

IMPLEMENTASI ALGORITMA *K-MEANS* UNTUK *CLUSTERING* PENYAKIT HIV/AIDS DI INDONESIA

IMPLEMENTATION OF *K-MEANS* ALGORITHM FOR *CLUSTERING OF HIV/AIDS DISEASE IN INDONESIA*

Andi Andrea Lesmana¹, Yudha Purwanto, S.T., M.T.², Ashri Dinimaharawati, S.pd., M.T.³

^{1,2,3}Prodi S1 Teknik Komputer, Fakultas Teknik Elektro, Universitas Telkom

andremanapra@telkomuniversity.ac.id¹, omyudha@telkomuniveristy.ac.id², ashridini@telkomuniversity.ac.id³

Abstrak

Peningkatan kasus HIV / AIDS di Indonesia telah menjadi kasus yang tidak pernah lolos dari kota. Perhatian terhadap angka kematian terus meningkat membuat orang khawatir tentang penyebaran HIV / AIDS. HIV / AIDS hanya dapat ditularkan melalui hubungan seksual dari orang yang terinfeksi, transfusi darah yang terkontaminasi dengan HIV / AIDS, dan dari perempuan yang terinfeksi HIV kepada anaknya selama kehamilan, persalinan, dan menyusui. Di Indonesia, jumlah kasus HIV / AIDS membuat instansi terkait harus bertindak cepat dan akurat. Konseling dan tes HIV / AIDS dapat dilakukan di Indonesia berdasarkan provinsi dengan pasien tertinggi. Selain itu, ini akan diintegrasikan oleh jumlah kluster orang dengan HIV / AIDS menggunakan algoritma K-Means yang akan membantu pihak terkait untuk mengambil tindakan terhadap HIV / AIDS

Pada penelitian ini, membuat aplikasi berbasis web. Aplikasi dibuat dengan menggunakan PHP sebagai bahasa pemrograman, MySql sebagai ruang penyimpanan data dan CodeIgneter3 sebagai framework dari aplikasi ini. Diharapkan bisa mempermudah dalam pembuatan aplikasi tersebut. Aplikasi ini akan mengolah data dengan menggunakan algoritma K-Means. Selanjutnya, aplikasi berbasis web ini akan diintegrasikan dengan mengelompokkan jumlah orang dengan HIV / AIDS menggunakan algoritma K-Means yang akan membantu pihak terkait untuk mengambil tindakan terhadap HIV / AIDS di Indonesia.

Keyword : clustering, HIV/AIDS, K-MEANS, data mining

Abstract

Increased cases of HIV / AIDS in Indonesia have become cases that have never escaped from the city. Attention to the increasing mortality rate makes people worried about the spread of HIV / AIDS. HIV / AIDS can only be transmitted through sexual contact from an infected person, blood transfusions contaminated with HIV / AIDS, and from HIV-infected women to their children during pregnancy, childbirth and breastfeeding. In Indonesia, the number of HIV / AIDS cases makes the relevant agencies must act quickly and accurately. HIV / AIDS counseling and testing can be done in province in Indonesia based province with the highest patients. In addition, will be integrated by the number of clusters of people with HIV / AIDS using the K-Means algorithm which will help related parties take action against HIV / AIDS.

In this study, creating a web-based application. The application is created using PHP as a programming language, MySql as data storage space and CodeIgneter3 as a framework of this application. Expected to be expected in making the application. This application will process data using the K-Means algorithm. Furthermore, this web-based application will be integrated by grouping the number of people with HIV / AIDS using a K-Mean algorithm that will help relevant parties to take action against HIV / AIDS in Indonesia..

Keyword: clustering, HIV/AIDS, K-MEANS, data mining

Pendahuluan

Menurut data dari World Health Organization(WHO) dan Joint United Nations Programme on HIV/AIDS(UNAIDS), sebanyak 33.2 juta orang yang hidup dengan HIV yang terdiri daripada 30.8 juta orang dewasa, 15.4 juta orang wanita dan 2.1 juta orang anak – anak di bawah usia 15 tahun. Lebih kurang 6800 infeksi HIV baru dalam sehari dalam tahun 2007 yang terdiri dari 5800 dewasa di mana hampir 50% adalah wanita, dan 40% terdiri dari golongan muda yang berumur antara 15–24 tahun. Jumlah penderita lebih kurang 1200 orang anak–anak berumur di bawah 15 tahun dan lebih 96% dari negara golongan pendapatan rendah dan sederhana [1].

Pada anak–anak yang didiagnosa AIDS ketika berumur kurang dari 13 tahun, 90% dari mereka mendapat infeksi melalui ibu mereka yang terinfeksi HIV ke fetus atau anak yang baru lahir[2].

Di Indonesia, jumlah kasus kumulatif menurut faktor resiko yang terbanyak adalah transmisi melalui heteroseksual ke heteroseksual yaitu sebanyak 9166 kasus dan menurut golongan umur pula menunjukkan umur di antara 20 hingga 29 tahun yang terbanyak yaitu 9142 kasus dari data sehingga 2009[3].

Menurut UNAIDS, di Indonesia ada sekitar 690 ribu orang pengidap HIV sampai tahun 2015. Dari jumlah tersebut, setengah persennya berusia antara 15 hingga 49 tahun. Wanita usia 15 tahun ke atas yang hidup dengan kondisi HIV sekitar 250 ribu jiwa. Angka kematian akibat AIDS mencapai 35 ribu orang.

1. Dasar Teori

1.1 HIV/AIDS

HIV (*Human Immuno Deficiency Virus*) adalah virus penyebab AIDS yang menyerang system kekebalan tubuh manusia sehingga tidak mampu melindungi dari serangan penyakit lain[4]. HIV yaitu virus yang merusak sistem kekebalan tubuh manusia [4]. AIDS (Acquired Immuno Deficiency Syndrome) adalah kumpulan dari beberapa gejala penyakit akibat menurunnya sistem kekebalan tubuh yang disebabkan oleh HIV [4]. AIDS adalah sekumpulan gejala yang diakibatkan oleh menurunnya sistem kekebalan tubuh manusia karena terinfeksi HIV [4].

AIDS adalah suatu sindrom penyakit defisiensi imunitas selular yang didapat, yang pada penderitannya tidak dapat ditemukan penyebab defisiensi tersebut [5].

AIDS merupakan gangguan immunodefisiensi yang sekunder yang disebabkan oleh retrovirus (HIV) yang telah terisolasi dalam cairan tubuh orang yang terinfeksi [6].

1.2 Clustering

Clustering data dapat dibedakan menjadi dua tujuan [9], yaitu *clustering* untuk pemahaman dan *clustering* untuk penggunaan. Jika tujuan untuk pemahaman maka kluster yang terbentuk harus menangkap struktur alami data.

Biasanya proses *clustering* dalam tujuan ini hanya sebagai proses awal untuk kemudian dilanjutkan dengan pekerjaan ini seperti *summarization* (rata-rata, standar deviasi), pelabelan kelas pada setiap kelompok untuk kemudian digunakan sebagai data latih klasifikasi, dan sebagainya. Sementara jika tujuannya untuk penggunaan, biasanya tujuan utama untuk mencari *prototype* kluster yang paling representative terhadap data dan memberikan abstraksi dan setiap objek data dalam *cluster* di mana sebuah data terletak didalamnya.

Banyak metode *clustering* yang sudah dikembangkan oleh para ahli. Masing – masing metode mempunyai karakter, kelebihan, dan kekurangan. *Clustering* dapat dibedakan menurut stuktur cluster, keanggotaan data dalam cluster dan kekompakan data dalam cluster.

Metode *clustering* menurut strukturnya dibagi menjadi dua yaitu pengelompokan hirarki dan *partitioning*. Pengelompokan hirarki memiliki aturan satu data tunggal bisa dianggap sebagai sebuah kelompok, dua atau lebih kelompok kecil dapat bergabung menjadi satu kelompok besar dan begitu seterusnya hingga semua data dapat bergabung menjadi satu kelompok. Metode *clustering* hirarki merupakan satu-satunya metode yang masuk ke dalam kategori pengelompokan hirarki. Metode *clustering* *partitioning* membagi set data ke dalam sejumlah kelompok yang tidak tumpang tindih (*overlap*) antara satu kelompok dengan kelompok yang lain artinya setiap data hanya menjadi anggota satu kelompok. Metode seperti K-Means dan DBSCAN masuk dalam kategori pengelompokan *partitioning*.

Metode *clustering* menurut keanggotaan dalam kelompok dibagi menjadi dua, yaitu eksklusif dan tumpang-tindih. Metode tersebut termasuk kategori eksklusif jika sebuah data hanya menjadi anggota satu kelompok dan tidak menjadi anggota kelompok yang lain. Metode *clustering* yang masuk dalam kategori ini adalah K-Means dan DBSCAN sedangkan yang masuk kategori tumpang – tindih adalah metode *clustering* yang membolehkan sebuah data menjadi anggota di lebih dari satu kelompok, misalnya Fuzzy C-Means.

Clustering terbagi menjadi beberapa yaitu[6]:

1.Partitioning clustering

- a) Disebut juga exclusive clustering
- b) Setiap data harus termasuk kedalam cluster tertentu
- c) Mengungkinkan bagi setiap data yang termasuk cluster tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke cluster yang lain.

2.Hierarchical clustering

- a) Setiap data harus masuk kedalam cluster tertentu
- b) Suatu data termasuk ke dalam cluster tertentu pada suatu proses, tidak dapat berpindah ke cluster lain.
- c) contoh: *Single Linkage, Centroid Linkage, Complete Linkage, Average Linkage*.

3.Overlapping Clustering

- a) mempunyai nilai keanggotaan pada beberapa cluster
- b) Setiap Data data memungkinkan termasuk ke beberapa cluster
- c) Contoh : Fuzzy C-means, Gaussian Mixture

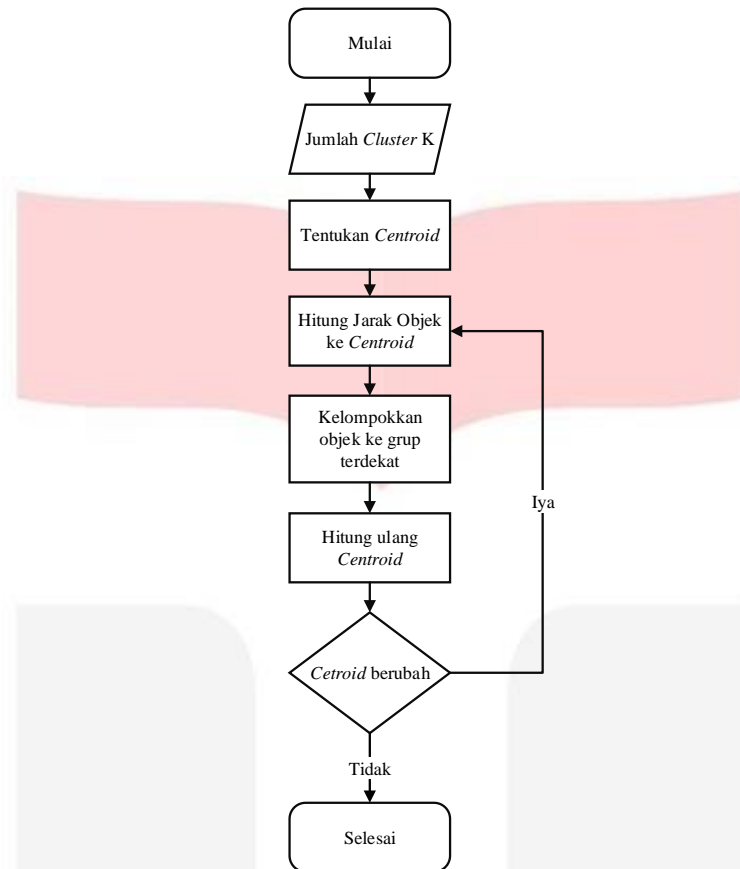
4. Hybrid

- a) Merupakan kombinasi dari karakteristik *partitioning*, *overlapping* dan *hierarchical*[3].

1.3 K-means Clustering

Partitioning clustering memiliki sifat semua data akan masuk ke dalam cluster-cluster yang terbentuk. Setiap data yang masuk kedalam suatu *cluster* memungkinkan pada suatu tahapan proses berikutnya untuk berpindah ke cluster lain. K-Means termasuk kedalam *partitioning clustering*. Means dalam hal ini berarti nilai suatu rata-rata dari suatu grup data yang dalam hal ini didefinisikan sebagai *cluster*[6].

K-means merupakan salah satu dari metode data *clustering* non hirarki yang berusaha mempartisi data yang ada ke dalam satu atau lebih *cluster* atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan kedalam satu *cluster* yang sama dan data yang memiliki karakteristik berbeda dikelompokkan ke dalam kelompok yang lain. Adapun tujuan dari K-Means *Clustering* ini adalah meminimalisasikan *objective function* yang di set dalam proses *clustering*, yang pada umumnya berusaha meminimalisasikan variasi dalam suatu *cluster* dan meminimalisasikan variasi antar *cluster*[7].



Gambar 2.1 Flowchart dari algoritma K-Means

1. Pilih jumlah *cluster* “K”.
2. Menentukan pusat *cluster*. Hal ini dapat dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat *cluster* diberi nilai awal dengan angka-angka random.
3. Mengalokasikan semua data/objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *cluster*. Jarak paling antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster* mana. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak Euclidean .
4. Melakukan penghitungan kembali pusat cluster dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/objek dalam *cluster* tertentu.
5. Langkah kelima adalah menugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah proses *clustering* selesai.

Algoritma untuk melakukan K-Means *Clustering* dapat dilihat pada gambar 2.1 dimulai dengan pemilihan K buah titik pusat (*centroid*) secara acak. Setelah itu, kelompokkan data sehingga terbentuk K buah *cluster* dengan titik *centroid* dari setiap cluster merupakan titik *centroid* yang telah dipilih sebelumnya dan perbaharui nilai titik *centroid*. Lakukan mengulangi hingga nilai dari titik *centroid* tidak lagi berubah.

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[8] :

.....(1)

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

Keterangan :

- dik : jarak data ke-i
 n : jumlah variable
 x_{ij} : data yang akan di *cluster*
 c_{kj} : pusat dari *cluster*

Algoritma K-Means merupakan algoritma pengelompokan *iterative* yang melakukan partisi set data ke dalam sejumlah K *cluster* yang sudah ditetapkan di awal. Algoritma K-Means sederhana untuk diimplementasikan dan dijalankan, *relative* cepat, mudah beradaptasi, umum penggunaannya dalam praktek. Secara historis, K-Means menjadi salah satu algoritma yang paling penting dalam bidang data *mining* [11].

K-means merupakan salah satu algoritma *clustering*[12]. Tujuan algoritma ini yaitu untuk membagi data menjadi beberapa kelompok. [13].

Pada algoritma pembelajaran ini, komputer mengelompokkan sendiri data-data yang menjadi masukannya tanpa mengetahui terlebih dulu target kelasnya[12].

Pembelajaran ini termasuk dalam *unsupervised learning*. Masukan yang diterima adalah data atau objek dan k buah kelompok (*cluster*) yang diinginkan. Algoritma ini akan mengelompokkan data atau objek ke dalam k buah kelompok tersebut. Pada setiap cluster terdapat titik pusat (*centroid*) yang merepresentasikan *cluster* tersebut.

K-means ditemukan oleh beberapa orang yaitu Lloyd (1957, 1982), Forgey (1965) , Friedman and Rubin (1967) , and McQueen (1967) [12]. Ide dari *clustering* pertama kali ditemukan oleh Lloyd pada tahun 1957, namun hal tersebut baru dipublikasi pada tahun 1982. Pada tahun 1965, Forgey juga mempublikasikan teknik yang sama sehingga terkadang dikenal sebagai Lloyd-Forgey pada beberapa sumber.

3.Pembahasan

3.1 Gambaran Umum Sistem

Metode yang di gunakan pada sistem ini yaitu metode pengelompokan data K-means *Clustering* lalu di implementasi berbasis web.

Aplikasi ini di buat dengan Bahasa pemograman PHP dan MYSQL sebagai sumber penyimpanan data. Menggunakan data yang sudah ada , lalu data akan di olah dengan metode K-Means *Clustering* yang dimana nanti nya akan di analisa yang di bagi menjadi beberapa *cluster* yang meliputi beberapa pengelompokan yaitu daerah-daerah yang tinggi, sedang, dan rendah nya serta menganalisa baik dan buruk nya daerah dalam penanganan penyakit HIV/AIDS di Indonesia

3.2 Spesifikasi perangkat

Perangkat yang digunakan dalam penelitian ini adalah laptop dengan spesifikasi sebagai berikut:

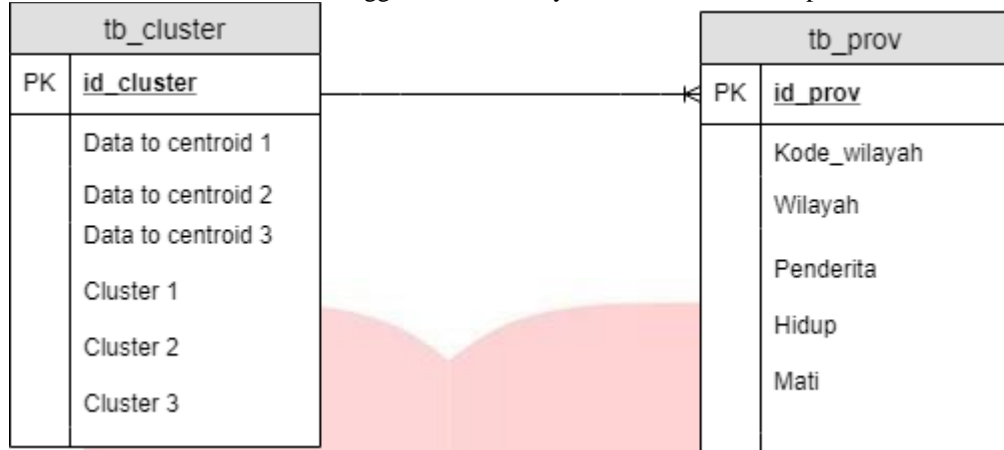
- *Processor* : Intel Core i5-7200 CPU @2.5 Ghz (4CPUs)~ 2.7 Ghz
- *Memory* : 8 GB 2400 MHz DDR4
- *Graphics* : Intel HD Graphics 620
- *Sistem Operasi* : Windows 10 Pro 64-bit (10.0, Built 17763)

Spesifikasi perangkat lunak :

- *Framework* : CodeIgneter 3
- *Database* : MYSQLI
- *Bahasa Pemograman* : PHP 7

3.3 Desain basis data

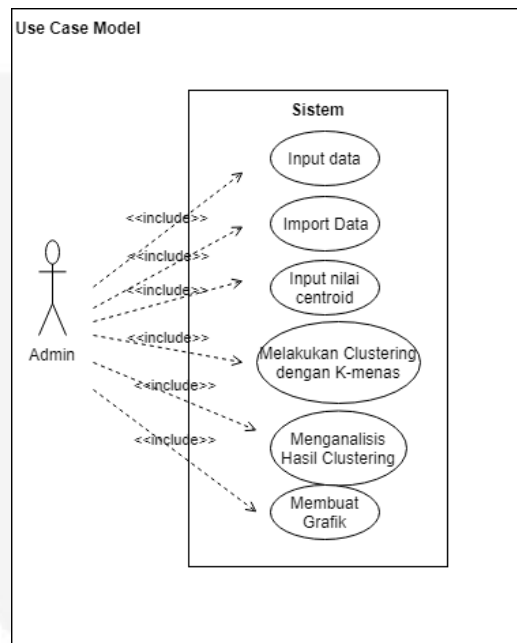
Penyimpanan dari basis data ini menggunakan MYSQL yang sebagai kontrol nya yaitu PHP dan codeigneter3 sebagai *framework*. Pada basis data kali ini menggunakan 2 tabel yaitu *tb_cluster* dan *tb_prov*.



Gambar 3.3.1 *tb cluster_tb_prov*

Pada gambar diatas menunjukkan relasi antara tabel *cluster* dan tabel provinsi

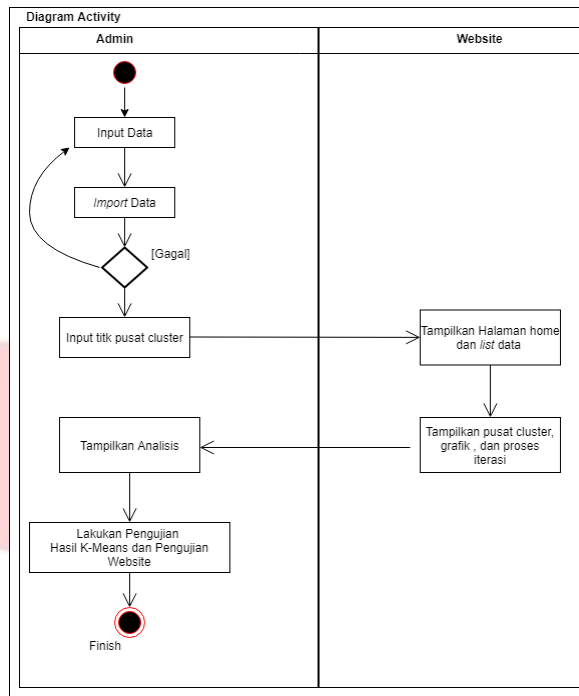
3.4 Use Case Model



Gambar 3.4.1. *Use case*

Gambar 3.4.1 menunjukkan bahwa admin dapat melakukan pengoprasian fitur yang terdapat pada aplikasi yaitu input data, *import* data, input nilai centroid, melakukan *clustering*, menganalisis hasil *clustering* dan membuat grafik

3.5. Diagram Activity

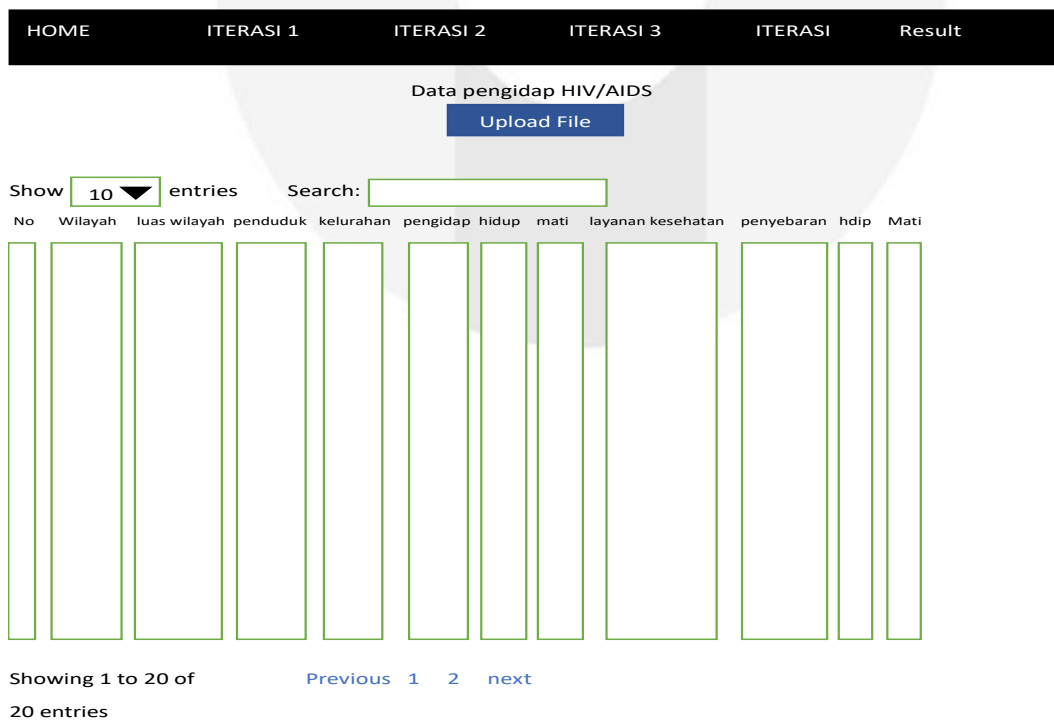


Gambar 3.7.1. Diagram Activity

Gambar 3.7.1 merupakan aktivitas yang di lakukan user dan sistem untuk mendapatkan hasil *clustering* menggunakan K-means dan melakukan analisis dari hasil tersebut.

3.6 Rancangan Interface

Tampilan awal pada menu *home*



Gambar 3.6.1 Interface Home

Gambar 3.8.1 merupakan tampilan pada halaman utama yang memiliki fitur.

1. *Home* untuk menampilkan halaman utama
2. Iterasi merupakan proses *cluster* data hingga mencapai hasil
3. *Show* untuk menampilkan data daerah yang meliputi pengidap, hidup dan mati , pada menu show ini memiliki beberapa pilihan untuk menampilkan dalam jumlah baris yang banyak nya 10,20,50,100.
4. *Upload file* untuk memperbarui data atau mengunggah data yang baru.
5. *Previous dan next* untuk merupakan bagian dari fitur untuk menampilkan data daerah.
6. Diagram untuk menunjukan jumlah cluster yang paling tinggi pengidap nya dan cluster daerah yang rata-rata tinggi jumlah pengidap nya.

Tampilan pada menu result

HOME	ITERASI 1	ITERASI 2	ITERASI 3	ITERASI	Result
GRAFIK					
HASIL CLUSTER	Cluster		Penyebaran	Hidup	Mati
	1				
	2				
Notes	3				
Showing 1 to 3 of 3 entries					Previous 1 Next

Gambar 3.6.2 Interface Result

Fitur yang terdapat pada tampilan *result* :

1. Diagram Grafik untuk menunjukan jumlah cluster yang paling tinggi pengidap nya dan cluster daerah yang rata-rata tinggi jumlah pengidap nya.
2. Hasil Cluster merupakan hasil pengelompokan data yang anggota nya merupakan wilayah kota dan kabupaten.
3. Notes merupakan keterangan dari penyebaran , hidup dan mati.
4. Penyebaran , hidup , mati berupa persentase hasil dari korelasi dari dataset

Tampilan pada proses iterasi awal sampai akhir .

HOME	ITERASI 1	ITERASI 2	ITERASI 3	ITERASI	ITERASI ~
------	-----------	-----------	-----------	---------	-----------

CENTROID	DATA HASIL CLUSTER
CLUSTER 1 = ANGKA CENTROID CLUSTER 2 = ANGKA CENTROID CLUSTER 2 = ANGKA CENTROID	CLUSTER 1 = DAERAH YANG SUDAH DI KELOMPOKAN CLUSTER 2 = DAERAH YANG SUDAH DI KELOMPOKAN CLUSTER 2 = DAERAH YANG SUDAH DI KELOMPOKAN

Show entries Search:

Data to centroid 1	Data to centroid 2	Data to centroid 3	Provinsi	Penderita	Hidup	Mati	Cluster 1	Cluster 2	Cluster 3
ANGKA JARAK CENTROID	ANGKA JARAK CENTROID	ANGKA JARAK CENTROID	NAMA PROVINSI	JUMLAH PENDERITA	JUMLAH PENDERITA	JUMLAH PENDERITA	DATA VISUAL COLORED BOX		

Showing 1 to 20 of [Previous](#) [1](#) [2](#) [next](#)
20 entries

Gambar 3.8.2 *Interface* Iterasi lanjutan

Gambar 3.8.2 menunjukkan menu selanjut dengan penjelasan sebagai berikut:

1. *Centroid* adalah angka yang di gunakan sebagai titik pusat awal yang di gunakan untuk mendapatkan nilai *data to centroid* dan centroid ini di pilih secara acak oleh sistem.
2. Data hasil *cluster* merupakan daerah-daerah yang sudah di *cluster* namun pada iterasi pertama hingga sebelum terakhir daerah yang di tampilan bersifat *non fixed* hanya tampilan proses untuk mendapatkan hasil daerah yang sudah berhasil di cluster yang terdapat pada iterasi akhir.

4. Analisis

Dari hasil penelitian yang sudah di lakukan bahwa program ini melakukan cluster data menjadi seperti berikut : Pada analisis menggunakan 5 data sampel pada setiap cluster.

Tabel 4.1.1 Cluster dengan tingkat kerawanan yang tinggi

No	Wilayah	Luas Wilayah	Penduduk	Kelurahan	Pengidap HIV/AIDS	Hidup	Mati	Layanan Kesehatan	Penyebaran	Hidup	Mati
1	Kab. Mimika	21,633 km ²	306,517 orang	152	1732	1553	179	13	0.002%	0.005%	0.001%
2	Kota Jayapura	11,157 km ²	165,404 orang	144	863	808	55	19	0.001%	0.010%	0.002%
3	Kota Surabaya	351 km ²	2,827,892 orang	154	2167	1189	978	63	0.012%	0.027%	0.025%
4	Kab. Nabire	11,113 km ²	166,463 orang	81	2120	1869	251	26	0.003%	0.011%	0.004%
5	Kab. Jayawijaya	7,031 km ²	268,137 orang	332	1769	1048	721	13	0.001%	0.009%	0.006%

Tabel 4.1.2 Cluster dengan tingkat kerawanan yang sedang

No	Wilayah	Luas Wilayah	Penduduk	Kelurahan	Pengidap HIV/AIDS	Hidup	Mati	Layanan Kesehatan	Penyebaran	Hidup	Mati
1	Kab. Bogor	2,711 km ²	4,246,307 orang	435	26	19	7	101	0.009%	0.036 %	0.018 %
2	Kota Bandar Lampung	296 km ²	1,175,397 orang	126	609	554	55	12	0.003%	0.034 %	0.006 %
3	Kab. Bandung	1,768 km ²	3,522,724 orang	280	30	26	4	62	0.013%	0.062 %	0.020 %
4	Kota Jambi	104 km ²	60,962 orang	62	524	363	161	20	0.032%	0.098 %	0.061 %
5	Kota Bekasi	207 km ²	2,409,083 orang	56	629	569	60	22	0.019%	0.063 %	0.024 %

Tabel 4.1.3 Cluster dengan tingkat kerawanan yang rendah

No	Wilayah	Luas Wilayah	Penduduk	Kelurahan	Pengidap HIV/AIDS	Hidup	Mati	Layanan Kesehatan	Penyebaran	Hidup	Mati
1	Kota Jakarta Selatan	141 km ²	2,185,711 orang	65	202	178	24	78	0.850%	1.473 %	0.934 %
2	Kota Jakarta Timur	188 km ²	2,843,816 orang	65	411	339	72	88	0.722%	1.159 %	0.814 %
3	Kota Jakarta Pusat	48 km ²	914,182 orang	44	267	233	34	42	1.984%	3.796 %	2.25%
4	Kota Jakarta Barat	130 km ²	2,463,560 orang	56	223	192	31	75	1.034%	1.699 %	1.141 %
5	Kota Jakarta Utara	147 km ²	1,747,315 orang	31	119	100	19	49	1.078%	1.651 %	1.187 %

Berdasarkan hasil penelitian dapat di analisis bahwa wilayah Kab. Mimika, Kota Jayapura, Kota Surabaya, Kab. Nabire, Kab. Jayawijaya merupakan daerah yang paling tinggi rawan mengidap penyakit HIV/AIDS, dilihat dari jumlah penderitanya yang rata-rata ribuan dan juga hasil korelasi yang di dapat menunjukkan bahwa penanganan terhadap penyakit HIV/AIDS pada wilayah-wilayah tersebut kurang baik karena dari hasil korelasi antara luas wilayah, jumlah penduduk, jumlah kelurahan serta jumlah layanan kesehatan terhadap persentase hidup yang rendah dan korelasi mati yang cukup tinggi jika di bandingkan dengan daerah-daerah lainnya. Persentase hidup yang rendah menunjukkan bahwa persentase kemungkinan untuk hidup bagi penderita juga rendah begitupun juga persentase mati. Mungkin hal ini disebabkan karena layanan kesehatan pada daerah-daerah tersebut sangat kurang jika di bandingkan dengan luas wilayah

pada daerah-daerah tersebut yang sangat besar.

Lalu dilihat dari Tabel 4.1.3 Cluster dengan tingkat kerawanan yang rendah yang anggota wilayah Kota Jakarta Selatan, Kota Jakarta Timur, Kota Jakarta Pusat, Kota Jakarta Barat, Kota Jakarta Utara mendapatkan hasil korelasi hidup yang sangat tinggi dan korelasi mati yang rendah, hal ini disebabkan karena dengan luas wilayah yang kecil namun memiliki layanan kesehatan yang banyak sehingga pengidap penyakit HIV/AIDS dapat di tangani dengan baik.

Dari hasil penelitian ini dapat disimpulkan bahwa angka korelasi tersebut dapat menjadi tolak ukur untuk evaluasi pemerintah terhadap pembangunan infrastruktur kesehatan secara merata serta menyeluruh pada daerah-daerah di Indonesia agar dapat mencegah dan menanggulangi aspek-aspek yang menyebabkan penyebaran penyakit HIV AIDS di Indonesia.

Tabel 4.2.3 Nilai rata-rata persentase korelasi

Percobaan	Penyebaran%			Hidup%			Mati%		
	Rendah	Sedang	Tinggi	Rendah	Sedang	Tinggi	Rendah	Sedang	Tinggi
1	0.002811	0.005465	0.005154	0.02182	0.03986	0.0173	0.00719	0.01352	0.01445
2	0.002814	0.0054658	0.005154	0.02189	0.03986	0.0173	0.00719	0.01352	0.01445
3	0.002808	0.005465	0.005154	0.02176	0.03981	0.0173	0.00706	0.01357	0.01445
4	0.002813	0.005467	0.005154	0.02178	0.03979	0.0173	0.00708	0.01351	0.01445
5	0.002813	0.005469	0.005154	0.02182	0.03827	0.0173	0.00721	0.01288	0.01445
6	0.002813	0.005469	0.005154	0.02182	0.03827	0.0173	0.00721	0.01288	0.01445
7	0.002805	0.005603	0.005154	0.02189	0.03896	0.0173	0.00707	0.01328	0.01445
8	0.002785	0.005773	0.005154	0.02176	0.03981	0.0173	0.00706	0.01357	0.01445
9	0.002785	0.005773	0.005154	0.02176	0.03981	0.0173	0.00706	0.01357	0.01445
10	0.002805	0.005603	0.005154	0.02189	0.03896	0.0173	0.00707	0.01328	0.01445
11	0.002813	0.005469	0.005154	0.02182	0.03827	0.0173	0.00721	0.01288	0.01445
12	0.002813	0.005469	0.005154	0.02182	0.03827	0.0173	0.00721	0.01288	0.01445
13	0.002805	0.005603	0.005154	0.02189	0.03896	0.0173	0.00707	0.01328	0.01445
14	0.002813	0.005469	0.005154	0.02182	0.03827	0.0173	0.00721	0.01288	0.01445
15	0.002785	0.005773	0.005154	0.02176	0.03981	0.0173	0.00706	0.01357	0.01445
16	0.002781	0.005763	0.005154	0.02172	0.03987	0.0173	0.00708	0.01352	0.01445
17	0.002781	0.00571	0.005154	0.02172	0.03985	0.0173	0.00708	0.01352	0.01445
18	0.002782	0.005921	0.005154	0.02169	0.04061	0.0173	0.00708	0.01392	0.01445
19	0.002789	0.005928	0.005154	0.02166	0.04069	0.0173	0.00708	0.01392	0.01445
20	0.002781	0.005925	0.005154	0.02162	0.04063	0.0173	0.00708	0.01392	0.01445
21	0.002813	0.005469	0.005154	0.02182	0.03827	0.0173	0.00721	0.01288	0.01445
22	0.002809	0.00542	0.005154	0.02185	0.03823	0.0173	0.00723	0.01281	0.01445
23	0.002813	0.005469	0.005154	0.02182	0.03827	0.0173	0.00721	0.01288	0.01445
24	0.002811	0.005467	0.005154	0.02184	0.03825	0.0173	0.00722	0.01286	0.01445
25	0.002808	0.005461	0.005154	0.02174	0.03822	0.0173	0.00729	0.01283	0.01445
26	0.002785	0.005773	0.005154	0.02176	0.03981	0.0173	0.00706	0.01357	0.01445
27	0.002813	0.005469	0.005154	0.02182	0.03827	0.0173	0.00721	0.01288	0.01445
28	0.002811	0.005466	0.005154	0.02184	0.03828	0.0173	0.00724	0.01286	0.01445
29	0.002816	0.005463	0.005154	0.02181	0.03826	0.0173	0.00722	0.01284	0.01445
30	0.002813	0.005469	0.005154	0.02182	0.03827	0.0173	0.01288	0.00721	0.01445
RATA-RATA	0.002802 9	0.0055836 27	0.005154	0.02179 4333	0.03909 2	0.0173	0.00733 7667	0.01304 9667	0.01445
STD DEV	0.000013 08421	0.0001660 2369	0.000000 00000	0.00006 698859	0.00069 158489	0.0000 00000	0.00104 948816	0.00116 612676	0.00000 000000

Pada tabel 4.2.3 dapat dilihat rata-rata nilai persentase korelasi yang didapatkan dari hasil korelasi penyebaran terhadap penyakit HIV AIDS, korelasi penderita HIV AIDS yang hidup dan korelasi penderita HIV AIDS yang meninggal dunia dari 30 kali percobaan. Lalu masing masing korelasi dan label yang sama di jumlahkan hingga mendapatkan nilai rata-rata korelasi tersebut.

Untuk nilai korelasi penyebaran penyakit HIV AIDS didapatkan dari rumus (2).

Penyebaran = ((Jumlah penderita/total jumlah penduduk) + (Jumlah tempat layanan kesehatan/Jumlah kelurahan))/Luas

wilayah)*100(2)

Untuk nilai korelasi penderita yang hidup terhadap penyakit HIV AIDS di dapatkan dari rumus(3).

Hidup = ((Jumlah penderita hidup/total jumlah penderita) + (Jumlah tempat layanan kesehatan/Jumlah kelurahan))/Luas wilayah)*100(3)

Untuk nilai korelasi penderita yang hidup terhadap penyakit HIV AIDS di dapatkan dari rumus(4).

Mati = ((Jumlah penderita mati/total jumlah penderita) + (Jumlah tempat layanan kesehatan/Jumlah kelurahan))/Luas wilayah)*100(4)

Berdasarkan tabel 4.2.3 hasil diatas menunjukkan bahwa perhitungan yang dilakukan program cukup stabil dilihat dari standar deviasi yang kecil pada hasil korelasi penyebaran pada cluster rendah nilai standar deviasi 0.00001308421 dan cluster sedang dengan nilai standar deviasi 0.00016602369 dan pada cluster tinggi mendapatkan nilai varians yang sangat stabil yaitu 0.00000000000.

Pada hasil korelasi hidup nilai standar deviasi yang didapat kurang stabil karena angka standar deviasi pada cluster rendah yaitu 0.00006698859 dan pada cluster sedang 0.00069158489 namun pada cluster tinggi nilai varians stabil dengan nilai 0.00000000000

Pada korelasi mati cukup stabil karena nilai standar deviasi pada cluster rendah mendapatkan nilai 0.00104948816 dan pada cluster sedang mendapat nilai varians 0.00116612676namun pada cluster tinggi stabil karena dilihat dari hasil nilai varians yaitu 0.00000000000

5. Pengujian

4.1 Pengujian *BlackBox*

Tabel 4.4.1 Hasil Pengujian *BlackBox*

ID Pengujian	Deskripsi Pengujian	Prosedur Pengujian	Keluaran yang diharapkan	Hasil yang Didapat	Hasil uji	
					Diterima	Ditolak
A.1	Tombol <i>Home</i>	Menekan tombol <i>Home</i>	Aktif	Home aktif	√	
A.2	Tombol Iterasi	Menekan tombol Iterasi	Aktif	Iterasi aktif	√	
A.3	Tombol <i>Upload File</i>	Menekan <i>Upload File</i>	Aktif	<i>Upload File</i> aktif	√	
A.5	Tombol <i>Search</i>	Menekan tombol <i>Search</i>	Aktif	Search Aktif	√	
A.5.5	Tombol <i>Previous</i>	Menekan <i>Previous</i>	Aktif	Kembali kehalaman sebelumnya	√	
A.6	Tombol <i>Next</i>	Menekan tombol <i>Next</i>	Aktif	Lanjut Kehalaman berikut	√	
A.7	Tombol <i>Show</i>	Menekan tombol <i>Show</i>	Aktif	Menampilkan jumlah baris pilihan	√	
A.8	Diagram	Arahkan kursor ke diagram	Aktif	Menampilkan jumlah penderita	√	

Berdasarkan tabel 4.4.1 menunjukkan hasil pengujian program dalam penggunaan fitur yang terdapat pada website telah berhasil dilakukan dan berjalan dengan baik sesuai dengan yang diharapkan

4.2 Pengujian Beta

Hasil pengujian beta

A. Aspek Sistem

1. Fitur website sudah lengkap.

Tabel 4.5.1 Pengujian Beta Pertanyaan 1

Point Jawaban	Frekuensi Jawaban	Jumlah Sampel	Persentase (%)	Keterangan
Cukup Lengkap	6	30	20	Tidak Baik
Lengkap	21		70	Baik
Sangat Lengkap	3		10	Baik

Berdasarkan tabel 4.5.1 hasil responden terhadap website ini yaitu baik karena 80% responden dengan keterangan baik hal itu menandakan sebagian besar respon menilai website ini sudah cukup lengkap.

2. Tampilan dari web program *clustering*.

Tabel 4.5.2 Pengujian Beta Pertanyaan 2

Point Jawaban	Frekuensi Jawaban	Jumlah Sampel	Persentase (%)	Keterangan
Sangat buruk	0	30	0	Tidak Menarik
Buruk	0		0	Tidak Menarik
Cukup Baik	2		6.7	Menarik
Baik	18		60	Menarik
Sangat Baik	10		33.3	Menarik

Berdasarkan tabel 4.5.2 hasil responden terhadap *website* ini yaitu menarik karena 100% responden dengan keterangan mengerti hal itu menandakan bahwa tampilan dari *website* ini menarik

B. Aspek Pengguna

3. tampilan *website* mudah di mengerti

Tabel 4.5.3 Pengujian Beta Pertanyaan 3

Point Jawaban	Frekuensi Jawaban	Jumlah Sampel	Persentase (%)	Keterangan
Sangat Sulit	0	30	0	Tidak Mengerti
Sulit	0		0	Tidak Mengerti
Cukup Mudah	4		13.3	Mengerti
Mudah	18		60	Mengerti
Sangat Mudah	8		26.7	Mengerti

Berdasarkan tabel 4.5.3 hasil responden terhadap *website* ini yaitu mengerti karena 100% responden dengan keterangan mengerti hal itu menandakan bahwa tampilan dari *website* ini mudah di mengerti

4. Fitur yang ada di *website* mudah di pahami.

Tabel 4.5.4 Pengujian Beta Pertanyaan 4

Point Jawaban	Frekuensi Jawaban	Jumlah Sampel	Persentase (%)	Keterangan
Cukup Mudah	3	30	10	Tidak paham
Mudah	6		70	Pahami
Sangat Mudah	21		20	Pahami

Berdasarkan tabel 4.5.4 hasil responden terhadap *website* ini yaitu dapat di pahami karena 90% responden dengan keterangan di pahami hal itu menandakan bahwa sebagian besar responden memahami fitur yang ada di *website*.

5. Tampilan web mudah di amati.

Tabel 4.5.5 Pengujian Beta Pertanyaan 5

Point Jawaban	Frekuensi Jawaban	Jumlah Sampel	Persentase (%)	Keterangan
Sangat Sulit	0	30	0	Tidak
Sulit	0		0	Tidak
Cukup Mudah	4		13.3	Ya
Mudah	16		53.3	Ya
Sangat Mudah	10		33.3	Ya

Berdasarkan tabel 4.5.5 hasil responden terhadap *website* ini yaitu mudah karena 100% responden dengan keterangan Ya menandakan bahwa *website* ini dapat di amati dengan mudah.

C. Aspek Interaksi

6. *website* ini dapat di akses dengan mudah.

Tabel 4.5.6 Pengujian Beta Pertanyaan 6

Point Jawaban	Frekuensi Jawaban	Jumlah Sampel	Persentase (%)	Keterangan
Sangat Sulit	0	30	0	Tidak
Sulit	0		0	Tidak
Cukup Mudah	3		10	Ya
Mudah	20		66.7	Ya
Sangat Mudah	7		23.3	Ya

Berdasarkan tabel 4.5.6 hasil responden terhadap *website* ini yaitu mudah karena 100% responden dengan keterangan Ya menandakan bahwa *website* ini dapat di amati dengan mudah.

7. *website* akan memudahkan pengguna untuk pengelompokan data.

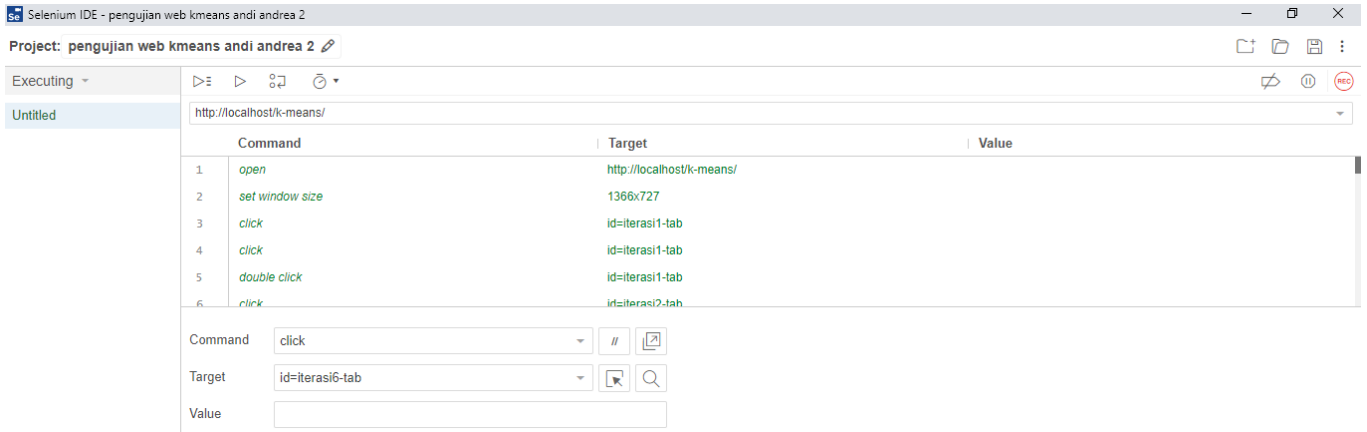
Tabel 4.5.7 Pengujian Beta Pertanyaan 7

Point Jawaban	Frekuensi Jawaban	Jumlah Sampel	Persentase (%)	Keterangan
Tidak	0	30	0	Tidak
Mungkin	3		10	Ya
Ya	27		90	Ya

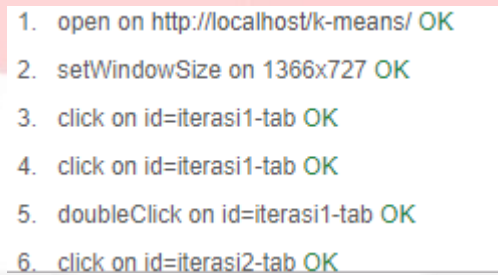
Berdasarkan tabel 4.5.7 hasil responden terhadap *website* ini yaitu memudahkan karena 100% responden dengan keterangan Ya menandakan bahwa *website* ini dapat membantu dan memudahkan pengguna untuk pengelompokan data.

5.2 Pengujian Website

Pada tahap pengujian fitur-fitur yang terdapat pada website menggunakan Selenium IDE , pengujian website meliputi fitur seperti *upload file* , *home* , *iterasi* , *search* , *previous* , *next* dan fitur *showing*.



Gambar 4.2.1 Tampilan pengujian website menggunakan Selenium IDE



Gambar 4.2.2 Tampilan hasil pengujian website

Gambar 4.2.2 merupakan hasil dari proses pengujian fitur yang terdapat pada website , semua fitur “OK” yang menunjukkan bahwa fitur yang di uji sudah berjalan dengan baik

No	Fitur	Hasil Pengujian Selenium IDE
1	Home	OK
2	Iterasi	OK
3	Upload File	OK
4	Search	OK
5	Previous	OK
6	Next	OK
7	Showing	OK
8	Diagram	OK

Tabel 4.2.1 Hasil Pengujian website menggunakan Selenium IDE

Berdasarkan hasil pengujian website menggunakan Selenium IDE semua fitur yang terdapat dalam website berjalan dengan baik.

5. Kesimpulan dan Saran

A Kesimpulan

1. Berdasarkan hasil yang di dapatkan maka di dapatkan hasil dari analisis tersebut bahwa kota/kabupaten di Indonesia di bagi menjadi 3 *cluster*.
2. Berdasarkan penelitian yang telah di lakukan dapat di simpulkan bahwa algoritma K-means dapat melakukan pengklasteran data tersebut.
3. Berdasarkan hasil penelitian yang di dapatkan bahwa program ini melakukan klasifikasi menjadi beberapa kelompok yaitu kelompok daerah tinggi, sedang, dan rendah . Hasil tersebut mencakup angka korelasi

yang dapat di gunakan untuk melakukan evaluasi dan pencegahan terhadap daerah yang rawan terhadap penyakit HIV AIDS di Indonesia.

B. Saran

1. Pengolahan data di gunakan lebih besar atau beberapa tahun kebelakang sehingga aplikasi dapat benar-benar berguna dalam pengklasteran data-data.
2. Mengembangkan GUI (*Graphical User Interface*) lebih menarik , mudah dan nyaman untuk user dalam menggunakan aplikasi ini



DAFTAR PUSTAKA

- [1] World Health Organization(WHO) dan Joint United Nations Programme on HIV/AIDS(UNAIDS), "Laporan perkembangan HIV & AIDS di dunia tahun 2007," Laporan Perkembangan HIV-AIDS Tahun 2007.
- [2] Kumar et al, " Penularan penyakit HIV/AIDS," Tahun 2007.
- [3] Kementerian Kesehatan RI, "Laporan Situasi Perkembangan HIV & AIDS di Indonesia Tahun 2013," Laporan Perkembangan HIV-AIDS Tahun 2009, Jakarta, 2009.
- [4] Dinas Kesehatan Nganjuk, " Pengertian HIV/AIDS," no 18 & 31 , Ngajuk , Jawa Timur , 2018.
- [5] Unandar B , " Definisi AIDS ," no 401 , 1999 .
- [6] C Long Barbara , " Penyebab AIDS ," 1996 .
- [7] Turban dkk , " Pengertian dari Data Mining ," 2005.
- [8] Dr. Daniel T. Larose , " an introduction to data mining ," Pengertian dan perbedaan data mining , 2005 .
- [9] Tan Et Al , " aim Clustering ," Perbedaan tujuan dari Clustering, 2006.
- [10] Eko Prasetyo , " Metode Clustering ," Metode clustering menurut kategori kekompakan ,Yogyakarta, 2014.
- [11] Xindong Wu & Vipin Kumar , "Algorithm in Data Mining ," Definisi Algoritma K-Means , 2009.
- [12] X. Wu and V. Kumar, eds., "The Top Ten Algorithms in Data Mining.Chapman and Hall," Data Mining, Algoritma K-Means, 2009.
- [13] S. Russell and P. Norvig, "Artificial Intelligence A Modern Approach. Upper Saddle River," Tujuan Algoritma, Vol 3, New Jersey, 2010.

