KOMPARASI ALGORITMA *K-MEANS* DAN *K-MEDOIDS CLUSTERING* PADA DATA PENYEBARAN KASUS HIV DI PROVINSI JAWA BARAT

Moh Soni, Nining Rahaningsih, Raditya Danar Dana
Teknik Informatika, Sekolah Tinggi Manajemen dan Informatika IKMI Cirebon
Jalan Perjuangan No. 10B Karyamulya Kec. Kesambi Kota Cirebon Jawa Barat 45135, Indonesia
mohsoni2001@gmail.com

ABSTRAK

Dinas Kesehatan Jawa Barat mencatat kasus HIV pada tahun 2022 di Jawa Barat terus mengalami peningkatan. Situasi ini mencerminkan penularan HIV di Jawa Barat masih berlangsung di masyarakat hingga saat ini dan memerlukan upaya pencegahan. Berdasarkan permasalahan tersebut, penelitian ini melakukan pengelompokkan dengan metode perbandingan algoritma *K-Means* dan *K-Medoids* untuk melihat algoritma mana yang paling optimal dari segi performa nilai *Davies-Boudien Index* (DBI) dan waktu komputasi. Analisis *cluster* dilakukan berdasarkan wilayah, jenis kelamin dan umur. Proses penelitian menggunakan metode KDD *(Knowledge Discovery in Databases)* mencakup tahap *Data selection, Pre-processing, Transformation, Data mining*, dan *Evalution/ interpretation*. Hasil penelitian menunjukkan algoritma *K-Means* sebagai metode yang paling baik dan optimal dalam melakukan pengelompokkan dibandingkan dengan algoritma *K-Medoids*. Hasil pengujian optimal pada algoritma *K-Means* didapat pada K=4 dengan nilai validitas DBI 0,102, berbeda dengan algoritma *K-Medoids* yang menunjukkan klaster optimal pada K=3 dengan nilai DBI 0,130. Selain itu, dalam hal perbandingan kinerja kecepatan didapatkan bahwa *K-Means* memiliki waktu komputasi lebih cepat dibandingkan dengan algoritma *K-Medoids* yaitu 5 detik pada algoritma *K-Means* dan 5 menit pada *K-Medoids*. Hasil penelitian diharapkan dapat membantu pemerintah dalam perencanaan dan pengembangan strategi pencegahan HIV/AIDS di Jawa Barat serta dapat digunakan sebagai panduan untuk menentukan prioritas dalam penanganan penyebaran HIV.

Kata kunci: *Data Mining, K-Means, K-Medoids, Clustering*, HIV

1. PENDAHULUAN

Angka kasus HIV di Indonesia masih menjadi permasalahan kesehatan yang signifikan, data Kementrian Kesehatan Republik Indonesia menunjukkan bahwa jumlah estimasi pengidap HIV di Indonesia tahun 2020 mencapai sebanyak 543.100 orang. Selain itu, data tersebut mencatat 29.557 diantaranya sebagai infeksi baru, hal ini menunjukkan bahwa penularan HIV masih berlangsung di masyarakat hingga saat ini. Di samping itu, angka kematian sebanyak 30.137 orang yang terkait dengan HIV pada tahun 2020 menjadi perhatian serius dalam menjaga kualitas hidup dan kesejahteraan individu yang terkena dampak penyakit ini.[1] Kasus HIV di Jawa Barat sendiri menunjukkan peningkatan yang cukup tinggi, sebagaimana menurut data Dinas Kesehatan Jawa Barat pada tahun 2022, jumlah HIV positif meningkat dibandingkan tahun 2021 mencatat sejumlah 5.444 kasus, sebagian besar kasus terjadi pada kelompok usia 25 sampai 49 tahun dengan angka 69,7%. Dari sisi jenis *gender*, kasus HIV positif terbagi menjadi 74% laki-laki dan perempuan di angka 26%.[2]

Peningkatan kasus penyebaran HIV di Jawa Barat menjadi sebuah perhatian serius yang memerlukan tindakan pencegahan. Pemahaman yang mendalam tentang pola penyebaran kasus HIV sangat penting untuk merancang strategi intervensi yang efektif. Dalam hal ini, analisis data mining dengan metode *clustering* dapat menjadi salah satu pencegahan yang efektif untuk menekan angka kasus HIV/AIDS di Jawa Barat. Pengelompokkan data menjadi langkah kritis untuk mengidentifikasi pola dan tren yang mungkin tidak terlihat secara langsung. *Clustering* merupakan metode pengelompokkan data berdasarkan kondisi tertentu. Prinsip dasar *clustering* adalah menggabungkan beberapa objek ke dalam *cluster*, di mana kualitas *cluster* diukur berdasarkan tingginya kemiripan di antara objek dalam *cluster* yang sama dan tingginya perbedaan antara *cluster* yang lain.[3] Adapun algoritma yang dapat digunakan untuk melakukan pengelompokan diantaranya yaitu *K-Means* dan *K-Medoids*. Algoritma *K-Means* membagi data ke dalam kelompok atau *cluster*, dengan setiap *cluster* terdiri dari data yang mempunyai ciri serupa, namun berbeda dengan data yang ada di *cluster* lain.[4] Algoritma *K-Medoids* masih serupa dengan *K-Means*, karena keduanya termasuk dalam jenis algoritma partitional yang membagi data menjadi beberapa kelompok. *K-Medoids* merupakan metode yang berfokus pada pengurangan jarak antara titik yang ada pada suatu *cluster* dan titik yang dijadikan pusat *cluster*.[5] Berdasarkan tingkat keefektifan dalam melakukan pengelompokkan data, penelitian ini menerapkan metode algoritma *K-Means* dan *K-Medoids*.

Penelitian ini dilakukan untuk mengidentifikasi *cluster* atau kelompok penyebaran penyakit HIV/AIDS di Provinsi Jawa Barat sebagai upaya untuk membantu pemerintah dalam mengurangi angka kasus HIV/AIDS. Penelitian ini akan membandingkan algoritma *K-Means* dan *K-Medoids* dalam mencari

keakuratan *cluster* sebagai upaya untuk membantu pemerintah dalam memberikan wawasan tentang mana dari kedua algoritma yang lebih cocok untuk pengelompokkan data penyebaran kasus HIV di Jawa Barat, dan diharapkan dapat bermanfaat dalam perencanaan dan pengembangan strategi pencegahan serta mengurangi angka kasus HIV/AIDS di Jawa Barat.. Dalam prosesnya, eksperimen ini menggunakan metode KDD (*Knowledge Discovery in Databases*) sebagai landasan kerangka kerja. Proses tersebut mencakup tahap *Data Selection*, *Pre-processing*, *Transformation*, *Data Mining*, dan *Evaluation/Interpretation*.

Hasil penelitian ini nantinya akan menentukan pengelompokkan mana yang lebih optimal antara kedua algoritma pada data Kasus Penyebaran HIV di Provinsi Jawa Barat. Perolehan *cluster* yang dominan akan menentukan jumlah daerah paling banyak berdasarkan jenis kelamin dan umur dalam hal kasus penyebaran HIV. Penentuan hasil *cluster* akan disesuaikan dengan hasil akhir evaluasi DBI (*Davies-Bouldin Index*) guna menentukan jumlah klaster terbaik. Penelitian ini dilakukan dengan bantuan aplikasi RapidMiner versi 10.3.

2. TINJAUAN PUSTAKA

Tinjauan pustaka (*literature review*) dilakukan dengan cara meneliti artikel-artikel terdahulu berkaitan dengan topik penelitian perbandingan algoritma *K-Means* dan *K-Medoids*. Proses pencarian sumber literature dilakukan dengan menggunakan alat bantu "*publish or perish*" untuk mengakses database akademik *Google Scholar*. Dari hasil pencarian, terpilih 10 artikel jurnal tahun 2019-2023 yang benar-benar relevan dengan topik sebagai berikut:

Paper [6] melakukan pengelompokan daerah produksi kakao guna mengidentifikasi Kabupaten/Kota di Provinsi Sulawesi Selatan yang mencapai hasil produksi kurang optimal, cukup baik, berlimpah, dan sangat berlimpah dengan algoritma *K-Means* dan *K-Medoids*. Hasil penelitian menunjukkan algoritma *K-Means* terbukti lebih efektif dibandingkan *K-Medoids* dengan *Davies-Bouldin Index* (DBI) 0,292, sementara untuk *K-Medoids* adalah 0,365.

Paper [3] mengelompokkan distribusi obat di Puskesmas Karangsambung ke dalam tiga kategori, yaitu lambat, sedang, dan cepat. Hasil penelitian menunjukkan algoritma *K-Means* memiliki *Silhouette Coefficient* mencapai 0,627, sementara *K-Medoids* mencapai 0,536. Hal ini menandakan bahwa hasil pengelompokkan yang diperoleh dari algoritma *K-Means* lebih unggul dibandingkan dengan *K-Medoids*.

Paper [7] berfokus pada pengelompokan data untuk menentukan metode pengolahan data terbaik antara *K-Means* dan *K-Medoids* terkait virus korona (Covid-19) di Indonesia. Nilai *indeks Davies-Bouldin* dijadikan pedoman untuk melakukan pengelompokan data. Berdasarkan jumlah klaster yaitu K2 hingga K9, metode *K-Means* merupakan metode terbaik untuk mengelompokkan penyebaran virus korona dengan nilai DBI terkecil pada K-5 sebesar 0,064, sedangkan *K-Medoids* pada nilai K-2 sebesar 0,411.

Paper [8] mencari keakuratan *cluster* algoritma *K-Means* dan *K-Medoids* dalam klasifikasi skripsi mahasiswa berdasarkan nilai *Davies Bouldin Index*. Hasil pengujian menunjukkan *K-Medoids* sebagai metode yang unggul dibanding *K-Means*. Nilai DBI yang dihasilkan oleh *K-Medoids* adalah 1,56, sedangkan *K-Means* mencapai 2,79. Evaluasi waktu komputasi dalam pengelompokkan 50 dokumen, *K-Means* menunjukkan keunggulan dengan rata-rata waktu komputasi sekitar 1 detik, sedangkan *K-Medoids* memerlukan rata-rata waktu komputasi sekitar 26,7778 detik.

Paper [9] mengoptimalkan dan meningkatkan produksi buah dengan mengelompokkan jenis produksi buah-buahan di Kabupaten Kotawaringin Timur menjadi klaster tinggi, klaster sedang, dan klaster terendah. Hasil pengujian DBI menunjukkan bahwa pada percobaan keempat, *K-Means* menghasilkan nilai DBI sebesar 0,296 dengan enam buah klaster sedangkan algoritma *K-Medoids* yang memiliki nilai DBI sebesar 0,507, penelitian menyimpulkan algoritma terbaik dalam penelitian ini adalah algoritma *K-Means*.

Paper [10] dibuat untuk memudahkan peserta baru yang akan mendaftar program BPJS Ketenagakerjaan dengan memanfaatkan metode *clustering*, khususnya membandingkan dua algoritma, yaitu *K-Means* dan *K-Medoids*. Penelitian menunjukkan *K-Medoids* menghasilkan kelompok yang lebih stabil dan optimal dibandingkan dengan *K-Means*. Dengan nilai DBI pada algortima *K-Medoids* lebih kecil dibandingkan dengan *K-Means* yaitu -51,9.

Paper [11] melakukan pengelompokan tingkat kelembapan udara, suhu, dan tegangan listrik (*humidity, temperature,* d*an voltage)* menggunakan algoritma *K-Means* dan *K-Medoids*. Kesimpulan penelitian menunjukkan *K-Means* lebih efektif dalam mengelompokkan data. Keunggulan *K-Means* terlihat dengan nilai *Davies-Bouldin Index* (DBI) sebesar 0,306 pada percobaan K=2, dan waktu proses yang relatif singkat, yaitu hanya 1 menit 22 detik.

Paper [12] mengelompokkan data penduduk miskin dengan metode perbandingan antara dua algoritma, yaitu *K-Means* dan *K-Medoids*. Hasil penelitian terlihat bahwa algoritma *K-Means* menunjukkan keunggulan dibandingkan dengan *K-Medoids*. Fakta ini didukung oleh nilai DBI terbaik dari *K-Means* sebesar 0.041 pada K=8.

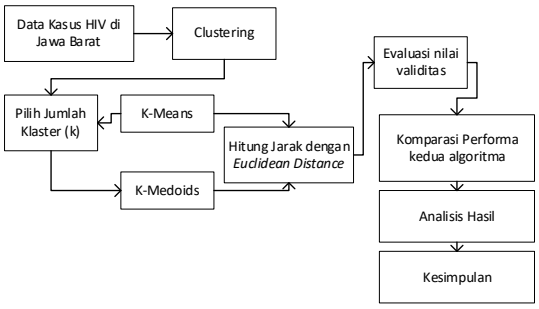
Paper [13] mengelompokkan data armada kendaraan truk berdasarkan produktivitas, dilakukan perbandingan antara algoritma *K-Means* dan *K-Medoids*. *Davies Bouldin Index* digunakan sebagai metode analisis klaster, dengan hasil nilai validitas sebesar 0,67 untuk *K-Means clustering* dan 1,78 untuk *K-Medoids*. Berdasarkan hasil tersebut, dibuat aplikasi *clustering* armada kendaraan berbasis web dengan algoritma *K-Means* karena nilai DBI yang lebih rendah dibandingkan *K-Medoids*. Pada aplikasi web tersebut

mencapai tingkat keakuratan sebesar 97%, baik dengan menggunakan tool RapidMiner ataupun perhitungan manual.

Paper [14] melakukan perbandingan antara algoritma K-eans dan *K-Medoids*. Fokus utama penelitian ini adalah pemetaan risiko bencana, khususnya di daerah kabupaten/kota Provinsi Jawa Barat yang sering mengalami bencana tanah longsor. Metode pengelompokkan algoritma *K-Means* terbukti lebih optimal dibandingkan dengan menggunakan algoritma *K-Medoids*. Jumlah klaster yang paling optimal ditemukan pada nilai k=6. Hasil identifikasi menunjukkan klaster 2 menjadi klaster dengan jumlah daerah terbanyak. Klaster 5 juga ditemukan memiliki jumlah kejadian terbanyak, yaitu 106 kejadian, yang terjadi di 4 daerah.

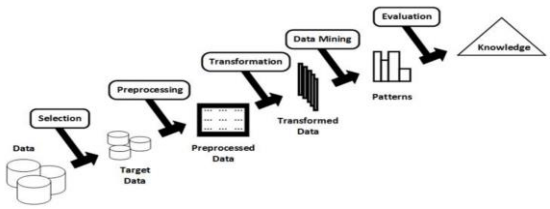
3. METODE PENELITIAN

Penelitian ini melakukan pengujian terhadap data kasus HIV di Jawa Barat dengan motode komparasi/perbandingan algoritma *K-Means* dan *K-Medoids clustering*. Proses komparasi dilakukan untuk memastikan hasil mana yang lebih optimal dalam melakukan pengelompokan data. Alur pada metode ini terlihat pada gambar berikut:



Gambar 1. Proses metode komparasi Algoritma K-Means dan K-Medoids

Gambar tersebut menunjukkan alur proses metode komparasi kedua algoritma, sebagai langkah untuk proses dan hasil yang lebih terperinci penelitian ini menggunakan metode *Knowledge Discovery in Databases* (KDD) dalam analisis dan pemrosesan data.



Gambar 2. *Proses Knowledge Discovery in Databases (KDD)*

Dari gambar diatas diketahui tahapan KDD terdiri dari:

3.1. *Data Selection*

Tahap ini melibatkan pemilihan data yang relevan untuk dianalisis. Pada proses data selection ini peneliti mengambil data jumlah kasus HIV berdasarkan kelompok umur di Provinsi Jawa Barat dalam kurun tahun 2019-2021 dengan format file excel.

3.2. *Preprocessing*

Setelah data terpilih, langkah selanjutnya adalah melakukan *pre-processing* pada data, yaitu proses pembersihan data seperti penanganan missing values, dan menghilangkan beberapa atribut yang tidak diperlukan dalam perhitungan algoritma *clustering K-Means* dan *K-Medoids* nantinya.

3.3. *Transformation*

Transformasi data dilakukan untuk meningkatkan efektivitas analisis data. Pada proses transformasi peneliti mengubah data yang sebelumnya nominal menjadi data bertipe numerik.

3.4. *Data mining*

Algoritma *K-Means* dan *K-Medoids* diterapkan pada data yang sebelumnya telah di proses untuk dilakukan pengelompokkan. *Clustering* dimulai dari klaster K=3 sampai K=10 di kedua algoritma. Pemilihan rentang jumlah K=3 hingga K=10 dalam penelitian ini didasari oleh tingkat karakteristik data kasus HIV di Jawa Barat yang memiliki kompleksitas data meliputi wilayah, jenis kelamin, umur, dan tahun, hal ini mencerminkan keberagaman potensi kelompok dalam data yang ingin dieksplorasi dan emilihan nilai K yang luas memungkinkan integrasi yang lebih baik antara hasil analisis klaster. Pemilihan K=3 sebagai jumlah klaster minimum bertujuan memberikan dasar untuk pengelompokkan data menjadi tingkat kasus HIV yang relatif tinggi, sedang dan rendah. Sedangkan pemilihan K=10 sebagai klaster maksimum memberikan ruang untuk mengembangkan klaster tambahan jika hasil analisis menunjukkan adanya kompleksitas yang memerlukan pengelompokkan lebih rinci.

3.5. *Evaluation/ Interpretation*

Pada tahap evaluasi, peneliti mengukur performa kedua algoritma berdasarkan nilai evaluasi *Davies-Bouldin Index* (DBI) dan evaluasi waktu komputasi dalam pengelompokan kasus HIV di Jawa Barat. Perbandingan nilai DBI dan waktu komputasi dilakukan untuk menentukan algoritma yang memberikan hasil *clustering* yang paling optimal. *Indeks Davies-Bouldin* (DBI) sendiri merupakan metrik yang digunakan untuk menilai kualitas klaster dalam analisis data. Nilai DBI menggabungkan konsep pemisahan dan kohesi antar klaster. Nilai DBI yang lebih rendah menunjukkan klasterisasi data yang lebih baik. [15]

4. HASIL DAN PEMBAHASAN

Hasil penelitian akan menghasilkan *clustering* kasus HIV Jawa Barat berdasarkan wilayah, jenis kelamin dan usia melalui proses KDD menggunakan algoritma *K-Means* dan *K-Medoids* dengan bantuan tools RapidMiner.

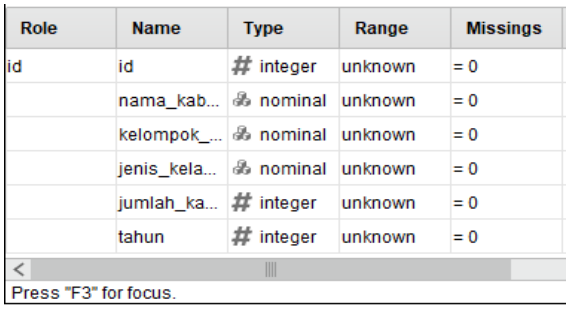
4.1. *Data Selection*

Tahap pertama yang dilakukan adalah pemilihan data kasus HIV Provinsi Jawa Barat tahun 2019-2021 berdasarkan usia. Data diperoleh melalui situs resmi https://opendata.jabarprov.go.id/. Terdapat sebanyak 971 data dengan atribut sebagai berikut:

Tabel 1. Atribut dataset kasus HIV di Jawa Barat

Nama Attribut	Type	Missings
kode_provinsi	*Interger*	0
nama_provinsi	*Polynominal*	0
kode_kabupaten_kota	*Interger*	0
nama_kabupaten_kota	*Polynominal*	0
jumlah_kasus	*Interger*	0
kelompok_umur	*Polynominal*	0
Satuan	*Polynominal*	0
Tahun	*Interger*	0

Tabel 1. Menunjukkan nilai attribut berdasarkan tipe datanya. Dari attribut tersebut dipilih sebanyak 6 atribut sebagai berikut:



Role	Name	Type	Range	Missings
id	id	integer	unknown	= 0
	nama_kab...	nominal	unknown	= 0
	kelompok_...	nominal	unknown	= 0
	jenis_kela...	nominal	unknown	= 0
	jumlah_ka...	integer	unknown	= 0
	tahun	integer	unknown	= 0

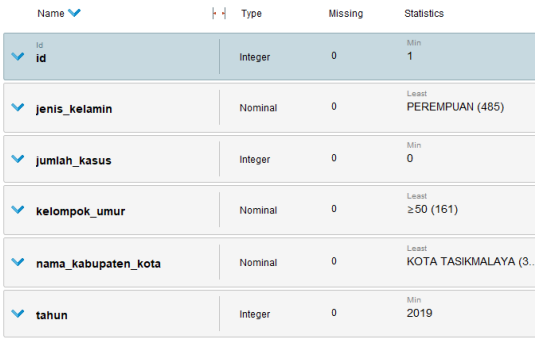
Press "F3" for focus.

Gambar 3. Atribut yang dipakai

Gambar diatas menunjukkan hasil pemilihan 6 attribut yang dipakai yaitu id, nama_kabupaten_kota, kelompok_umur, jenis_kelamin, jumlah_kasus, dan tahun. Attribut id dikategorikan sebagai 'id'. Agar pada proses *clustering* data 'id' tidak ikut serta terhitung dan merubah hasil.

4.2. *Preprocessing*

Tahap ini dilakukan pengecekan terhadap nilai *missig value*, hal ini dilakukan agar saat melakukan proses *clustering* pada RapidMiner tidak terjadi pesan error.



Name	Type	Missing	Statistics
id	Integer	0	Min 1
jenis_kelamin	Nominal	0	Least PEREMPUAN (485)
jumlah_kasus	Integer	0	Min 0
kelompok_umur	Nominal	0	Least ≥50 (161)
nama_kabupaten_kota	Nominal	0	Least KOTA TASIKMALAYA (3..
tahun	Integer	0	Min 2019

Gambar 4. Cek *Missing Value*

Pada gambar 4 terlihat kolom *missing* menunjukkan 0 pada setiap attribut hal ini menandakan tidak adanya *missing value* dan dapat di proses ke tahap berikutnya yaitu tahap *transformation*.

4.3. *Transformation*

Tahap ini akan meng-*konversi* nilai attribut pada tahun yang sebelumnya bertipe *numerik* menjadi *polynominal* yang kemudian di transformasi lagi bersama dengan attribut kelompok_umur dan jenis_kelamin menjadi nilai numerik dengan parameter *unique intergers*.

Tabel 2. Tranformasi nilai pada atribut nama_kabupaten_kota

No	Nama Kabupaten Kota	Transformasi
1.	KABUPATEN BOGOR	0
2.	KABUPATEN SUKABUMI	1
3.	KABUPATEN CIANJUR	2
4.	KABUPATEN BANDUNG	3
5.	KABUPATEN GARUT	4
6.	KABUPATEN TASIKMALAYA	5
7.	KABUPATEN CIAMIS	6
8.	KABUPATEN KUNINGAN	7
9.	KABUPATEN CIREBON	8
10.	KABUPATEN MAJALENGKA	9
11.	KABUPATEN SUMEDANG	10
12.	KABUPATEN INDRAMAYU	11
13.	KABUPATEN SUBANG	12
14.	KABUPATEN PURWAKARTA	13
15.	KABUPATEN KARAWANG	14
16.	KABUPATEN BEKASI	15
17.	KABUPATEN BANDUNG BARAT	16
18.	KABUPATEN PANGANDARAN	17
19.	KOTA BOGOR	18
20.	KOTA SUKABUMI	19
21.	KOTA BANDUNG	20
22.	KOTA CIREBON	21
23.	KOTA BEKASI	22
24.	KOTA DEPOK	23
25.	KOTA CIMAHI	24
26.	KOTA TASIKMALAYA	25
27.	KOTA BANJAR	26

Terlihat pada tabel 2 perubahan tipe nilai numerik pada attribut nama_kabupaten_kota

Tabel 3. Tranformasi nilai pada atribut kelompok_umur

No	Kelompok Umur	Transformasi
1.	0-4	0
2.	5-14	1
3.	15-19	2
4.	20-24	3
5.	25-49	4
6.	≥50	5

Terlihat pada tabel 3 perubahan tipe nilai numerik pada attribut kelompok_umur

Tabel 4.Tranformasi nilai pada atribut jenis_kelamin

No	Jenis Kelamin	Transformasi
1.	LAKI-LAKI	0
2.	PEREMPUAN	1

Terlihat pada tabel 4 perubahan tipe nilai numerik pada attribut jenis_kelamin

Tabel 5.Tranformasi nilai pada atribut tahun

S	Tahun	Transformasi
1.	2019	0
2.	2020	1
3.	2021	2

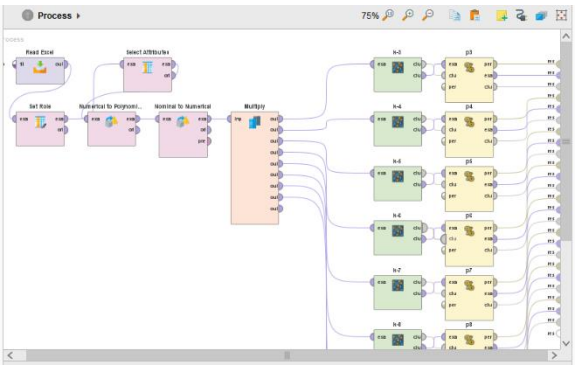
Terlihat pada tabel 5 perubahan tipe nilai numerik pada attribut tahun

4.4. *Data Mining*

Proses *Clustering* menggunakan algoritma *K-Means* dan *K-Medoids*, untuk menentukan acuan pengelompokkan *cluster* yang paling optimal, dilihat dari nilai *Davies-Boudien Index* (DBI) terkecil dan waktu komputasi RapidMiner dalam melakukan *clustering* pada kedua algoritma.

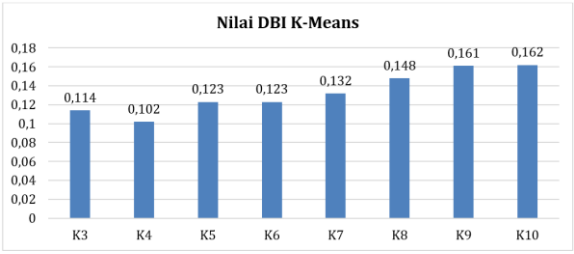
4.4.1. *K-Means*

Perhitungan *clustering* dilakukan dengan menggunakan metode jarak *Euclidean Distance* dengan settingan *measure tipes numericalmeasure* pada RapidMiner.



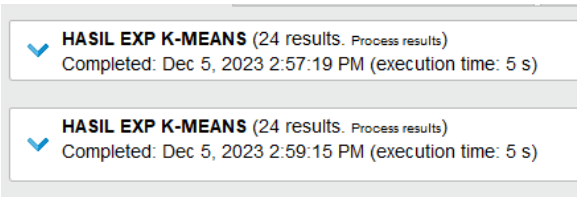
Gambar 5. Tampilan Model Algoritma *K-Means* pada RapidMiner

Pada gambar 5. terlihat pengelompokan data pada kedua algoritma dilakukan pada jumlah *cluster* K=3 sampai dengan K=10. Perolehan nilai *Davies-Boudien Index* dari proses *clustering* data *K-Means* terlihat pada gambar berikut:



Gambar 6. Nilai DBI *K-Means*

Berdasarkan gambar 6, evaluasi *Davies Bouldin Index*, ditemukan bahwa K=4 merupakan jumlah klaster terbaik untuk algoritma *K-Means* dengan nilai DBI terkecil 0,102. Hasil *clustering* menunjukkan pembagian data dengan 813 anggota untuk klaster 1, 40 anggota untuk klaster 2, 107 anggota untuk klaster 3, dan 11 anggota untuk klaster 4. Dan waktu komputasi yang dibutuhkan dalam pengelompokkan data pada RapidMiner terlihat sebagai berikut:

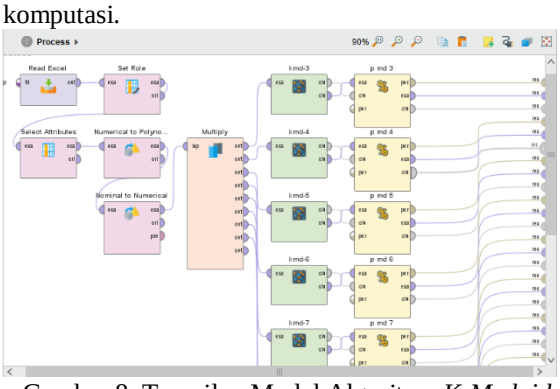


Gambar 7. Waktu Komputasi *K-Means*

Pada gambar diatas terlihat waktu komputasi yang dilakukan pada jumlah K=3 hingga K=10 algortima *K-Means* sebanyak dua kali percobaan *running*, diketahui membutuhkan waktu 5 detik,

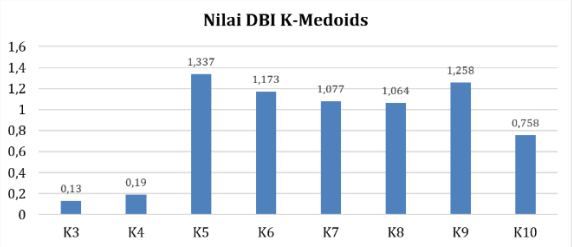
4.4.2. *K-Medoids*

Sama seperti metode sebelumnya dalam pengujian algoritma *K-Medoids*, pengelompokkan data dilakukan dengan menggunakan metode jarak *Euclidean Distance* dan pengukuran performa jumlah klaster terbaik dilihat dari nilai DBI terkecil dan waktu komputasi.



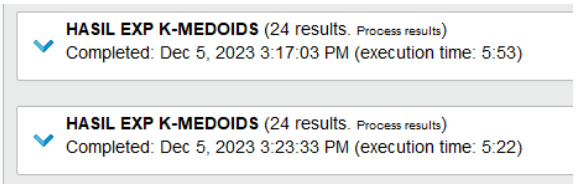
Gambar 8. Tampilan Model Algoritma *K-Medoids* pada RapidMiner

Gambar 8 menunjukkan pemilihan jumlah *cluster* yang sama dengan algoritma *K-Means* yaitu dari K=3 sampai K=10. Perolehan nilai DBI dari proses pengelompokan data dengan algoritma *K-Medoids* sebagai berikut:



Gambar 9. Nilai DBI *K-Medoids*

Dengan melihat nilai DBI yang ditunjukkan pada gambar 9, diperoleh klaster terbaik pada algoritma *K-Medoids* terdapat pada K=3 dengan nilai DBI yaitu 0,13 menunjukkan pembagian data pada klaster 1 yaitu 116 anggota, klater 2 sebanyak 16 anggota, dan klaster 2 sebanyak 839 anggota.



Gambar 10. Waktu Komputasi *K-Medoids*

Berdasarkan gambar 10, waktu komputasi yang dibutuhkan dalam pengelompokkan *K-Medoids* yaitu 5 menit, pada percobaan pertama membutuhkan waktu 5 menit 53 detik dan percobaan kedua 5 menit 22 detik.

4.5. *Evalution/Interpretation*

Tahap ini dilakukan analisis komparasi pada algoritma *K-Means* dan *K-Medoids* dengan melakukan evaluasi keakuratan klaster melalui perhitungan nilai DBI kedua algoritma. Semakin kecil nilai DBI, maka semkain baik hasil klasternya. Perbandingan juga dilakukan dengan melihat waktu komputasi *clustering* RapidMiner, hal ini dilakukann untuk mengukur hasil *performance* kecepatan pengelompokkan antara algoritma *K-Means* dan *K-Medoids*.
Adapun perbandingan nilai pengujian DBI yang didapatkan dari kedua algoritma pada tabel berikut ini:

Tabel 6. Perbandingan Nilai DBI *K-Means* dan *K-Medoids*

Klaster	DBI *K-Means*	DBI *K-Medoids*
K=3	0, 114	0, 13
K=4	0, 102	0, 19
K=5	0, 123	1, 337
K=6	0, 123	1, 173
K=7	0, 132	1, 077
K=8	0, 148	1, 064
K=9	0, 161	1, 258
K=10	0, 162	0, 758

Dari hasil pengujian nilai DBI tabel di atas dan hasil perbandingan waktu komputasi kedua algoritma, ditemukan bahwa algoritma *K-Means* merupakan metode yang paling baik dan optimal dalam melakukan pengelompokkan data penyebaran HIV di Provinsi Jawa Barat, hal ini dapat dibuktikan dengan nilai DBI yang dihasilkan *K-Means* lebih baik dibandingkan dengan algoritma *K-Medoids*. Nilai DBI pada algoritma *K-Means* mendekati 0 pada tiap-tiap *cluster*nya, sedangkan pada algoritma *K-Medoids* nilai DBI pada tiap-tiap *cluster* terlihat lebih besar dibandingkan dengan algoritma *K-Means*. Hasil pengujian menunjukkan bahwa klaster optimal pada algoritma *K-Means* didapat pada pengujian kedua dengan k=4 senilai 0,102, sedangkan pada algoritma *K-Medoids* klaster optimal ditunjukkan pada pengujian pertama dengan K=3 senilai 0,130. Selain itu, dalam hal perbandingan kinerja kecepatan diketahui bahwa algoritma *K-Means* memiliki waktu yang relatif lebih cepat dibandingkan dengan algoritma *K-Medoids*. Hal ini dibuktikan dengan waktu perhitungan pada rata-raata algoritma *K-Means* sebesar 5 detik, sedangkan pada algoritma *K-Medoids* memerlukan waktu perhitungan lebih lama yaitu sekitar 5 menit.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis *cluster* pengelompokkan data penyebaran HIV di Jawa Barat dengan algoritma *K-Means* dan K-Medois, disimpulkan bahwa *K-Means* merupakan metode yang paling baik dan optimal. Hal ini dibuktkan dengan nilai validasi *Davies Bouldin Index* (DBI) dan waktu komputasi rapidminer. Hasil pengujian pada algoritma *K-Means* didapat pada pengujian kedua yaitu K=4 dengan nilai DBI 0,102, sedangkan pada algoritma *K-Medoids* klaster optimal ditunjukkan pada pengujian pertama dengan K=3 senilai 0,130. Selain itu, dalam hal perbandingan kinerja kecepatan diketahui bahwa algoritma *K-Means* memiliki waktu yang relatif lebih cepat dibandingkan dengan algoritma *K-Medoids*. Hal ini dibuktikan dengan waktu komputasi pada algoritma *K-Means* sekitar 5 detik, sedangkan pada algoritma *K-Medoids* memerlukan waktu perhitungan lebih lama yaitu sekitar 5 menit. Berdasarkan hasil penelitian yang didapat, penelitian ini menyarankan pemerintah Daerah Jawa Barat untuk menngunakan algoritma *K-Means* dalam melakukan pengelompokkan data penyebaran kasus penyebaran HIV di Jawa Barat.

DAFTAR PUSTAKA

[1] R. I. Kemenkes, “Profil Kesehatan Indonesia Tahun 2021. Kementerian Kesehatan Republik Indonesia,” *Kementerian Kesehatan Republik Indonesia*. Depkes Ri, Jakarta, 2021. [Online]. Available: https://www.kemkes.go.id/downloads/resources/download/pusdatin/profil-kesehatan-indonesia/Profil-Kesehatan-2021.pdf

[2] Dinkesjabar, *Profil Kesehatan Jawa Barat 2022*. Dinas Kesehatan Provinsi Jawa Barat, 2022. [Online]. Available: https://diskes.jabarprov.go.id

[3] R. A. Farissa, R. Mayasari, and Y. Umaidah, "Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokkan Data Obat dengan Silhouette Coefficient di Puskesmas Karangsambung," *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 109–116, 2021, doi: 10.30871/jaic.v5i1.3237.

[4] S. Hasyrif, Rismayani, and S. Asrul, "PROSIDING SEMINAR ILMIAH SISTEM INFORMASI DAN TEKNOLOGI INFORMASI Pusat Penelitian dan Pengabdian Pada Masyarakat (P4M) STMIK Dipanegara Makassar Data Mining Menggunakan Algoritma K-Means Pengelompokan Penyebaran Diare Di Kota Makassar," *SISITI Semin. Ilm. Sist. Inf. dan Teknol. Inf.*, vol. VIII, no. 1, pp. 73–82, 2019.

[5] N. Puspitasari, G. Lempas, H. Hamdani, H. Haviuddin, and A. Septiarini, "Perbandingan Algoritma K-Means dan Algoritma K-Medoids Pada Kasus Covid-19 di Indonesia," *Build. Informatics, Technol. Sci.*, vol. 4, no. 4, pp. 2015–2027, 2023, doi: 10.47065/bits.v4i4.2994.

[6] N. A. S. Z. Abidin, R. D. Avila, A. Hermatyar, and R. Rismayani, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Daerah Produksi Kakao," *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 2, pp. 383–391, 2022, doi: 10.28932/jutisi.v8i2.4897.

[7] W. Utomo, "The comparison of k-means and k-medoids algorithms for clustering the spread of the covid-19 outbreak in Indonesia," *Ilk. J. Ilm.*, vol. 13, no. 1, pp. 31–35, 2021, doi: 10.33096/ilkom.v13i1.763.31-35.

[8] S. Ramadhani, D. Azzahra, and T. Z, "Perbandingan Algoritma K-Means Dan K-Medoids Pada Text Mining untuk Klasifikasi Tesis Mahasiswa Berdasarkan Pengujian Davies Bouldin Index," *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 13, no. 1, pp. 24–33, 2022, doi: 10.31849/digitalzone.v13i1.9292.

[9] E. ; Prasetyaningrum and P. Susanti, "Perbandingan Algoritma K-Means Dan K-Medoids Untuk Pemetaan Hasil Produksi Buah-Buahan," *J. MEDIA Inform. BUDIDARMA*, vol. 7, no. 4, pp. 1775–1783, 2023, doi: 10.30865/mib.v7i4.6477.

[10] A. Meiriza, E. Ali, Rahmiati, and Agustin, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Program BPJS Ketenagakerjaan," *Indones. J. Comput. Sci.*, vol. 12, no. 2, pp. 714–728, 2023, doi: 10.33022/ijcs.v12i2.3184.

[11] N. T. Luchia and M. Mustakim, "Perbandingan Algoritma K-Means Dan K-Medoids Pada Pengelompokan Humidity, Temperature, Dan Voltage Di Data Center Perawang," *J. Inf. Syst. Res.*, vol. 4, no. 1, pp. 184–190, 2022, doi: 10.47065/josh.v4i1.2385.

[12] N. T. Luchia, H. Handayani, F. S. Hamdi, D. Erlangga, and S. F. Octavia, "Perbandingan K-Means dan K-Medoids Pada Pengelompokan Data Miskin di Indonesia," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 2, pp. 35–41, 2022, doi: 10.57152/malcom.v2i2.422.

[13] I. Supriyadi, Aceng;Triayudi Ageng;Diana Sholihati, "Perbandingan algoritma k-means dengan k-medoids pada pengelompokan armada kendaraan truk berdasarkan produktivitas," vol. 06, pp. 229–240, 2021.

[14] T. K. Herviany, Mufidah;Putri Delima, Saleha;Nurhidayah, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokkan Daerah Rawan Tanah Longsor di Provinsi Jawa Barat," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 34–40, 2021, [Online]. Available: https://journal.irpi.or.id/index.php/malcom/article/view/60

[15] Y. A. Wijaya, D. A. Kurniady, E. Setyanto, W. S. Tarihoran, D. Rusmana, and R. Rahim, "Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities," *TEM J.*, vol. 10, no. 3, pp. 1099–1103, 2021, doi: 10.18421/TEM103.