

IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING PADA KLASTERISASI KASUS HIV DI JAWA BARAT

Koko Kodratul Munawar, Ade Irma Purnamasari

Program Studi Manajemen Informatika D3

STMIK IKMI Cirebon, Jl. Perjuangan No. 10B, Cirebon, Indonesia

kodratulkoko2002@gmail.com

ABSTRAK

HIV atau *Human Immunodeficiency Virus* merupakan penyakit yang menyebabkan kerusakan sistem imun dengan menginfeksi dan menghancurkan sel *CD4*. Penyakit ini sangat berbahaya dan harus diwaspadai karena penyebaran penyakit ini sangat cepat. Penyakit HIV ini dapat menular melalui hubungan seksual dari orang yang terkena HIV, transfusi darah yang terkontaminasi HIV, dan dari perempuan yang terkena HIV pada anaknya selama masa kehamilan, persalinan, dan menyusui. Di Jawa Barat, angka kasus HIV sudah terbilang sangat tinggi. Maka dari itu, diperlukan analisis tentang penyebaran kasus HIV di Provinsi Jawa Barat. Pada analisis penyebaran penyakit HIV ini menggunakan data yang berasal dari Dinas Kesehatan Provinsi Jawa Barat yang diperoleh dari *website* Open Data Jabar. Proses analisis menggunakan metode algoritma *K-Means Clustering* untuk mengelompokkan penyebaran penyakit HIV di setiap kabupaten dan kota yang ada di Provinsi Jawa Barat ke dalam beberapa *cluster* yaitu berdasarkan kelompok umur dengan tingkatan penyebaran kasus HIV yang berkategori tinggi dan berkategori rendah. Hasil pengelompokkan ini dapat digunakan untuk memetakan penyebaran penyakit HIV sebagai upaya penanggulangan penyebaran yang lebih luas lagi.

Kata Kunci: Algoritma, *CD4*, *Clustering*, HIV, *K-Means*

1. PENDAHULUAN

Human Immunodeficiency Virus (HIV) merupakan virus yang dapat menyerang terhadap sistem imun sehingga sel darah putih akan mengalami kerusakan terutama bagian *limfosit* pada tubuh. Hal ini mengakibatkan turunnya sistem kekebalan tubuh penderita HIV. Beberapa orang bahkan mampu hidup dengan tidak tampak seperti seseorang yang terkena penyakit HIV dan beberapa kasus sampai tidak membutuhkan pengobatan [1]. Pada tahun 2019 sampai tahun 2021, angka kasus HIV/AIDS telah menjadi sorotan publik. Karena merujuk pada data Komisi Penanggulangan AIDS (KPA), jumlah kasus yang ada di Ibu Kota Provinsi Jawa Barat telah mencapai 12.358 kasus hingga tahun 2021 [2]. Sementara di Jawa Barat sendiri, angka kasus HIV juga termasuk tinggi. Data tersebut dipublikasikan melalui dokumen Provinsi Jawa Barat oleh BPS dalam angka yang berasal dari Dinas Kesehatan (Dinkes) Provinsi Jawa Barat.

Meningkatnya kasus HIV di Indonesia ini menjadi kasus yang mayoritas banyak terjadi di wilayah kota. HIV dapat menular melalui hubungan seksual dari orang yang terkena HIV, transfusi darah yang terkontaminasi HIV, dan dari perempuan yang terkena HIV pada anaknya selama masa kehamilan, persalinan, dan menyusui [3]. Sehingga peningkatan angka kasus HIV di Indonesia ini, membuat banyak instansi harus bertindak secara signifikan dan akurat.

Berdasarkan penelitian terdahulu yang dilakukan oleh Rut Indra Lita Sinaga, dkk., tentang “Pengelompokan Jumlah Kasus Penyakit Aids Berdasarkan Provinsi Menggunakan Metode *K-Means*” yang melakukan pengelompokkan penyakit HIV/AIDS berdasarkan provinsi pada tahun 2016-2018 menggunakan algoritma *K-Means*, di mana hasil pemodelan terbagi menjadi 2 *cluster*, *cluster* 1 merupakan kategori tinggi dan *cluster* 2 merupakan kategori rendah [4]. Penelitian ini merupakan salah satu referensi penulis untuk melakukan analisis masalah sehingga penulis dapat mengkaji teori yang diperlukan dalam analisis yang dilakukan.

Data yang digunakan dalam penelitian ini berasal dari Dinas Kesehatan Provinsi Jawa Barat yang diambil dari *website* Open Data Jabar. Data ini memiliki beberapa atribut yaitu id, kode provinsi, nama provinsi, kode kabupaten kota, nama kabupaten kota, kelompok umur, jenis kelamin, jumlah kasus, satuan dan tahun dimulai dari data tahun 2019 sampai data tahun 2021 mengenai data jumlah kasus HIV menurut umur di setiap kabupaten dan kota di Provinsi Jawa Barat. Metode dalam penelitian ini menggunakan algoritma *K-Means Clustering* untuk mengelompokkan penyebaran kasus HIV. Data yang telah diperoleh dikelompokkan ke dalam beberapa *cluster*, di mana dalam analisis ini untuk proses implementasi algoritma *K-Means Clustering* menggunakan *tools RapidMiner*.

Tabel 1. Data Jumlah Kasus HIV Berdasarkan Kelompok Umur di Jawa Barat

Id	Kode Provinsi	Nama Provinsi	Kode Kabupaten Kota	Nama Kabupaten Kota	Kelompok Umur	Jenis Kelamin	Jumlah Kasus	Satuan	Tahun
1	32	JAWA BARAT	3201	KABUPATEN BOGOR	0-4	LAKI-LAKI	4	ORANG	2019
2	32	JAWA BARAT	3201	KABUPATEN BOGOR	0-4	PEREMPUAN	7	ORANG	2019
3	32	JAWA BARAT	3201	KABUPATEN BOGOR	5-14	LAKI-LAKI	3	ORANG	2019
4	32	JAWA BARAT	3201	KABUPATEN BOGOR	5-14	PEREMPUAN	0	ORANG	2019
5	32	JAWA BARAT	3201	KABUPATEN BOGOR	15-19	LAKI-LAKI	8	ORANG	2019
6	32	JAWA BARAT	3201	KABUPATEN BOGOR	15-19	PEREMPUAN	5	ORANG	2019
7	32	JAWA BARAT	3201	KABUPATEN BOGOR	20-24	LAKI-LAKI	48	ORANG	2019
8	32	JAWA BARAT	3201	KABUPATEN BOGOR	20-24	PEREMPUAN	21	ORANG	2019
9	32	JAWA BARAT	3201	KABUPATEN BOGOR	25-49	LAKI-LAKI	217	ORANG	2019
10	32	JAWA BARAT	3201	KABUPATEN BOGOR	25-49	PEREMPUAN	133	ORANG	2019
11	32	JAWA BARAT	3201	KABUPATEN BOGOR	≥50	LAKI-LAKI	16	ORANG	2019
12	32	JAWA BARAT	3201	KABUPATEN BOGOR	≥50	PEREMPUAN	13	ORANG	2019
...	...	...	...	...	...	...	...	...	...
960	32	JAWA BARAT	3279	KOTA BANJAR	0-4	LAKI-LAKI	0	ORANG	2021
961	32	JAWA BARAT	3279	KOTA BANJAR	0-4	PEREMPUAN	1	ORANG	2021
962	32	JAWA BARAT	3279	KOTA BANJAR	5-14	LAKI-LAKI	1	ORANG	2021
963	32	JAWA BARAT	3279	KOTA BANJAR	5-14	PEREMPUAN	0	ORANG	2021
964	32	JAWA BARAT	3279	KOTA BANJAR	15-19	LAKI-LAKI	1	ORANG	2021
965	32	JAWA BARAT	3279	KOTA BANJAR	15-19	PEREMPUAN	0	ORANG	2021
966	32	JAWA BARAT	3279	KOTA BANJAR	20-24	LAKI-LAKI	3	ORANG	2021
967	32	JAWA BARAT	3279	KOTA BANJAR	20-24	PEREMPUAN	1	ORANG	2021
968	32	JAWA BARAT	3279	KOTA BANJAR	25-49	LAKI-LAKI	23	ORANG	2021
969	32	JAWA BARAT	3279	KOTA BANJAR	25-49	PEREMPUAN	7	ORANG	2021
970	32	JAWA BARAT	3279	KOTA BANJAR	≥50	LAKI-LAKI	2	ORANG	2021
971	32	JAWA BARAT	3279	KOTA BANJAR	≥50	PEREMPUAN	1	ORANG	2021

Setiap tahunnya, kasus HIV selalu meningkat, di kala terjadi penurunan pun tidak terlalu banyak angka kasus HIV yang berkurang. Dilansir dari [website detik.com](https://www.detik.com), dijelaskan hingga bulan Januari 2019, kasus HIV/AIDS secara kumulatif sebanyak 12.033 kasus. Dengan bertambah kasus baru HIV/AIDS pada 2019 sebanyak 1.077 kasus. Sedangkan pada tahun 2020, kasus HIV/AIDS secara kumulatif telah menurun menjadi 11.543 kasus. Akan tetapi, jumlah kasus baru HIV/AIDS di tahun 2020 meningkat

menjadi 1.370 kasus. Sementara pada tahun 2021, kasus HIV/AIDS secara kumulatif melonjak secara drastis sebanyak total 19.860 kasus, dengan kasus baru yang mencapai 2.216 kasus [2]. Semakin meningkatnya kasus HIV, dikhawatirkan semakin sulitnya upaya penanggulangan terhadap penyebaran kasus HIV ini, khususnya di Provinsi Jawa Barat. Sehingga pemerintah dan masyarakat harus sigap dalam mengupayakan pencegahan dan penanggulangan secara intensif dan akurat.

Penelitian ini bertujuan untuk melakukan pengelompokan tingkat penyebaran kasus HIV yang berkategori tinggi dan berkategori rendah berdasarkan kelompok umur di setiap kabupaten dan kota di Provinsi Jawa Barat. Metode dalam penelitian ini menggunakan algoritma *K-Means Clustering* dengan menentukan terlebih dahulu jumlah *cluster* yang akan dibentuk. Penelitian ini mengelompokkan wilayah dengan kasus HIV berkategori tinggi dan berkategori rendah untuk menjadi acuan dalam upaya penanggulangan penyebaran dan pencegahan secara luas dan merata di wilayah dengan kasus HIV berkategori tinggi.

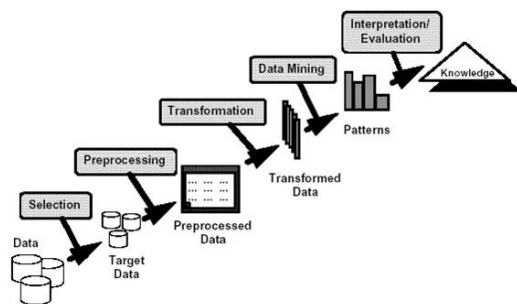
2. TINJAUAN PUSTAKA

Penelitian ini memerlukan landasan teori, guna menjelaskan definisi dan deskripsi dari metode yang digunakan. Landasan teori yang digunakan adalah *Data Mining*, *Clustering*, Algoritma *K-Means Clustering* dan *HIV*.

2.1. Data Mining

Data mining adalah proses penambangan atau penemuan informasi baru dengan melakukan pencarian suatu aturan atau pola tertentu dari sekumpulan data yang sangat besar. *Data mining* merupakan rangkaian proses untuk menggali nilai tambah berupa pengetahuan yang tidak diketahui selama ini secara manual dari sekumpulan data.

Proses *Data Mining* berdasarkan KDD atau *Knowledge Discovery In Database* memiliki tahapan sebagai berikut:



Gambar 1. Tahapan Proses *Data Mining* Berdasarkan *Knowledge Discovery In Database* (KDD) [5]
(Sumber: Dosbing, 2022)

- Seleksi Data (*Data Selection*)**
Pada database sering kali tidak semua data yang ada di dalamnya digunakan, maka yang akan diambil dari database tersebut hanya data yang telah sesuai untuk dianalisis.
- Pembersihan Data/Praproses Data (*Data Cleaning/Data Preprocessing*)**
Pembersihan data merupakan suatu proses yang dilakukan untuk menghilangkan *noise* serta data yang tidak relevan ataupun tidak konsisten.
- Transformasi data (*Data Transformation*)**

Mengubah atau menggabungkan data untuk diproses dalam data mining menjadi format yang sesuai.

- Data Mining / Proses Mining**

Merupakan proses utama saat metode digunakan untuk menemukan pengetahuan yang berharga dan tersembunyi dari data.

- Evaluasi pola (*Pattern Evaluation*)**

Untuk mengidentifikasi suatu pola yang menarik ke dalam *knowledge based* yang ditemukan [6].

2.2. Clustering

Clustering merupakan suatu proses untuk mengelompokkan titik data ke dalam dua kelompok atau lebih sehingga titik data di kelompok yang sama satu sama lain lebih mirip daripada di kelompok yang berbeda, hanya berdasarkan pada informasi yang tersedia dengan titik data [7].

Menurut Han dan Kamber (2006), *clustering* atau dikenal juga dengan istilah *cluster analysis* merupakan suatu teknik atau metode pada data mining yang tepat untuk digunakan dalam pengolahan data yang label ataupun kelasnya tidak diketahui. Cara kerja dari metode ini adalah melakukan pengelompokan kumpulan data menjadi kluster-kluster atau kelas-kelas, di mana jika dilakukan perbandingan dengan objek lain dari kelas tersebut, objek-objeknya memiliki similaritas yang tinggi. Namun jika dilakukan perbandingan dengan objek pada kelas/kluster lain, maka akan memiliki similaritas yang rendah [8].

2.3. Algoritma K-Means Clustering

K-means adalah suatu algoritma untuk mengelompokkan yang memisahkan data secara partisi menjadi berbeda-beda kelompok. Algoritma ini mampu meminimalisir jarak antara data ke klusternya. Pada dasarnya, digunakannya algoritma ini pada proses *clustering* tergantung pada konklusi yang ingin dicapai pada akhir proses dan data yang diperoleh. Penelitian ini menggunakan algoritma *K-Means* daripada algoritma *K-Medoids*, karena *K-Means* lebih mudah dan cepat dibanding *K-Medoids* yang prosesnya lama [4].

2.4. HIV (*Human Immunodeficiency Virus*)

Menurut Kementerian Kesehatan Republik Indonesia (2010), HIV adalah retrovirus yang menyerang sistem imun manusia sehingga mudah diserang oleh penyakit bawaan lain [9].

Menurut Rosella (2013), HIV adalah retrovirus (suatu golongan virus yang mengandung satu benang tunggal RNA (rangkain nukleotida seperti rantai yang saling terikat). Setelah sel terinfeksi, virus akan membentuk suatu replika DNA dari RNA-nya menggunakan *enzim reverse transcriptase*/transkripsi balik) yang menyebabkan infeksi sel-sel dari sistem imun, dan mengakibatkan kerusakan sel darah putih [6].

3. METODE PENELITIAN

3.1. Sumber Data

Data pada penelitian ini berasal dari Dinas Kesehatan Provinsi Jawa Barat yang diperoleh dari *website* Open Data Jabar melalui *link* [10]. Dinas Kesehatan Provinsi Jawa Barat terletak di Jl. Pasteur No. 25, Pasir Kaliki, Kec. Cicendo, Kota Bandung, Jawa Barat 40171. Data diambil pada hari Kamis, 19 Januari 2023. Data tersebut berisi Data Jumlah Kasus HIV Berdasarkan Kelompok Umur di Provinsi Jawa Barat.

Data pada penelitian ini termasuk ke dalam Data Sekunder yakni Data yang diperoleh melalui *website* dari suatu instansi atau perusahaan. Data diambil langsung dari *website* Open Data Jabar yang berasal dari Dinas Kesehatan Provinsi Jawa Barat.

3.2. Teknik Pengumpulan Data

3.2.1. Studi Pustaka

Teknik yang pertama pada penelitian ini diperoleh melalui studi pustaka. Dilansir dari *website* katadata, studi pustaka merupakan bagian dari suatu

karya ilmiah yang membahas penelitian terdahulu untuk digunakan sebagai referensi. Studi pustaka penting sebagai landasan teori yang relevan dengan tema penelitian [11]. Pada metode studi pustaka, referensi pada tugas akhir ini mengambil dari jurnal dan internet.

3.2.2. Studi Dokumentasi

Teknik yang kedua pada penelitian ini diperoleh melalui studi dokumentasi. Dokumen adalah rekaman kejadian di masa lalu yang dicetak atau ditulis, dapat berupa surat, catatan anekdot, buku harian serta dokumen-dokumen lain dalam bentuk audio, visual, maupun audio visual [12]. Dengan teknik studi dokumentasi ini, peneliti mendapatkan informasi dari sumber-sumber tertulis atau dari data dalam bentuk video maupun foto, bukan dari orang sebagai narasumber. Data yang digunakan pada pengelompokan tingkat penyebaran kasus HIV ini merupakan data sekunder yang diambil dari Open Data Jabar.

Tabel 2. Studi Dokumentasi Data Jumlah Kasus HIV Berdasarkan Kelompok Umur di Jawa Barat pada *website* Open Data Jabar

nama_provinsi	kode_kabupaten_kota	nama_kabupaten_kota	kelompok_umur	jenis_kelamin	jumlah_kasus	satuan	tahun
Filter	Filter	Filter	Filter	Filter	Cari	Filter	Filter
JAWA BARAT	3201	KABUPATEN BOGOR	0-4	LAKI-LAKI	4	ORANG	2019
JAWA BARAT	3201	KABUPATEN BOGOR	0-4	PEREMPUAN	7	ORANG	2019
JAWA BARAT	3201	KABUPATEN BOGOR	5-14	LAKI-LAKI	3	ORANG	2019
JAWA BARAT	3201	KABUPATEN BOGOR	5-14	PEREMPUAN	0	ORANG	2019
JAWA BARAT	3201	KABUPATEN BOGOR	15-19	LAKI-LAKI	8	ORANG	2019
JAWA BARAT	3201	KABUPATEN BOGOR	15-19	PEREMPUAN	5	ORANG	2019
JAWA BARAT	3201	KABUPATEN BOGOR	20-24	LAKI-LAKI	48	ORANG	2019
JAWA BARAT	3201	KABUPATEN BOGOR	20-24	PEREMPUAN	21	ORANG	2019
JAWA BARAT	3201	KABUPATEN BOGOR	25-49	LAKI-LAKI	217	ORANG	2019
JAWA BARAT	3201	KABUPATEN BOGOR	25-49	PEREMPUAN	133	ORANG	2019
JAWA BARAT	3201	KABUPATEN BOGOR	≥50	LAKI-LAKI	16	ORANG	2019
JAWA BARAT	3201	KABUPATEN BOGOR	≥50	PEREMPUAN	13	ORANG	2019
JAWA BARAT	3202	KABUPATEN SUKABUMI	0-4	LAKI-LAKI	2	ORANG	2019
JAWA BARAT	3202	KABUPATEN SUKABUMI	0-4	PEREMPUAN	0	ORANG	2019
JAWA BARAT	3202	KABUPATEN SUKABUMI	5-14	LAKI-LAKI	1	ORANG	2019
JAWA BARAT	3202	KABUPATEN SUKABUMI	5-14	PEREMPUAN	0	ORANG	2019
JAWA BARAT	3202	KABUPATEN SUKABUMI	15-19	LAKI-LAKI	4	ORANG	2019
JAWA BARAT	3202	KABUPATEN SUKABUMI	15-19	PEREMPUAN	2	ORANG	2019
JAWA BARAT	3202	KABUPATEN SUKABUMI	20-24	LAKI-LAKI	20	ORANG	2019
JAWA BARAT	3202	KABUPATEN SUKABUMI	20-24	PEREMPUAN	11	ORANG	2019
JAWA BARAT	3202	KABUPATEN SUKABUMI	25-49	LAKI-LAKI	50	ORANG	2019
JAWA BARAT	3202	KABUPATEN SUKABUMI	25-49	PEREMPUAN	19	ORANG	2019
JAWA BARAT	3202	KABUPATEN SUKABUMI	≥50	LAKI-LAKI	2	ORANG	2019
JAWA BARAT	3202	KABUPATEN SUKABUMI	≥50	PEREMPUAN	1	ORANG	2019
JAWA BARAT	3203	KABUPATEN CIANJUR	0-4	LAKI-LAKI	0	ORANG	2019

(Sumber: Open Data Jabar)

Data ini berisi data jumlah kasus HIV berdasarkan kelompok umur di Jawa Barat. Data tersebut merupakan data tahun 2019-2021 dengan sumber data dari Dinas Kesehatan Provinsi Jawa Barat. Data yang telah diperoleh diolah menggunakan tools *RapidMiner*.

3.3. Tahapan Perancangan

Tahapan perancangan pada penelitian ini menggunakan KDD atau *Knowledge Discovery In Databases*. Proses KDD memiliki tahapan sebagai berikut:

3.3.1. Seleksi Data (Data Selection)

Data diambil dari website Open Data Jabar yang bersumber dari Dinas Kesehatan Provinsi Jawa Barat mengenai Data Jumlah Kasus HIV Berdasarkan Kelompok Umur di Jawa Barat yang mencakup data dari tahun 2019-2021.

3.3.2. Pembersihan Data/Praproses Data (Data Cleaning/Data Preprocessing)

Pembersihan data/praproses data adalah tahapan untuk menyeleksi data dengan tujuan untuk memperoleh data yang bersih serta siap digunakan pada tugas akhir. Tahapan-tahapan yang dilakukan untuk mengetahui jumlah kasus HIV menurut umur di Provinsi Jawa Barat adalah melakukan pemeriksaan Data HIV yang tidak konsisten, melakukan pembersihan Data HIV, serta evaluasi terhadap kesalahan Data HIV.

3.3.3. Transformasi Data (Data Transformation)

Dataset yang akan diolah dilakukan transformasi terlebih dahulu yakni dengan melakukan perubahan data pada variabel kelompok umur yang awalnya berbentuk teks diubah menjadi numerik agar dapat dilakukan proses *data mining* pada aplikasi *RapidMiner*.

3.3.4. Data Mining / Proses Mining

Penggunaan teknik data mining untuk melakukan pengelompokan Data Kasus HIV dengan pemilihan algoritma *K-Means Clustering*. Jumlah data yang akan diolah sebanyak 971 data. Proses pengolahan data dan pemodelan dengan algoritma *K-Means Clustering* menggunakan aplikasi *RapidMiner*.

3.3.5. Evaluasi pola (Pattern Evaluation)

Setelah diperoleh hasil dari proses *data mining*, selanjutnya tahap evaluasi pola dengan melihat kualitas dari cluster yang telah dibentuk. Pada tahapan ini memeriksa pola atau informasi yang ditemukan, apakah bertentangan atau tidak dengan fakta yang ada sebelumnya.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Perancangan

Berikut ini adalah hasil berdasarkan tahapan perancangan menggunakan KDD atau *Knowledge Discovery In Databases*.

4.1.1. Seleksi Data (Data Selection)

Proses penyelesaian pengelompokan penyebaran kasus HIV yang pertama adalah seleksi data yakni melakukan pemilihan data yang berasal dari Dinas Kesehatan Provinsi Jawa Barat yang diperoleh dari website Open Data Jabar. Dataset terdiri dari 971 record dengan 10 atribut yakni id, kode provinsi, nama provinsi, kode kabupaten kota, nama kabupaten kota, kelompok umur, jenis kelamin, jumlah kasus, satuan dan tahun.

4.1.2. Pembersihan Data/Praproses data (Data Cleaning/Data Preprocessing)

Melakukan pengecekan dan pembersihan data missing value dan data noise dari Data Kasus HIV melalui tools *RapidMiner*.

Tabel 3. Pengecekan Missing Value pada Data Kasus HIV di Jawa Barat

Name	Type	Missing	Statistics	Filter (5 attributes)
nama_kabupaten_kota	Nominal	0	Count: KOTA TASIKMALAYA (3), KABUPATEN BANDUNG, KABUPATEN BANDUNG (36), KABUPATEN	
kelompok_umur	Nominal	0	Count: 350 (161), 0-4 (162), 5-14 (162), 15-19 (162), 20-24 (162), 25-49 (162), ≥50 (162)	
jenis_kelamin	Nominal	0	Count: PEREMPUAN (485), LAKI-LAKI (486), LAKI-LAKI (486), PEREMPUAN (485)	
jumlah_kasus	Integer	0	Min: 0, Max: 254, Average: 13.958	
tahun	Integer	0	Min: 2019, Max: 2021, Average: 2019.999	

Data Kasus HIV yang diolah tidak memiliki missing value yang ditandai dengan angka 0 di bagian Missing.

4.1.3. Transformasi Data (Data Transformation)

Melakukan transformasi data atau mengubah data pada dataset awal Data Kasus HIV menurut kabupaten/kota serta menurut umur agar dapat diolah menggunakan aplikasi *RapidMiner*. Selanjutnya dilakukan inisialisasi atribut kelompok umur dan jenis kelamin pada data tersebut.

Tabel 4. Inisialisasi Atribut Kelompok Umur

Kelompok Umur	Inisialisasi
0-4 Tahun	0
5-14 Tahun	1
15-19 Tahun	2
20-24 Tahun	3
25-49 Tahun	4
≥50 Tahun	5

Tabel 5. Inisialisasi Atribut Jenis Kelamin

Jenis Kelamin	Inisialisasi
Laki-laki	0
Perempuan	1

4.1.4. Data Mining / Proses Mining

Melakukan pengelompokan Data Kasus HIV berdasarkan kelompok umur pada aplikasi RapidMiner menggunakan Algoritma *K-Means Clustering*.



Gambar 2. Data Mining Menggunakan Algoritma *K-Means Clustering*

Proses *data mining* menggunakan algoritma *K-Means Clustering* pada aplikasi *RapidMiner*.

4.1.5. Evaluasi pola (Pattern Evaluation)

Tahap terakhir setelah proses *data mining* pada aplikasi *RapidMiner* selanjutnya adalah menekan tombol *run*, maka akan muncul sebuah tab *result* berisi pengelompokan berdasarkan kabupaten/kota dan berdasarkan kelompok umur yakni $k=2$.

Tabel 6. Hasil Evaluasi *Cluster* Dataset Kasus HIV Berdasarkan Kelompok Umur

Row No.	nama_kabupaten_kota	cluster	kelompok_u...	jenis_kelamin	jumlah_kasus	tahun
1	KABUPATEN BOGOR	cluster_0	0	0	4	2019
2	KABUPATEN BOGOR	cluster_0	0	1	7	2019
3	KABUPATEN BOGOR	cluster_0	1	0	3	2019
4	KABUPATEN BOGOR	cluster_0	1	1	0	2019
5	KABUPATEN BOGOR	cluster_0	2	0	8	2019
6	KABUPATEN BOGOR	cluster_0	2	1	5	2019
7	KABUPATEN BOGOR	cluster_0	3	0	48	2019
8	KABUPATEN BOGOR	cluster_0	3	1	21	2019
9	KABUPATEN BOGOR	cluster_1	4	0	217	2019
10	KABUPATEN BOGOR	cluster_1	4	1	133	2019
11	KABUPATEN BOGOR	cluster_0	5	0	16	2019
12	KABUPATEN BOGOR	cluster_0	5	1	13	2019
13	KABUPATEN SUKABUMI	cluster_0	0	0	2	2019
14	KABUPATEN SUKABUMI	cluster_0	0	1	0	2019

ExampleSet (971 examples, 2 special attributes, 4 regular attributes)

4.2. Hasil Pengelompokan

4.2.1. Pengelompokan Penyebaran Kasus HIV Berdasarkan Kelompok Umur

Berikut ini merupakan hasil dari implementasi algoritma *K-Means Clustering* pada pengelompokan penyebaran kasus HIV berdasarkan umur di Jawa Barat. Setelah proses *running*, diperoleh DBI sebagai berikut :

Tabel 7. Hasil Cluster DBI Kasus HIV Berdasarkan Kelompok Umur

Cluster/k	DBI
2	0.093
3	0.121
4	0.096
5	0.109
6	0.123
7	0.125
8	0.126
9	0.146
10	0.141

Berdasarkan tabel 7, diperoleh *cluster* atau k terbaik yaitu *cluster* atau $k=2$ dengan DBI 0.093.

PerformanceVector

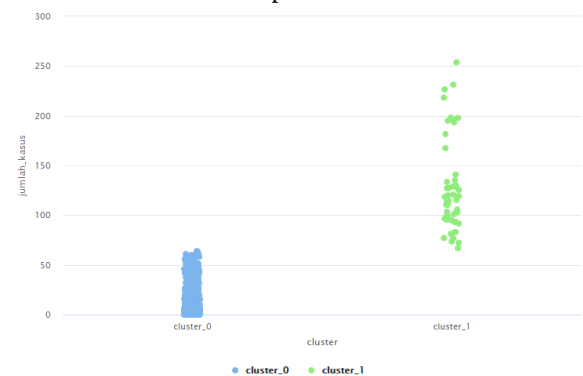
PerformanceVector:

Avg. within centroid distance: 70.569
 Avg. within centroid distance_cluster_0: 44.655
 Avg. within centroid distance_cluster_1: 519.413
 Davies Bouldin: 0.093

Gambar 3. Hasil *PerformanceVector* Berdasarkan Kelompok Umur

Dari gambar 3, dapat dilihat hasil *PerformanceVector* dari pengelompokan penyebaran kasus HIV berdasarkan kelompok umur diperoleh DBI 0.093.

Selanjutnya hasil *cluster* dapat dilihat pada *Visualizations* dalam *RapidMiner*.



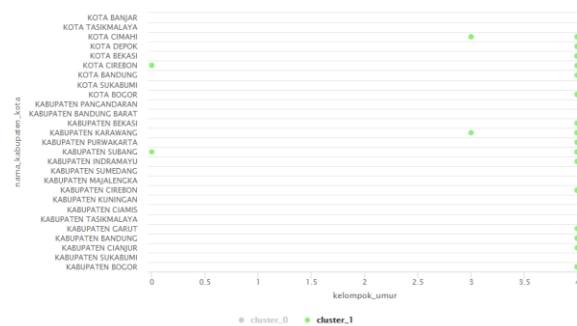
Gambar 4. Hasil *Visualizations* dari Penyebaran Kasus HIV Kategori Tinggi dan Kategori Rendah

Berdasarkan gambar 4, dapat dilihat penyebaran kasus HIV berdasarkan kelompok umur di Jawa Barat yang terbagi ke dalam 2 *cluster*. *Cluster_0* yang berwarna biru adalah *cluster* dengan kategori rendah dengan angka kasus yang berkisar antara 0-65 kasus, sedangkan *cluster_1* yang berwarna hijau adalah *cluster* dengan kategori tinggi dengan angka kasus yang berkisar antara 68-254 kasus. Penjelasan yang secara rinci dapat dilihat pada gambar 5.



Gambar 5. Hasil *Visualizations* dari Penyebaran Kasus HIV Kategori Rendah Berdasarkan Kelompok Umur

Berdasarkan gambar 5, kasus HIV berkategori rendah pada kelompok umur 0-4 tahun, kelompok umur 5-14 tahun, kelompok umur 15-19 tahun, kelompok umur 20-24 tahun, dan kelompok umur ≥ 50 tahun terdapat di setiap wilayah yang berjumlah 27 kabupaten/kota, sedangkan pada kelompok umur 25-49 tahun terdapat di 26 kabupaten/kota, karena 1 wilayah yang lain termasuk ke dalam kategori tinggi. Untuk kasus HIV berkategori tinggi akan dijelaskan pada gambar 6.



Gambar 6. Hasil Visualizations dari Penyebaran Kasus HIV Kategori Tinggi Berdasarkan Kelompok Umur

Berdasarkan gambar 6, kasus HIV berkategori tinggi pada kelompok umur 0-4 tahun terdapat di 2 wilayah antara lain Kabupaten Subang dan Kota Cirebon. Pada kelompok umur 5-14 tahun, kelompok umur 15-19 tahun, dan kelompok umur ≥ 50 tahun tidak ada kasus HIV yang termasuk kategori tinggi. Pada kelompok umur 20-24 tahun kasus HIV berkategori tinggi terdapat di 2 wilayah antara lain Kabupaten Karawang dan Kota Cimahi. Pada kelompok umur 25-49 tahun terdapat di 16 wilayah antara lain Kabupaten Bogor, Kabupaten Cianjur, Kabupaten Bandung, Kabupaten Garut, Kabupaten Cirebon, Kabupaten Indramayu, Kabupaten Subang, Kabupaten Purwakarta, Kabupaten Karawang, Kabupaten Bekasi, Kota Bogor, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, dan Kota Cimahi.

Maka, dapat disimpulkan bahwa kasus HIV berkategori tinggi lebih sedikit dibanding dengan kasus HIV berkategori rendah. Oleh karena itu, pemerintah ataupun instansi kesehatan dapat lebih mudah dalam upaya penanggulangan terhadap penyebaran HIV di wilayah dengan kasus HIV berkategori tinggi.

5. KESIMPULAN DAN SARAN

Berdasarkan analisis di atas, kesimpulan yang diperoleh sebagai berikut: Hasil pengelompokan tingkat penyebaran HIV berdasarkan kelompok umur menggunakan algoritma *K-Means Clustering* diperoleh 2 cluster. *Cluster_0* merupakan kategori rendah, sedangkan *cluster_1* merupakan kategori tinggi yang terdapat pada kelompok umur 0-4 tahun di Kabupaten Subang dan Kota Cirebon, kelompok umur 20-24 tahun di Kabupaten Karawang dan Kota

Cimahi, dan kelompok umur 25-49 tahun di Kabupaten Bogor, Kabupaten Cianjur, Kabupaten Bandung, Kabupaten Garut, Kabupaten Cirebon, Kabupaten Indramayu, Kabupaten Subang, Kabupaten Purwakarta, Kabupaten Karawang, Kabupaten Bekasi, Kota Bogor, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, dan Kota Cimahi.

Adapun saran dari penelitian ini yakni sebagai berikut: Untuk penelitian selanjutnya dapat menggunakan data yang lebih banyak lagi agar memperoleh akurasi yang lebih baik dan hasil yang lebih luas lagi. Pada penelitian selanjutnya, dapat dilakukan menggunakan algoritma lain untuk melakukan pengelompokan data agar dapat dilakukan perbandingan hasil yang lebih akurat.

DAFTAR PUSTAKA

- [1] A. W. P. Samudra, R. A. Susanto, A. R. Putra, F. I. Kurniadi, dan B. Juarto, "Klasifikasi HIV AIDS dengan Aplikasi Rapid Miner," *Sist. Komput. dan Kecerdasan Buatan*, no. 1, hal. 15–19, 2022.
- [2] R. Alhamidi, "Membaca Grafik Kasus HIV/AIDS di Jabar: Naik 3 Tahun Terakhir!," *detikjabar*, 2022. <https://www.detik.com/jabar/berita/d-6262009/membaca-grafik-kasus-hivaidis-di-jabar-naik-3-tahun-terakhir> (diakses 18 Januari 2023).
- [3] A. A. Lesmana *et al.*, "IMPLEMENTASI ALGORITMA K-MEANS UNTUK CLUSTERING PENYAKIT HIV / AIDS DI INDONESIA IMPLEMENTATION OF K-MEANS ALGORITHM FOR CLUSTERING OF HIV / AIDS DISEASE IN INDONESIA," *e-Proceeding Eng.*, vol. 6, no. 2, hal. 5564–5580, 2019.
- [4] R. I. L. Sinaga, W. Saputra, dan H. Qurniawan, "Pengelompokan Jumlah Kasus Penyakit Aids Berdasarkan Provinsi Menggunakan Metode K-Means," *KESATRIA Jurnal Penerapan Sist. Inf. (Komputer Manajemen)*, vol. 2, no. 2, hal. 99–107, 2021.
- [5] dosbing.id, "Proses tahapan data mining," *dosbing.id*, 2023. <https://dosbing.id/2023/04/21/proses-tahapan-data-mining/> (diakses 4 Februari 2023).
- [6] A. Y. E. Dodu, D. W. Nugraha, dan M. A. Ayyub, "Penerapan Data Mining Untuk Mendeteksi Tingkatan Stadium Penyakit Human Immunodeficiency Virus / Acquired Immune Deficiency Syndrome (Hiv/Aids) Menggunakan Algoritma Naive Bayes Classifier (Studi Kasus Pada Rumah Sakit Umum Daerah (Rsud) Anutapura Palu)," *Sci. Comput. Sci. Informatics J.*, vol. 1, no. 1, hal. 33, 2019, doi: 10.22487/j26204118.2018.v1.i1.11900.
- [7] V. Herlinda dan D. Darwis, "Analisis Clustering Untuk Recredesialing Fasilitas Kesehatan Menggunakan Metode Fuzzy C-Means,"

- Teknol. dan Sist. Inf.*, vol. 2, no. 2, hal. 94–99, 2021, [Daring]. Tersedia pada: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [8] H. Noor, A. Dharmawati, dan T. W. Qur'ana, "Penerapan Algoritma K-Means Clustering Analysis Pada Kasus Penderita Hiv/Aids (Studi Kasus Kabupaten Banjar)," *Technol. J. Ilm.*, vol. 12, no. 2, hal. 72, 2021, doi: 10.31602/tji.v12i2.4573.
- [9] D. Y. Nugroho, A. Wibowo, F. Kesehatan, dan M. Universitas, "ANALISIS CLUSTER K - MEANS KABUPATEN / KOTA DI JAWA TIMUR BERDASARKAN LEVEL EPIDEMIK HIV CLUSTER K - MEANS ANALYSIS OF THE HIV EPIDEMIC LEVEL IN," *Biometrika dan Kependud.*, vol. 8, no. 2, hal. 108–117, 2019, doi: 10.20473/jbk.v8i2.2019.108.
- [10] opendata.jabarprov.go.id, "Jumlah Kasus HIV Berdasarkan Kelompok Umur di Jawa Barat," opendata.jabarprov.go.id, 2021. <https://opendata.jabarprov.go.id/id/dataset/jumlah-kasus-hiv-berdasarkan-kelompok-umur-di-jawa-barat> (diakses 18 Januari 2023).
- [11] D. Latifatul Fajri, "Studi Pustaka Adalah Referensi Penelitian, Ini Penjelasan Lengkapnya," *katadata*, 2022. <https://katadata.co.id/agung/berita/62e773e3da762/studi-pustaka-adalah-referensi-penelitian-ini-penjelasan-lengkapnya> (diakses 27 Januari 2023).
- [12] E. R. Wardhani, "MANAJEMEN PERUBAHAN DALAM IMPLEMENTASI REFORMASI BIROKRASI INTERNAL," 2018. [Daring]. Tersedia pada: http://repository.upi.edu/44420/7/T_ADPEN_1608369_Chapter3.pdf