

# Penerapan Metode *Clustering* dengan Algoritma K-Means untuk Pengelompokan Data Migrasi Penduduk Tiap Kecamatan di Kabupaten Rembang

Nayla Nur Afidah<sup>a,\*</sup>, Masrukan<sup>b</sup>

<sup>a, b</sup> Universitas Negeri Semarang, Gunungpati, Semarang, 50229, Indonesia

\* Alamat Surel: [naylaafidah24@students.unnes.ac.id](mailto:naylaafidah24@students.unnes.ac.id)

## Abstrak

Adanya kemajuan teknologi informasi saat ini menjadikan kebutuhan akan informasi dari sekumpulan data semakin tinggi. Salah satunya adalah informasi mengenai bidang kependudukan yaitu migrasi. Migrasi merupakan satu dari tiga faktor dasar yang mempengaruhi dinamika kependudukan, selain faktor kelahiran dan kematian. Migrasi diartikan sebagai perpindahan penduduk dari satu tempat ke tempat yang lain dengan tujuan untuk menetap. Peninjauan migrasi sangat penting untuk ditelaah mengingat adanya desentralisasi dan distribusi penduduk yang tidak merata. Data penelitian diperoleh dari Dinas Kependudukan dan Pencatatan Sipil Kabupaten Rembang. Metode pengolahan data yang digunakan dalam pembahasan ini adalah metode *clustering* dengan algoritma K-Means, dimana data akan dikelompokkan menjadi 3 *cluster* yaitu tinggi (C1), sedang (C2), dan rendah (C3). Tujuan dari penelitian ini adalah untuk mengelompokkan data migrasi penduduk tiap kecamatan di Kabupaten Rembang. Analisis data menggunakan algoritma K-Means dilakukan dengan cara perhitungan manual dan memanfaatkan aplikasi *RapidMiner*. Hasil yang diperoleh dari pembahasan ini adalah bahwa terdapat sedikit perbedaan antara perhitungan manual dengan aplikasi *RapidMiner* dimana pada perhitungan manual *cluster* 1 (C1) sebanyak 7 kecamatan, *cluster* 2 (C2) sebanyak 6 kecamatan, dan *cluster* 3 (C3) sebanyak 1 kecamatan. Sedangkan pada aplikasi *RapidMiner* *cluster* 1 (C1) sebanyak 10 kecamatan, *cluster* 2 (C2) sebanyak 3 kecamatan, dan *cluster* 3 (C3) sebanyak 1 kecamatan.

Kata kunci:

*Clustering*, Algoritma K-Means, Migrasi

©2023 Dipublikasikan oleh Jurusan Matematika, Universitas Negeri Semarang

## 1. Pendahuluan

Seiring dengan majunya teknologi informasi saat ini menjadikan kebutuhan akan informasi dari sekumpulan data semakin berkembang pesat. Hal ini karena adanya kemudahan teknologi yang mampu menyelesaikan masalah beraneka ragam dan kompleks yang meliputi segala bidang kehidupan, tak terkecuali dengan bidang kependudukan. Aspek bidang kependudukan yang sangat perlu diperhatikan salah satunya adalah migrasi. Migrasi merupakan satu dari tiga faktor dasar yang mempengaruhi dinamika kependudukan, selain faktor kelahiran dan kematian. Migrasi diartikan sebagai perpindahan penduduk dari satu tempat ke tempat yang lain dengan tujuan untuk menetap. Adanya migrasi tentu akan berdampak pada perubahan stuktur penduduk. Migrasi masuk yang lebih besar daripada migrasi keluar akan menyebabkan jumlah penduduk meningkat. Sebaliknya, jika migrasi keluar lebih besar daripada migrasi masuk akan berakibat pada pengurangan penduduk. Jika terjadi kasus seperti itu, maka berbagai kemungkinan buruk bisa terjadi apabila tidak segera dikendalikan.

Berdasarkan data Badan Pusat Statistik tahun 2015 menyatakan bahwa Provinsi Jawa Tengah menjadi peringkat kedua dengan tingkat migrasi risen tertinggi, baik itu migrasi masuk maupun migrasi keluar. Terhitung sebanyak 518.103 orang melakukan migrasi masuk ke Jawa Tengah dan migrasi keluar dari Jawa

To cite this article:

Afidah, N. N. & Masrukan. (2023). Penerapan Metode *Clustering* dengan Algoritma K-means untuk Pengelompokan Data Migrasi Penduduk Tiap Kecamatan di Kabupaten Rembang. *PRISMA, Prosiding Seminar Nasional Matematika* 6, 729-738

Tengah sebanyak 647.482 orang. Di Kabupaten Rembang yang merupakan salah satu kabupaten di Jawa Tengah tercatat hingga sampai dengan bulan oktober 2021 sudah sebanyak 20.240 orang melakukan migrasi masuk dan 27.427 orang melakukan migrasi keluar (Dinas Kependudukan dan Pencatatan Sipil, 2021). Menurut Munir (2000) peninjauan migrasi secara regional sangat penting untuk ditelaah mengingat adanya desentralisasi (kepadatan) dan distribusi penduduk yang tidak merata.

*Clustering* atau klasterisasi merupakan suatu teknik atau metode untuk mengelompokkan data. *Clustering* adalah salah satu alat yang penting dalam pengolahan data statistik untuk melakukan analisis data. Analisis *cluster* merupakan seperangkat metode yang digunakan untuk mengelompokkan objek ke dalam sebuah *cluster* berdasarkan informasi yang ditemukan pada data. Saat ini analisis *cluster* telah banyak digunakan di berbagai bidang ilmu pengetahuan seperti ekonomi, psikologi, kesehatan, sosial masyarakat dan kependudukan (Dani *et. al.*, 2019). Salah satu metode *clustering* yang paling terkenal adalah K-Means.

K-Means merupakan salah satu algoritma *clustering* non hirarki yang berusaha untuk mempartisi objek yang ada ke dalam satu atau lebih *cluster* atau kelompok objek berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama dikelompokkan dalam satu *cluster* yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokkan kedalam *cluster* yang lain (Musharyadi, 2017). K-means merupakan algoritma yang mengklasterisasikan berdasarkan partisi dan melakukan *clustering* melalui proses iterasi berkelanjutan sampai dengan bertemu kondisi akhir, proses iterasi berhenti dan hasil clustering adalah output (Fakhri *et. al.*, 2021).

Berbagai penelitian tentang penggunaan metode *clustering* dengan algoritma K-means telah banyak dilakukan sebelumnya. Beberapa diantaranya yaitu Mardalius (2018) membahas penggunaan algoritma K-means untuk pengelompokkan data penjualan aksesoris, Mudzakkir (2018) membahas penggunaan algoritma K-means untuk pengelompokkan data penjualan produk pada PT Advanta Seeds, dan Napitupulu *et al.* (2020) membahas penggunaan algoritma K-means untuk pengelompokkan dokumen akta kelahiran pada tiap kecamatan di kabupaten Simalungun. Ketiga penelitian tersebut menerapkan metode *clustering* dengan algoritma K-means untuk mengelompokkan suatu data. Sehingga diharapkan dengan adanya pengelompokkan data tersebut diperoleh informasi untuk dijadikan acuan dalam mengambil keputusan yang tepat.

Dalam penelitian ini akan membahas penerapan metode *clustering* dengan algoritma K-means dalam mengelompokkan data migrasi penduduk di Kabupaten Rembang dengan cara mengklasifikasikan tiap kecamatan pada *cluster* yang telah ditetapkan. Metode K-Means dipilih karena metode ini menggunakan data fisik tidak abstrak dan bersifat jelas, hal ini sesuai dengan data yang akan digunakan pada permasalahan di dalam pengelompokkan data migrasi penduduk. Selain itu, metode ini bersifat fleksibel sebab dapat menentukan jumlah *cluster* yang akan dibuat. Adapun tujuan dari penelitian ini adalah untuk mengetahui penerapan dan hasil *clustering* dengan algoritma K-Means dalam mengelompokkan data migrasi penduduk tiap kecamatan di Kabupaten Rembang.

---

## 2. Metode

Metode penelitian merupakan cara atau langkah yang digunakan penulis dalam melakukan suatu penelitian melalui tahapan-tahapan yang disusun secara ilmiah sehingga dapat dipergunakan untuk menganalisa, mengerjakan, dan mengatasi masalah yang dihadapi. Pada rancangan penelitian yang digunakan yaitu mempelajari literatur, mengumpulkan data, menganalisa data, menganalisa metode K-Means, implementasi software, menguji hasil, dan menganalisa hasil.

### 2.1. *Clustering*

Menurut Mardalius (2018) *clustering* merupakan metode yang digunakan untuk membagi rangkaian data menjadi beberapa kelompok berdasarkan kemiripan karakteristik yang telah ditentukan sebelumnya. Irwansyah & Faisal (2015) mengemukakan *clustering* adalah proses partisi atau membagi satu set objek data ke dalam himpunan bagian yang disebut dengan *cluster*.

### 2.2. *Algoritma K-Means*

K-Means adalah salah satu algoritma klastering dengan metode partisi (*partitioning method*) yang berbasis titik pusat (*centroid*). K-Means merupakan metode analisis kelompok yang mengarah pada pembagian  $n$  obyek pengamatan ke dalam  $k$  kelompok (*cluster*). Metode ini berusaha membagi data kedalam kelompok sehingga data yang berkarakteristik sama dimasukkan ke dalam satu kelompok sementara data yang berkarakteristik berbeda dimasukkan dalam kelompok yang lain (Suprayogi, 2013).

Menurut Wakhidah (2017) langkah-langkah melakukan clustering dengan algoritma K-Means adalah sebagai berikut :

- 1) Menentukan jumlah *cluster* ( $k$ )
- 2) Menentukan nilai *centroid*

Dalam menentukan nilai *centroid* untuk awal iterasi, nilai awal *centroid* dilakukan secara acak. Sedangkan jika menentukan nilai *centroid* yang merupakan tahap dari iterasi, maka digunakan rumus sebagai berikut.

$$v_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} x_{kj}$$

dimana :

$v_{ij}$  = centroid rata-rata *cluster* ke-I untuk variabel ke-j

$N_i$  = jumlah data yang menjadi anggota *cluster* ke-i

$i, k$  = indeks dari *cluster*

$j$  = indeks dari variabel

$x_{kj}$  = nilai data ke-k yang ada di dalam cluster tersebut untuk variable ke-j

- 3) Menghitung jarak antara titik *centroid* dengan titik tiap objek

Untuk menghitung jarak tersebut dapat menggunakan *Euclidean Distance*, yaitu:

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

dimana :

$D_e$  = *Euclidean Distance*

$i$  = banyaknya objek

$(x, y)$  = koordinat objek

$(s, t)$  = koordinat *centroid*

- 4) Pengelompokkan objek dengan memperhitungkan jarak minimum objek.

- 5) Kembali ke tahap 2, lakukan perulangan hingga nilai *centroid* yang dihasilkan tetap dan anggota *cluster* tidak berpindah ke *cluster* lain.



**Gambar 1.** Diagram alir algoritma K-Means

### 3. Hasil dan Pembahasan

Pada pengumpulan data ada dua parameter yang digunakan dalam pengolahan data yaitu migrasi masuk dan migrasi keluar. Pada penelitian ini data yang digunakan yaitu data pindah dan datang (migrasi masuk dan migrasi keluar) antar kabupaten/kota tiap kecamatan di Kabupaten Rembang dihitung mulai bulan Januari–Oktober 2021. Data tersebut diperoleh dari Dinas Kependudukan dan Pencatatan Sipil Kabupaten Rembang melalui Bidang Kependudukan bagian Pindah Datang dan Pendataan Penduduk. Data tersebut disajikan dalam bentuk tabel seperti terlihat pada tabel 1 berikut.

**Tabel 1.** Data migrasi masuk dan keluar tiap kecamatan di Kabupaten Rembang

No	Kecamatan	Migrasi	
		Masuk	Keluar
1	Sumber	100	175
2	Bulu	93	176
3	Gunem	90	110
4	Sale	182	225
5	Sarang	146	290
6	Sedan	127	259
7	Pamotan	176	245
8	Sulang	128	188
9	Kaliori	181	197
10	Rembang	454	662
11	Pancur	117	102
12	Kragan	215	275
13	Sluke	91	147
14	Lasem	283	379
<b>Jumlah</b>		<b>2383</b>	<b>3430</b>

Data di atas selanjutnya akan dilakukan pengolahan data dengan proses *clustering* menggunakan algoritma K-Means sehingga didapatkan hasil pengelompokan data yang diinginkan. Adapun langkah *clustering* dengan algoritma K-Means adalah sebagai berikut :

1) Menentukan jumlah *cluster* ( $k$ )

Pengelompokan data pada pengujian ini sebanyak 3 *cluster* ( $k = 3$ ), diantaranya yaitu *cluster* 1 menunjukkan tingkat migrasi rendah (C1), *cluster* 2 menunjukkan tingkat migrasi sedang (C2), dan *cluster* 3 menunjukkan tingkat migrasi tinggi (C3).

2) Menentukan *centroid* (pusat *cluster*)

Penentuan pusat awal *cluster* ditentukan secara random yang diambil dari data pada tabel 1 Sehingga terpilih untuk *cluster* 1 (C1) Kecamatan Sarang, *cluster* 2 (C2) Kecamatan Kragan, dan *cluster* 3 (C3) Kecamatan Rembang.

**Tabel 2.** Pusat *cluster* awal

Migrasi	C1	C2	C3
Masuk	146	215	454
Keluar	290	275	662

3) Menghitung jarak setiap objek ke *centroid* pada masing-masing *cluster*

Rumus untuk menghitung jarak dari *centroid* adalah :

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

Dengan dilakukan perhitungan menggunakan rumus di atas diperoleh tabel hasil perhitungan jarak *centroid* sebagai berikut.

**Tabel 3.** Perhitungan jarak *centroid* iterasi ke-1

Kecamatan	DC1	DC2	DC3
Sumber	123.859	152.398	602.067
Bulu	125.718	157.115	605.406
Gunem	188.510	207.002	661.211
Sale	74.303	59.908	514.736
Sarang	0.000	70.612	482.958
Sedan	36.359	89.443	518.978
Pamotan	54.083	49.204	501.172
Sulang	103.576	123.037	575.284
Kaliori	99.368	85.088	539.216
Rembang	482.958	454.852	0.000
Pancur	190.224	198.829	653.582
Kragan	70.612	0.000	454.852
Sluke	153.212	178.213	630.075
Lasem	163.371	124.258	330.651

4) Menentukan *cluster*

Alokasikan masing-masing objek (kecamatan) ke *cluster* yang paling terdekat berdasarkan nilai minimal dari nilai *cluster* dan letakkan pada *cluster* yang sesuai dengan nilai minimal pada iterasi 1.

**Tabel 4.** *Cluster* iterasi ke-1

Kecamatan	DC1	DC2	DC3	C1	C2	C3
Sumber	123.859	152.398	602.067	1	0	0
Bulu	125.718	157.115	605.406	1	0	0
Gunem	188.510	207.002	661.211	1	0	0
Sale	74.303	59.908	514.736	0	1	0
Sarang	0.000	70.612	482.958	1	0	0
Sedan	36.359	89.443	518.978	1	0	0
Pamotan	54.083	49.204	501.172	0	1	0
Sulang	103.576	123.037	575.284	1	0	0
Kaliori	99.368	85.088	539.216	0	1	0
Rembang	482.958	454.852	0.000	0	0	1
Pancur	190.224	198.829	653.582	1	0	0
Kragan	70.612	0.000	454.852	0	1	0
Sluke	153.212	178.213	630.075	1	0	0
Lasem	163.371	124.258	330.651	0	1	0

Dari tabel di atas didapatkan data dengan jarak terdekat C1 sebanyak 8, jarak terdekat C2 sebanyak 5, dan jarak terdekat C3 sebanyak 1.

5) Lakukan iterasi, kemudian tentukan posisi *centroid* baru dengan rumus,

$$v_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} x_{kj}$$

atau dalam hal ini dengan cara menjumlahkan nilai yang terpilih pada *cluster* tersebut kemudian membaginya sebanyak jumlah nilai. Sehingga data *centroid* baru iterasi ke-1 dapat dilihat pada tabel berikut.

**Tabel 5.** *Centroid* baru iterasi ke-1

Migrasi	C1	C2	C3
Masuk	111.5	207.4	454
Keluar	180.875	264.2	662

Dengan menggunakan langkah yang sama seperti sebelumnya untuk menentukan jarak dari *centroid* dengan menggunakan *centroid* baru iterasi ke-1, maka berikut hasil perhitungan jarak *centroid* dan *cluster* iterasi ke-2.

**Tabel 6.** Perhitungan jarak *centroid* dan hasil *cluster* iterasi ke-2

Kecamatan	DC1	DC2	DC3	C1	C2	C3
Sumber	12.914	139.612	602.067	1	0	0
Bulu	19.132	144.453	605.406	1	0	0
Gunem	74.064	193.805	661.211	1	0	0
Sale	83.170	46.710	514.736	0	1	0
Sarang	114.449	66.600	482.958	0	1	0
Sedan	79.648	80.568	518.978	1	0	0
Pamotan	90.952	36.805	501.172	0	1	0
Sulang	17.973	110.049	575.284	1	0	0
Kaliori	71.346	72.200	539.216	1	0	0
Rembang	590.582	468.035	0.000	0	0	1
Pancur	79.067	185.691	653.582	1	0	0
Kragan	139.899	13.206	454.852	0	1	0
Sluke	39.595	165.181	630.075	1	0	0
Lasem	262.042	137.457	330.651	0	1	0

Hasil dari *cluster* iterasi ke-1 dan *cluster* iterasi ke-2 menunjukkan bahwa terdapat 2 kecamatan yaitu Sarang dan Kaliori yang pindah ke *cluster* lain. Maka dilakukan perhitungan kembali sampai data *cluster* tidak berubah.

Selanjutnya mencari *centroid* dari iterasi ke-2. Maka diperoleh tabel *centroid* baru iterasi ke-2 sebagai berikut.

**Tabel 7.** *Centroid* baru iterasi ke-2

Migrasi	C1	C2	C3
Masuk	115.875	200.4	454
Keluar	169.25	282.8	662

**Tabel 8.** Perhitungan jarak *centroid* dan hasil *cluster* iterasi ke-3

Kecamatan	DC1	DC2	DC3	C1	C2	C3
Sumber	16.884	147.313	602.067	1	0	0
Bulu	23.850	151.463	605.406	1	0	0
Gunem	64.654	205.056	661.211	1	0	0
Sale	86.490	60.658	514.736	0	1	0
Sarang	124.451	54.874	482.958	0	1	0
Sedan	90.437	77.162	518.978	0	1	0
Pamotan	96.711	44.991	501.172	0	1	0
Sulang	22.329	119.285	575.284	1	0	0
Kaliori	70.791	87.966	539.216	1	0	0
Rembang	597.604	456.186	0.000	0	0	1
Pancur	67.259	199.109	653.582	1	0	0
Kragan	144.944	16.553	454.852	0	1	0
Sluke	33.374	174.385	630.075	1	0	0
Lasem	268.190	126.796	330.651	0	1	0

Dari tabel *cluster* iterasi ke-2 dan *cluster* iterasi ke-3 terlihat bahwa masih terdapat kecamatan yang berpindah *cluster* yaitu Kecamatan Sedan yang semula dari C1 pindah ke C2. Maka dari itu dilakukan perhitungan jarak dengan *centroid* baru.

**Tabel 9.** *Centroid* baru iterasi ke-3

Migrasi	C1	C2	C3
Masuk	114.2857	188.1667	454
Keluar	156.4286	278.8333	662

**Tabel 10.** Perhitungan jarak *centroid* dan hasil *cluster* iterasi ke-4

Kecamatan	DC1	DC2	DC3	C1	C2	C3
Sumber	23.430	136.216	602.067	1	0	0
Bulu	28.916	140.112	605.406	1	0	0
Gunem	52.397	195.298	661.211	1	0	0
Sale	96.370	54.185	514.736	0	1	0
Sarang	137.285	43.620	482.958	0	1	0
Sedan	103.356	64.302	518.978	0	1	0
Pamotan	107.952	35.954	501.172	0	1	0
Sulang	34.421	108.953	575.284	1	0	0
Kaliori	78.082	82.147	539.216	1	0	0
Rembang	609.104	466.352	0.000	0	0	1
Pancur	54.496	190.617	653.582	1	0	0
Kragan	155.572	27.106	454.852	0	1	0
Sluke	25.122	163.772	630.075	1	0	0
Lasem	279.289	137.937	330.651	0	1	0

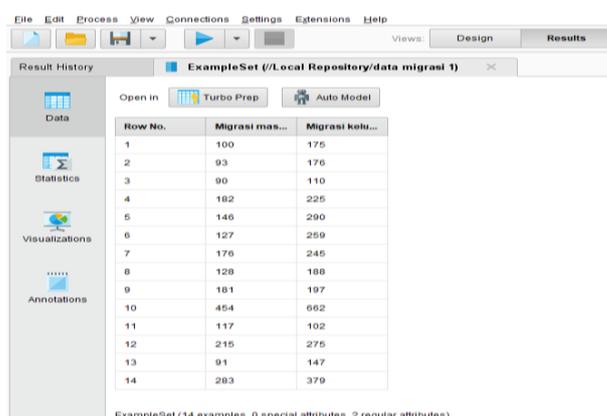
Dari tabel di atas dapat dilihat bahwa posisi *cluster* pada iterasi ke-4 sama dengan posisi *cluster* iterasi ke-3, artinya posisi *cluster* sudah tidak berubah. Maka perhitungan dihentikan dan diperoleh hasil akhir *clustering* sebagai berikut:

- Cluster* 1 (C1) memiliki 7 data yang berarti bahwa kelompok pertama merupakan kelompok dengan tingkat migrasi rendah selama periode Januari – Oktober 2021.
- Cluster* 2 (C2) memiliki 6 data yang berarti bahwa kelompok kedua merupakan kelompok dengan tingkat migrasi sedang selama periode Januari – Oktober 2021.
- Cluster* 3 (C3) memiliki 1 data yang berarti bahwa kelompok ketiga merupakan kelompok dengan tingkat migrasi tinggi selama periode Januari – Oktober 2021.

Pada implementasi dan pengujian dilakukan menggunakan aplikasi *RapidMiner Studio* v.9.10 dengan tujuan untuk membandingkan hasil pengolahan data secara manual dengan hasil pengolahan data menggunakan aplikasi. Adapun langkah-langkahnya adalah sebagai berikut.

#### 1) *Import* data

Tahap pertama adalah memasukkan data yang akan diuji dengan klik *import data* → *select the data location* → *select the cell to import* → *next* → *finish*.



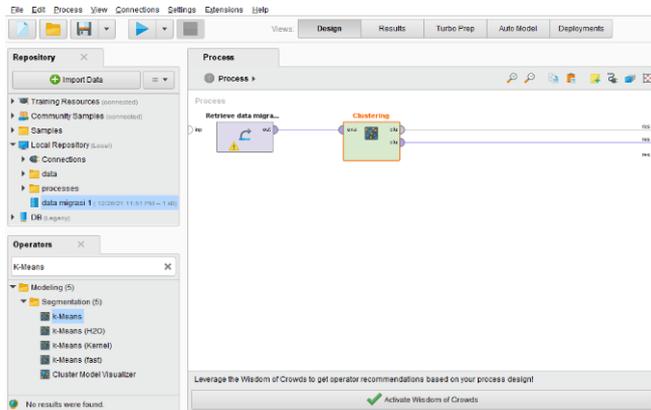
The screenshot shows the RapidMiner Studio interface with a table titled 'ExampleSet (/Local Repository/data migrasi 1)'. The table has three columns: 'Row No.', 'Migrasi mas...', and 'Migrasi kelu...'. The data is as follows:

Row No.	Migrasi mas...	Migrasi kelu...
1	100	175
2	93	176
3	90	110
4	182	225
5	146	290
6	127	259
7	176	245
8	128	188
9	181	197
10	454	662
11	117	102
12	215	275
13	91	147
14	283	379

**Gambar 2.** Tampilan *import data/ExampleSet*

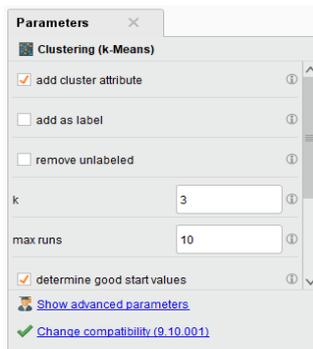
#### 2) Pengujian data dengan algoritma K-Means

Kemudian data tersebut dilakukan pengujian menggunakan *operators* K-Means dengan terlebih dahulu mengklik tombol *design* agar nampak tampilan layar *process*. Selanjutnya *drag and drop* data yang sebelumnya telah disimpan di *Local Repository* menuju ke layar *process*. Lalu tambahkan *operators* K-Means dengan cara menuliskan algoritma K-Means pada kolom pencarian *operators* lalu *drag and drop operators* K-Means pada layar *process* dan hubungkan data *ExampleSet* dengan *operators* K-Means agar proses *clustering* dapat berjalan, seperti gambar 2 berikut.



**Gambar 3.** Tampilan proses K-Means

Selanjutnya melakukan pengaturan algoritma K-Means berupa pengaturan nilai  $k$  yang menunjukkan jumlah *cluster* yang akan dibentuk. Di sini jumlah *cluster* yang akan dibentuk sebanyak 3 *cluster* yaitu rendah, sedang, dan tinggi. Lalu data di-*running* dengan mengklik tombol segitiga yang terletak pada bagian *toolbar*.



**Gambar 4.** Penentuan jumlah *cluster*

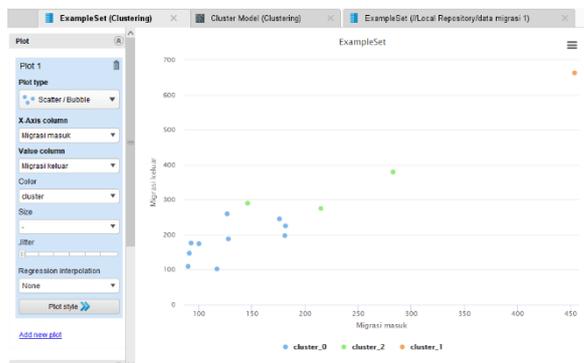
Setelah proses *running* selesai, maka akan muncul *output ExampleSet* hasil *clustering* K-Means, seperti pada gambar 4 berikut.

Row No.	id	cluster	Migrasi mas...	Migrasi kolu...
1	1	cluster_0	100	175
2	2	cluster_0	93	176
3	3	cluster_0	90	110
4	4	cluster_0	182	225
5	5	cluster_2	146	290
6	6	cluster_0	127	259
7	7	cluster_0	176	245
8	8	cluster_0	128	188
9	9	cluster_0	181	197
10	10	cluster_1	454	662
11	11	cluster_0	117	102
12	12	cluster_2	215	275
13	13	cluster_0	91	147
14	14	cluster_2	283	379

ExampleSet (14 examples, 2 special attributes, 2 regular attributes)

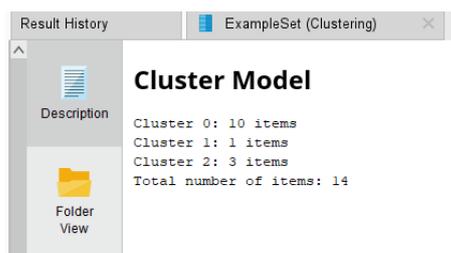
**Gambar 5.** Tampilan *ExampleSet* hasil *clustering*

Tampilan *ExampleSet* hasil *clustering* dapat dilihat dalam bentuk *scatter plot* seperti pada gambar 5 berikut.



**Gambar 6.** Tampilan *scatter plot* hasil *clustering*

Selain itu, hasil *clustering* juga dapat dilihat pada model *cluster* yang tampilannya berupa *text view* yang melihatkan hasil pengelompokkan berdasarkan *cluster* dan jumlah anggotanya, seperti pada gambar 6 berikut.



**Gambar 7.** Tampilan model *cluster*

Sehingga dari pengujian data menggunakan aplikasi *RapidMiner* diperoleh hasil *clustering* sebagai berikut.

- a. *Cluster 0* (C1) memiliki 10 data yang berarti bahwa kelompok pertama merupakan kelompok dengan tingkat migrasi rendah selama periode Januari – Oktober 2021.
- b. *Cluster 2* (C2) memiliki 3 data yang berarti bahwa kelompok kedua merupakan kelompok dengan tingkat migrasi sedang selama periode Januari – Oktober 2021.
- c. *Cluster 1* (C3) memiliki 1 data yang berarti bahwa kelompok ketiga merupakan kelompok dengan tingkat migrasi tinggi selama periode Januari – Oktober 2021.

#### 4. Simpulan

Berdasarkan hasil pembahasan yang telah diuraikan di atas, dapat ditarik kesimpulan bahwa metode *clustering* dengan algoritma K-Means dapat digunakan untuk mengelompokkan data migrasi penduduk dengan cara mengklasifikasikan tiap kecamatan pada *cluster* yang telah ditetapkan. Berdasarkan hasil *clustering* dengan algoritma K-Means antara perhitungan manual dengan aplikasi terdapat sedikit perbedaan dimana *Cluster 1* (C1), pada perhitungan manual diperoleh sebanyak 7 kecamatan meliputi Kecamatan Sumber, Bulu, Gunem, Sulang, Kaliori, Pancur, dan Sluke. Sedangkan pada aplikasi diperoleh sebanyak 10 kecamatan, meliputi Kecamatan Sumber, Bulu, Gunem, Sale, Sedan, Pamotan, Sulang, Kaliori, Pancur, dan Sluke. *Cluster 2* (C2), pada perhitungan manual diperoleh sebanyak 6 kecamatan, meliputi Kecamatan Sale, Sarang, Sedan, Pamotan, Kragan, dan Lasem. Sedangkan pada aplikasi diperoleh sebanyak 3 kecamatan, meliputi Kecamatan Sarang, Kragan, dan Lasem. *Cluster 3* (C3), baik perhitungan manual maupun aplikasi diperoleh hasil yang sama sebanyak 1 kecamatan yaitu Kecamatan Rembang.

---

**Daftar Pustaka**

- Badan Pusat Statistik (BPS). (2016). *Migrasi Risen (Recent Migration) Tahun 2015*.
- Dani, A. T. R., Wahyuningsih, S., & Rizki, N. A. (2019). Penerapan *Hierarchical Clustering* Metode *Agglomerative* pada Data Runtun Waktu. *Jambura Journal of Mathematics*, 1(2), 64-78.
- Dinas Kependudukan dan Pencatatan Sipil Kabupaten Rembang. (2021). *Pindah dan Datang Antar Kabupaten/Kota 2021*.
- Fakhri, D.A., Defit, S., & Sumijan. (2021). Optimalisasi Pelayanan Perpustakaan terhadap Minat Baca Menggunakan Metode K-Means *Clustering*. *Jurnal Informasi dan Teknologi*, 3(3), 160-166.
- Irwansyah, E. & Faisal, M. (2015). *Advanced Clustering : Teori dan Aplikasi*. Yogyakarta: DeePublish.
- Mardalius. (2018). Pemanfaatan *Rapid Miner Studio 8.2* Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma K-Means. *Jurnal Teknologi dan Sistem Informasi*, 4(2), 123-132.
- Mudzakkir, B. D. (2018). Pengelompokan Data Penjualan Produk Pada PT Advanta Seeds Indonesia Menggunakan Metode K-Means. *Jurnal Mahasiswa Teknik Informatika*, 2(2), 34-40.
- Munir, R. (2000). *Dasar-dasar Demografi*. Jakarta : Lembaga Demografi Fakultas Ekonomi Universitas Indonesia.
- Musharyadi, F. (2017). Tingkat Pemahaman Mahasiswa Terhadap Norma Agama Islam Menggunakan Algoritma K-Means *Clustering*. *Menara ilmu*, 11(78), 48-54.
- Napitupulu, F. S., Damanik, I. S., Saragih, I. S., & Wanto, A. (2020). Algoritma K-Means untuk Pengelompokan Dokumen Akta Kelahiran pada Tiap Kecamatan di Kabupaten Simalungun. *Building of Informatics Technology and Science (BITS)*, 2(1), 55-63.
- Suprayogi. (2013). *Data Mining Clustering*. Semarang : UDINUS.
- Wakhidah, N. (2017). *Clustering Menggunakan K-Means Algorithm (K-Means Algorithm Clustering)*. Fakultas Teknologi Informasi dan Komunikasi Universitas Semarang.