 14	BRINGIN	X		
0.093262539012182	1.0464585895496	1.1649939688941	Objek 15	BANCAN	X		
0.35137888037168	0.79033399916861	0.88268992305254	Objek 16	PRINGAPUS	X		
1.140188431868	0.72506246761975	0.1616060421017	Objek 17	BERGAS			X
1.0884209908274	0.80476200598802	0.25002642887411	Objek 18	UNGERAN BARAT			X
0.1549041073853	1.0876389157673	1.2078092571226	Objek 19	UNGERAN TIMUR	X		

Gambar 4.5 Iterasi Ketiga

Pada iterasi kedua terjadi perubahan cluster pada kecamatan Bringin. Selanjutnya dilakukan iterasi ketiga untuk mengetahui apakah terjadi perubahan cluster atau tidak. Pada iterasi ketiga sudah tidak terjadi perubahan cluster sehingga iterasi dihentikan. Dan didapatkan bahwa cluster 1 terdiri dari kecamatan Getasan, Tenganan, Susukan, Kaliwungu, Suruh, Pabelan, Tuntang, Banyubiru, Sumowono, Bringin, Bancak, Pringapus, dan Ungaran Timur. Cluster 2 terdiri dari kecamatan Jambu, Bandungan, Bawen. Dan cluster 3 terdiri dari kecamatan Ambarawa, Bergas, dan Ungaran Barat.

HASIL ANALISIS

Cluster 1: GETASAN TENGARAN SUSUKAN KALIWUNGU SURUH PABELAN TUNTANG BANYUBIRU SUMOWONO BRINGIN BANCAN PRINGAPUS UNGARAN TIMUR
 Cluster 2: JAMBU BANDUNGAN BAWEN
 Cluster 3: AMBARAWA BERGAS UNGARAN BARAT

Cluster 1 -> 0.16524216524217 0.5443786982485 0.16722408026756 0.087912087912088
 Cluster 2 -> 0.74074074074074 0.58974358974359 0.65217391304348 0.71428571428571
 Cluster 3 -> 0.91358024691358 0.69871794871795 1 0.23809523809524

Cluster 1 -> Jumlah penderita HIV/AIDS : rendah
 rata-rata umur : rendah
 jumlah penderita hidup : rendah
 jumlah penderita mati : rendah
 Cluster 2 -> Jumlah penderita HIV/AIDS : sedang
 rata-rata umur : sedang
 jumlah penderita hidup : sedang
 jumlah penderita mati : tinggi
 Cluster 3 -> Jumlah penderita HIV/AIDS : tinggi
 rata-rata umur : tinggi
 jumlah penderita hidup : tinggi
 jumlah penderita mati : sedang

Gambar 4.6 Hasil Analisis

Berdasarkan hasil clustering maka didapatkan hasil analisa yang menyatakan beberapa kecamatan di kabupaten Semarang terbagi menjadi 3 cluster yang sebelumnya

terdapat 3 kecamatan yang dijadikan titik pusat cluster yaitu kecamatan Getasan, Jambu, dan Bergas. Kecamatan yang masuk dalam suatu cluster dibagi berdasarkan jarak terdekat dengan titik pusat cluster.

Berdasarkan hasil analisis, maka kecamatan yang termasuk dalam cluster 1 adalah kecamatan Getasan, Tenganan, Susukan, Kaliwungu, Suruh, Pabelan, Tuntang, Banyubiru, Sumowono, Bringin, Bancak, Pringapus, dan Ungaran Timur. Kecamatan – kecamatan tersebut memiliki jumlah penderita HIV/AIDS rendah, rata-rata umur rendah, jumlah penderita hidup rendah, dan jumlah penderita mati rendah. Hal ini berdasarkan perbandingan ketiga titik pusat cluster terakhir yang menyatakan bahwa cluster 1 memiliki penderita HIV/AIDS yang rendah, rata-rata umur rendah, penderita hidup rendah, dan penderita mati juga rendah dibandingkan kedua cluster yang lain.

Anggota cluster 2 memiliki jumlah penderita HIV/ADIS sedang, rata-rata umur sedang, penderita hidup sedang, dan penderita mati tinggi. Hal ini berdasarkan perbandingan titik pusat cluster terakhir yang dibandingkan dengan cluster 1 dan cluster 3. Berdasarkan analisis diatas maka kecamatan yang menjadi anggota cluster 2 adalah kecamatan Jambu, Bandungan, Bawen.

Dari hasil perbandingan ketiga titik pusat cluster terakhir, maka cluster 3 memiliki penderita HIV/AIDS tinggi, rata-rata umur tinggi, dan penderita hidup tinggi dibandingkan cluster lain, sedangkan penderita mati sedang. Dari hasil analisis ini maka didapatkan kecamatan yang menjadi anggota cluster 3 yaitu, Ambarawa, Bergas, dan Ungaran Barat.



Gambar 4.7 Grafik Hasil Clustering

Setelah proses diatas selesai, maka dilakukan penghitungan kualitas cluster. Perhitungan kualitas cluster menggunakan SSE dan BCV. Semakin kecil nilai SSE maka semakin bagus kualitas cluster, sedangkan semakin besar nilai BCV semakin bagus kualitas suatu cluster. Dibawah ini merupakan hasil perhitungan kualitas cluster dengan menggunakan rumus dibawah ini :

$$SSE = \sum_{i=1}^k \sum_{p \in c_i} d(p, m_i)^2$$

$$BCV = \frac{1}{n_k} \sum_{i=1}^k d(m_i, m_i)$$

$$\frac{BCV}{WCV} = \frac{\frac{1}{n_k} \sum_{i=1}^k d(m_i, m_i)}{SSE}$$

```

SSE = (0.15863071248494)2 + (0.23990793664896)2 + (0.59542725924533)2 + (0.59945439232463)2 + (0.11757668391303)2 +
(0.59945439232463)2 + (0.14975114319788)2 + (0.1520376773471)2 + (0.28672285261912)2 + (0.092213153117633)2 +
(0.35429358369976)2 + (0.44399806357693)2 + (0.29522993225349)2 + (0.34923206379902)2 + (0.093262539012182)2 +
(0.35137880037168)2 + (0.18166064210377)2 + (0.25026442807413)2 + (0.1549041073853)2
= 0.02516370294348 + 0.057555818067161 + 0.35453362105241 + 0.3593455684773 + 0.013824276599985 + 0.3593455684773 +
0.022425404889072 + 0.023115455333102 + 0.082209994214044 + 0.0085032656078961 + 0.12552392927907 +
0.19713428046006 + 0.087160712898401 + 0.12196303438532 + 0.0086979011829988 + 0.12346711757126 + 0.026116512843777
+ 0.062513215135541 + 0.023995282484836
= 2.082594661903

BCV = √ { (-0.57549857549858)2 + (-0.045364891518738)2 + (-0.48494983277592)2 + (-0.62637362637363)2 }
+ { (-0.74833808167141)2 + (-0.1543392504931)2 + (-0.83277591973244)2 + (-0.15018315018315)2 }
+ { (-0.17283950617284)2 + (-0.10897435897436)2 + (-0.34783608895652)2 + (0.47619047619048)2 }
= √ { (0.96077684390924 + 0.62957823350835 + 0.62752026012584 + 0.39234391981645 )
+ ( 1.2999011998076 + 0.7388913532792 + 0.71607071108515 + 0.02254978989935 )
+ ( 0.38948926218994 + 0.35961576729587 + 0.347740356382 + 0.22675736961451 ) }
= 0.96077684390924 + 1.2999011998076 + 0.38948926218994
= 1.6279334463997

BCV / SSE = 1.6279334463997 / 2.082594661903
= 0.7816852103674

```

Gambar 4.19 Hasil Akurasi dengan SSE & BCV

V. PENUTUP

A. Kesimpulan

Kesimpulan yang didapat berdasarkan penelitian adalah :

1. Berdasarkan analisis yang dilakukan pada pembahasan maka didapatkan hasil dari analisis tersebut bahwa kecamatan-kecamatan pada Kabupaten Semarang di bagi menjadi 3 cluster.
2. Berdasarkan analisis tersebut maka data yang dihasilkan dari proses clustering dan analisis diharapkan dapat membantu menanggulangi penyebaran virus HIV/AIDS pada Kabupaten Semarang.
3. Dari penelitian yang telah dilakukan dapat disimpulkan bahwa algoritma k-means dapat diterima untuk pengklasteran data tersebut.
4. Membantu dinas kesehatan di Kabupaten Semarang dalam melakukan penyuluhan, konseling, dan juga test HIV/AIDS berdasarkan banyaknya penderita HIV/AIDS pada suatu cluster sehingga diharapkan lebih tepat sasaran.

B. Saran

Berdasarkan kesimpulan diatas, hal yang dapat diharapkan kedepan adalah :

1. Pengolahan data yang lebih besar sehingga aplikasi dapat benar-benar digunakan dan bermanfaat dalam menanggulangi penyebaran virus HIV/AIDS di Kabupaten Semarang.
2. Diharapkan aplikasi tidak hanya bermanfaat untuk Kabupaten Semarang saja namun untuk kabupaten atau kota lain di Indonesia.

3. Mengembangkan GUI (*Graphical User Interface*) yang lebih mudah dan nyaman untuk user dalam menggunakan sistem.

REFERENCES

- [1] Bakara, D.M., Esmianti, F., dan Wulandari, C., "Pengaruh Penyuluhan Kesehatan tentang HIV/AIDS terhadap Tingkat Pengetahuan Siswa di SMA Negeri 1 Selupu Renjang Tahun 2013," Buletin Penelitian Sistem Kesehatan, vol. 17, no. 3, Juli 2014.
- [2] Kementerian Kesehatan RI, "Laporan Situasi Perkembangan HIV & AIDS di Indonesia Tahun 2013," Laporan Perkembangan HIV-AIDS Triwulan I Tahun 2013, Jakarta, 2013.
- [3] [Online]. Available: <http://www.sv-europe.com/crisp-dm-methodology/>. [Accessed 20 November 2015].
- [4] An Ni'Mah, R.M.S., "Sistem Informasi Geografis Visualisasi Clustering Penyakit ISPA di Kecamatan Kaliwungu," Skripsi Teknik Informatika Universitas Dian Nuswantoro, Semarang, 2014.
- [5] Alfina, T., Santosa, B., dan Barakbah, A.R., "Analisa Perbandingan Metode Hierarchical Clustering, K-Means dan Gabungan Keduanya dalam Cluster Data," Jurnal Teknik ITS, vol.1, September 2012.
- [6] [Online]. Available: <http://www.who.int/gho/hiv/en/>. [Accessed 27 Oktober 2015].
- [7] Widyasari, "Analisis Penerapan Metode Single Linkage dan K-Means untuk Pewilayahan Pola Hujan Provinsi Sumatera Selatan Berbasis Arc View," Skripsi Teknik Informatika STMIK GI MDP, Palembang, 2012.
- [8] Hidayah, R., "Sistem Informasi Geografis Berbasis Web untuk Produksi Kedelai di Indonesia Menggunakan K-Means Clustering," Skripsi Departemen Ilmu Komputer Institut Pertanian Bogor, Bogor, 2014.
- [9] Tim Penyusun Buku Situasi HIV/AIDS di Indonesia Departemen Kesehatan RI, Situasi HIV/AIDS di Indonesia Tahun 1987-2006, Jakarta: Pusat Data dan Informasi Departemen Kesehatan RI, 2007.
- [10] Kristanto, O., "Penerapan Algoritma Klasifikasi Data Mining ID3 untuk Menentukan Penjurusan Siswa SMAN 6 Semarang," Skripsi Teknik Informatika Universitas Dian Nuswantoro, Semarang, 2014.
- [11] Hilman, G.Y., Sasmito, B., dan Wijaya, A.P., "Pemetaan Daerah Rawan Kriminalitas di Wilayah Hukum Poltabes Semarang Tahun 2013 dengan Menggunakan Metode Clustering," Jurnal Geodesi Universitas Diponegoro, vol. 4, no. 1, Januari 2015.
- [12] Suryana, N., "Penggunaan Metode Statistik K-Means Clustering pada Analisis Peruntukan Lahan Usaha Tambang Berbasis Sistem Informasi Geografi," Jurnal Statistika Pusat Penelitian dan Pengembangan Teknologi Mineral dan Batubara, vol. 11, no. 1, Mei 2011.
- [13] Hawari, D., Global Effect HIV/AIDS Dimensi Psikoreligi, Jakarta: Fakultas Kedokteran Universitas Indonesia, 2004.

- [14] Hermawati, F.A., Data Mining, Yogyakarta: Andi Offset, 2013.
- [15] Agusta, Y., "K-Means - Penerapan, Permasalahan, dan Metode Terkait," Jurnal Sistem dan Informatika STMIK STIKOM Bali, vol. 3, Februari 2007.
- [16] Suropto, M.A., Triyono, R.A., "Pembangunan Sistem Informasi Akta Kelahiran Dinas Kependudukan dan Perancangan Sipil Sragen," Indonesian Journal on Networking and Security, vol. 3, no. 3, Juli 2014.
- [17] Peranginangin, K., Aplikasi web dengan PHP dan MySQL, Yogyakarta: Andi Yogyakarta, 2006.
- [18] [Online] Available: <http://jordansyahreza.blogspot.co.id/p/pengertian-php-dan-my-sql.html> . [Accessed 29 Januari 2016].
- [19] Caturiyadi, W., "Analisis dan Perancangan Sistem Sales Marketing Tools (SMART) pada Unit Data dan Internet Sales & Promotion PT. Telkom Kandatel Semarang," Transient Universitas Diponegoro Semarang, vol. 1, no. 4, Desember 2012.
- [20] [Online] Available : meilisdasari.blogspot.co.id/2012/04/apa-itu-use-case-activity-diagram-dan.html?m=1. [Accessed 12 Mei 2016]
- [21] [Online] Available : <https://zarisbudi.wordpress.com/2013/05/13/kebijakan-tentang-pengendalian-hiv-aids-di-indonesia>. [Accessed 23 Juni 2016]