

PENGELOMPOKAN JUMLAH PENDUDUK BERDASARKAN KATEGORI USIA DENGAN METODE K-MEANS

Sahat Sonang S.¹, Arifin Tua Purba² Ferri Ojak Imanuel Pardede³

^{1,2,3} Teknik Komputer, Politeknik Bisnis Indonesia

Email : sahatsonangstg@gmail.com¹, arifintuaprb20@gmail.com², ferri_pardede@yahoo.com³

Abstract

Control of population is one of the tasks of the government in Indonesia. The increase and movement of population in each region makes a certain area to defeat changes in population surging, and this can affect the economic level of the area. This study aims to process the population of Pematangsiantar City in 2018 which is divided into age groups, namely: Toddlers, Young Children, Early Adolescents, Late Adolescents, Early Adulthood, Late Adulthood, Elderly, Late Elderly, and Upper Seniors. Data processing is done by using K-Means method clustering in accordance with the population of Pematangsiantar City per district. With this grouping, we can see that the number of population in each sub-district is based on each age group so that we can implement programs that are more appropriate in improving human resources.

Keywords: Clustering, Number of Population, K-Means.

1. PENDAHULUAN

Pertambahan jumlah penduduk di suatu kota ataupun kabupaten jika tidak diselesaikan dengan baik dan sistematis akan mengakibatkan dampak-dampak yang buruk seperti halnya banyaknya pengangguran, tingkat pencemaran lingkungan yang semakin meningkat, berkurangnya lahan terbuka hijau karena dipergunakan untuk pemukiman penduduk, pertumbuhan ekonomi, dan dampak-dampak lainnya. Pertambahan jumlah penduduk juga akan mempengaruhi pertumbuhan ekonomi.

Pengolahan Jumlah penduduk dalam penelitian ini dikelompokkan berdasarkan usia, yaitu : Balita, Kanak-Kanak, Remaja Awal, Remaja Akhir, Dewasa Awal, Dewasa Akhir, Lansia Awal, Lansia Akhir, dan Manula Atas yang diperoleh data jumlah penduduk per kecamatan di Kota Pematangsiantar tahun 2018. Melalui data ini peneliti akan mengelompokkan kecamatan yang jumlah penduduk yang lebih tinggi.

Teknologi data mining merupakan salah satu alat bantu untuk penggalian data pada basis data berukuran besar dan dengan spesifikasi tingkat kerumitan yang telah banyak digunakan pada banyak domain aplikasi seperti perbankan maupun bidang telekomunikasi [1], [2].

Metode K-Means adalah salah satu metode cluster non hirarki yang bertujuan untuk membantu pengelompokan variabel untuk dimasukkan kedalam kelas-kelas yang terletak pada hasil akhir perhitungan [3].

K-Means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*/kelompok. Metode ini mempartisi data ke dalam *cluster*/kelompok sehingga data yang memiliki karakteristik sama dikelompokkan ke dalam satu *cluster* yang sama [4].

2. METODE PENELITIAN

2.1. Clustering

Clustering adalah sebuah proses untuk mengelompokkan data ke dalam beberapa *cluster* atau kelompok sehingga data dalam satu *cluster* memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripan yang minimum. Analisis *cluster* (Clustering) merupakan usaha untuk mengidentifikasi kelompok obyek yang mirip-mirip dan membantu menemukan pola penyebaran dan pola hubungan dalam sekumpulan data yang besar [5].

2.2. Algoritma Klasifikasi K-Means

Algoritma *K-means* adalah implemen-tasi dari algoritma *clustering* partisional yang paling sederhana dan paling banyak digunakan. *K-means* juga menggunakan kuadrat *error criterion*. Algoritma ini mulai mempartisi ruang data secara acak sambil penunjukan (*assignment*) sampel yang ada ke dalam *kluster-kluster* berdasarkan kemiripan antara *kluster* dan sampel, sampai sebuah *criterion* yang *convergen*

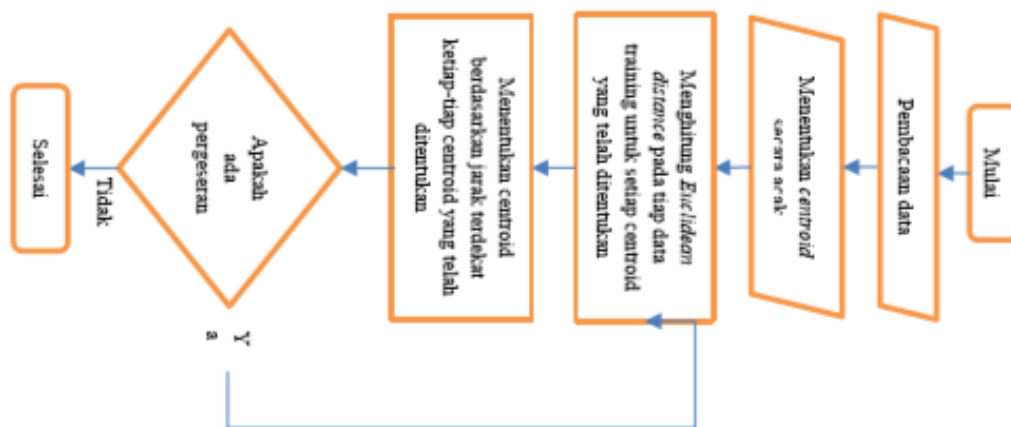
ditemukan. Syarat sebuah *criterion* telah ditemukan adalah ketika tidak ada lagi pemindahan sampel (*reassignment*) dari satu kluster ke kluster yang lain yang akan menyebabkan berkurangnya total *error* yang dikuadratkan (*errorsquare*). Algoritma ini populer digunakan karena kemudahan implementasinya, dan memiliki kecepatan yang cukup baik.

Tahapan-Tahapan dasar yang diambil oleh algoritma *K-means* adalah:

1. Tahap pertama: Tanyakan kepada pemakai algoritma *K-means*, catatan-catatan yang ada akan dibuat menjadi beberapa kelompok, sebutlah sebanyak *k* kelompok.
2. Tahap kedua: secara sembarang, pilih *k* buah catatan (dari sekian catatan yang ada) sebagai pusat-pusat kelompok awal.

3. Tahap ketiga: untuk setiap catatan, tentukan pusat kelompok terdekatnya dan tetapkan catatan tersebut sebagai anggota dari kelompok yang terdekat pusat kelompoknya. Hitung rasio antara besaran *Between Cluster Variation* dengan *Within Cluster Variation*, lalu bandingkan rasio tersebut dengan rasio sebelumnya(bila sudah ada). Jika rasio tersebut membesar, maka lanjutkan ke langkah keempat. Jika tidak hentikan prosesnya.
4. Tahap keempat: perbaharui pusat-pusat kelompok (berdasarkan kelompok yang didapat dari langkah ketiga) dan kembali ke langkah ketiga.

Flowchart mengenai Tahapan-Tahapan penerapan metode *K-Means* ditampilkan pada Gambar 1



Gambar 1. Flowchart K-Means Clustering

2.3. Kategori Usia

Menurut Departemen Kesehatan Republik Indonesia tahun 2009 kategori umur, yakni :

- a. Masa balita usia 0 – 5 tahun
- b. Masa kanak-kanak usia 5 – 11 tahun
- c. Masa remaja awal usia 12 – 16 tahun
- d. Masa remaja akhir usia 17 – 25 tahun
- e. Masa dewasa awal usia 26 – 35 tahun
- f. Masa dewasa akhir usia 36 – 45 tahun
- g. Masa lansia awal usia 46 – 55 tahun

- h. Masa lansia akhir usia 56 – 65 tahun
- i. Masa manula usia 65 – ke atas

3. HASIL DAN PEMBAHASAN

3.1. Data Pengujian

Pengelompokan data dilakukan dengan menggunakan *K-Means Clustering* dengan sampel data jumlah penduduk setiap kecamatan di kota Pematangsiantar yang dibagi dalam kategori umur menurut Depkes Republik Indonesia. yaitu : Balita, Kanak-Kanak, Remaja Awal, Remaja Akhir, Dewasa Awal, Dewasa Akhir, Lansia Awal, Lansia Akhir, dan Manula Atas

Tabel 1. Data Jumlah Penduduk per Kecamatan Kota Pematangsiantar tahun 2018

NO	Kecamatan	Balita (0 - 5 Thn)	Kanak- Kanak (5 - 11 Thn)	Remaja Awal (12 - 16 Thn)	Remaja Akhir (17 - 25 Thn)	Dewasa Awal (26 - 35 Thn)	Dewasa Akhir (36 - 45 Thn)	Lansia Awal (46 - 55 Thn)	Lansia Akhir (55 - 65 Thn)	Manula Atas (65 >= Thn)
1	Siantar Selatan	1,334	1,574	1,637	3,196	2,205	2,535	2,346	1,805	1,378
2	Siantar Barat	3,244	3,209	3,114	6,593	5,600	5,228	4,736	3,555	2,160
3	Siantar Marimbun	1,468	1,583	1,611	2,820	1,861	2,061	2,057	1,388	890
4	Siantar Marihat	1,741	1,930	1,969	3,559	2,396	2,466	2,328	1,743	1,125
5	Siantar Martoba	4,378	4,347	4,021	7,364	6,221	5,788	4,545	2,664	1,481
6	Siantar Sitalasari	2,717	2,826	2,625	5,229	4,106	4,161	3,594	2,255	1,246
7	Siantar Timur	3,240	3,613	3,550	9,527	5,347	5,305	4,448	3,193	2,319
8	Siantar Utara	4,599	4,648	4,485	9,387	7,094	6,552	5,776	3,907	2,502

3.2. Hasil Pelatihan

Tahapan-tahapan algoritma klasifikasi *K-Means* untuk mengelompokkan data yang ada.

1. Tahap Pertama dari *K-Means clustering* adalah pemakai akan menentukan jumlah kelompok yang akan dibuat dalam pengelompokan data. Sebagai contoh dalam penyelesaian kasus ini kita akan membagi dalam 3 kelompok atau $k = 3$.

2. Tahap Kedua. secara sembarang memilih 3 record dari data yang akan kita uji sebagai pusat data awal setiap kelompok. misalnya

a. Record ke-1 sebagai Cluster 1 sehingga $C1 = \{1.334, 1.574, 1.637, 3.196, 2.205, 2.535, 2.346, 1.805, 1.378\}$

b. Record ke-3 sebagai Cluster 2 sehingga $C2 = \{1.468, 1.583, 1.611, 2.820, 1.861, 2.061, 2.057, 1.388, 890\}$

c. Record ke-5 sebagai Cluster 3 sehingga $C3 = \{4.378, 4.347, 4.021, 7.364, 6.221, 5.788, 4.545, 2.664, 1.481\}$

3. Tahap ketiga. setiap record akan ditentukan pusat kelompok terdekatnya. Record tersebut akan ditetapkan sebagai kelompok yang

terdekat pusat kelompoknya. Hitung jarak setiap data yang ada terhadap setiap pusat cluster dengan persamaan berikut:

$$D(x,y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2}$$

Cluster 1 (C1):

$$\begin{aligned} D_{11} &= \sqrt{(1.334 - 1.334)^2 + (1.574 - 1.574)^2} \\ &+ (1.637 - 1.637)^2 + (3.196 - 3.196)^2 \\ &+ (2.205 - 2.205)^2 \\ &+ (2.535 - 2.535)^2 + (2.346 - 2.346)^2 \\ &+ (1.805 - 1.805)^2 + (1.378 - 1.378)^2 \\ D_{11} &= 0,00 \end{aligned}$$

$$\begin{aligned} D_{12} &= \sqrt{(3.244 - 1.334)^2 + (3.209 - 1.574)^2} \\ &+ (3.114 - 1.637)^2 + (6.593 - 3.196)^2 \\ &+ (5.600 - 2.205)^2 \\ &+ (5.228 - 2.535)^2 + (4.736 - 2.346)^2 \\ &+ (3.555 - 1.805)^2 + (2.160 - 1.378)^2 \\ D_{12} &= 6.943,12 \end{aligned}$$

Cluster 2 (C2):

$$\begin{aligned} D_{11} &= \sqrt{(1.334 - 1.468)^2 + (1.574 - 1.583)^2} \\ &+ (1.637 - 1.611)^2 + (3.196 - 2.820)^2 \\ &+ (2.205 - 1.861)^2 \\ &+ (2.535 - 2.061)^2 + (2.346 - 2.057)^2 \\ &+ (1.805 - 1.388)^2 + (1.378 - 890)^2 \\ D_{11} &= 999,33 \end{aligned}$$

$$D_{12} = \sqrt{(3.244 - 1.468)^2 + (3.209 - 1583)^2 + (3.114 - 1.611)^2 + (6.593 - 2.820)^2 + (5.600 - 1.861)^2 + (5.228 - 2.061)^2 + (4.736 - 2.057)^2 + (3.555 - 1.388)^2 + (2.160 - 890)^2}$$

$$D_{12} = 7.737,30$$

$$D_{11} = 8.503,26$$

$$D_{12} = \sqrt{(3.244 - 4.378)^2 + (3.209 - 4.347)^2 + (3.114 - 4.021)^2 + (6.593 - 7.364)^2 + (5.600 - 6.221)^2 + (5.228 - 5.788)^2 + (4.736 - 4.545)^2 + (3.555 - 2.664)^2 + (2.160 - 1.481)^2}$$

$$D_{12} = 2.447,19$$

Cluster 3 (C3):

$$D_{11} = \sqrt{(1.334 - 4.378)^2 + (1.574 - 4.347)^2 + (1.637 - 4.021)^2 + (3.196 - 7.364)^2 + (2.205 - 6.221)^2 + (2.535 - 5.788)^2 + (2.346 - 4.545)^2 + (1.805 - 2.664)^2 + (1.378 - 1.481)^2}$$

Berikut Hasil perhitungan selengkapnya pada Tabel 2.

Tabel 2. Hasil Perhitungan Jarak Setiap Kecamatan

NO	Kecamatan	Balita (0 - 5 Thn)	Kanak- Kanak (5 - 11 Thn)	Remaja Awal (12 - 16 Thn)	Remaja Akhir (17 - 25 Thn)	Dewasa Awal (26 - 35 Thn)	Dewasa Akhir (36 - 45 Thn)	Lansia Awal (46 - 55 Thn)	Lansia Akhir (55 - 65 Thn)	Manula Atas (65 >= Thn)	C1	C2	C3
2	Siantar Barat	3.244	3.209	3.114	6.593	5.600	5.228	4.736	3.555	2.160	6.943,12	7.732,30	2.447,19
3	Siantar Marimbun	1.468	1.583	1.611	2.820	1.861	2.061	2.057	1.388	890	999,33	0,00	9.145,03
4	Siantar Marihat	1.741	1.930	1.969	3.559	2.396	2.466	2.328	1.743	1.125	802,37	1.254,66	7.939,82
5	Siantar Martoba	4.378	4.347	4.021	7.364	6.221	5.788	4.545	2.664	1.481	8.503,26	9.145,03	0,00
6	Siantar Sitalasari	2.717	2.826	2.625	5.229	4.106	4.161	3.594	2.255	1.246	4.077,32	4.756,82	4.452,71
7	Siantar Timur	3.240	3.613	3.550	9.527	5.347	5.305	4.448	3.193	2.319	8.735,39	9.470,86	2.953,45
8	Siantar Utara	4.599	4.648	4.485	9.387	7.094	6.552	5.776	3.907	2.502	11.137,08	11.870,59	3.145,78

Setiap record akan menjadi anggota dari suatu cluster yang memiliki jarak terkecil dari pusat cluster-nya. Misalnya untuk record pertama, jarak terkecil diperoleh pada cluster C1, maka data pertama akan menjadi anggota dari cluster

C1. Demikian juga untuk data kedua, jarak terkecil ada pada cluster C3, maka data tersebut akan masuk pada cluster C3. Posisi cluster selengkapnya dapat dilihat pada Tabel 3.

Tabel 3. Posisi Cluster Pada Iterasi Pertama

NO	Kecamatan	Balita (0 - 5 Thn)	Kanak- Kanak (5 - 11 Thn)	Remaja Awal (12 - 16 Thn)	Remaja Akhir (17 - 25 Thn)	Dewasa Awal (26 - 35 Thn)	Dewasa Akhir (36 - 45 Thn)	Lansia Awal (46 - 55 Thn)	Lansia Akhir (55 - 65 Thn)	Manula Atas (65 >= Thn)	C1	C2	C3	JTK
2	Siantar Barat	3.244	3.209	3.114	6.593	5.600	5.228	4.736	3.555	2.160	6.943,12	7.732,30	2.447,19	C3
3	Siantar Marimbun	1.468	1.583	1.611	2.820	1.861	2.061	2.057	1.388	890	999,33	0,00	9.145,03	C2

4	Siantar Marihah	1.741	1.930	1.969	3.559	2.396	2.466	2.328	1.743	1.125	802,37	1.254,66	7.939,82	C1
5	Siantar Martoba	4.378	4.347	4.021	7.364	6.221	5.788	4.545	2.664	1.481	8.503,26	9.145,03	0,00	C3
6	Siantar Sitalasari	2.717	2.826	2.625	5.229	4.106	4.161	3.594	2.255	1.246	4.077,32	4.756,82	4.452,71	C1
7	Siantar Timur	3.240	3.613	3.550	9.527	5.347	5.305	4.448	3.193	2.319	8.735,39	9.470,86	2.953,45	C3
8	Siantar Utara	4.599	4.648	4.485	9.387	7.094	6.552	5.776	3.907	2.502	11.137,08	11.870,59	3.145,78	C3

Keterangan

JTK : Jarak Terdekat ke Kelompok

Dari Tabel 3 didapat keanggotaan *Cluster* sebagai berikut:

1. Kelompok 1 (atau C1) = { 1, 4, 6 }
2. Kelompok 2 (atau C2) = { 3 }
3. Kelompok 3 (atau C3) = { 2, 5, 7, 8 }

Pada tahap ini akan mengitung rasio antara besaran *Between Cluster Variation (BCV)* dengan *Within Cluster Variation (WCV)* sebagai berikut:

$$BCV = D(C1.C2)+D(C1.C3)+D(C2.C3)$$

$$BCV = (999,33-0,00)+(8.503,26-0,00)+(9.145,03-0,00)$$

$$BCV = 18.647,62$$

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

$$WCV = 0,00^2 + 2.447,19^2 + 0,00^2 + 802,37^2 + 0,00^2 + 4.077,32^2 + 2.953,45^2 + 3.145,78^2$$

$$WCV = 41.875.874,00$$

$$BCV/WCV = 18.647,62/41.875.874,00= 0.00044$$

Berhubung pada tahapan sebelumnya belum mendapat rasio ini, maka perbandingan rasio belum dapat dilakukan dan proses *K-Means clustering*. Maka selanjutnya *tahap* keempat.

4. Tahap keempat (Iterasi-2)

Pada tahap ini. perubahan pusat-pusat kelompok akan dilakukan seperti berikut:

- a. Untuk *Cluster 1*, ada 3 data yaitu record ke-1, 4, dan 6. sehingga:
 $C_{11} = (1.334 + 1.741 + 2.717)/3 = 5.792$
 $C_{12} = (1.574 + 1.930 + 2.826)/3 = 6.330$
 $C_{13} = (1.637 + 1.969 + 2.625)/3 = 6.231$
 $C_{14} = (3.196 + 3.559 + 5.229)/3 = 11.984$
 $C_{15} = (2.205 + 2.396 + 4.106)/3 = 8.707$
 $C_{16} = (2.535 + 2.466 + 5.305)/3 = 9.162$

$$C_{17} = (2.346 + 2.328 + 3.594)/3 = 8.268$$

$$C_{18} = (1.805 + 1.743 + 2.255)/3 = 5.803$$

$$C_{19} = (1.378 + 1.125 + 1.246)/3 = 3.749$$

Sehingga C1 = {5.792, 6.330, 6.231, 11.984, 8.707, 9.162, 8.268, 5.803, 3.749}

- b. Untuk *Cluster 2*, ada 1 record, yaitu record ke-3, sehingga:
 Sehingga C2 = {1.468, 1.583, 1.611, 2.820, 1.861, 2.061, 2.057, 1.388, 890}
 - c. Untuk *Cluster 3*, ada 4 record, yaitu record ke- 2, 5, 7, dan 8. sehingga:
 $C_{31} = (3.244 + 4.378 + 3.240 + 4.599)/4 = 15.461$
 $C_{32} = (3.209 + 4.347 + 3.613 + 4.648)/4 = 15.817$
 $C_{33} = (3.114 + 4.021 + 3.550 + 4.485)/4 = 15.170$
 $C_{34} = (6.593 + 7.364 + 9.527 + 9.387)/4 = 32.871$
 $C_{35} = (5.600 + 6.221 + 5.347 + 7.094)/4 = 24.262$
 $C_{36} = (5.228 + 5.788 + 5.305 + 6.552)/4 = 22.873$
 $C_{37} = (4.736 + 4.545 + 4.448 + 5.776)/4 = 19.505$
 $C_{38} = (3.555 + 2.664 + 3.193 + 3.907)/4 = 13.319$
 $C_{39} = (2.160 + 1.481 + 2.319 + 2.502)/4 = 8.462$
 Sehingga C3 = {15.461, 15.817, 15.170, 32.871, 24.262, 22.873, 19.505, 13.319, 8.462}
5. Ulangi tahap ke-3 hingga nilai rasio antara besaran *Between Cluster Variation (BCV)* dengan *Within Cluster Variation (WCV)* tidak mengalami kenaikan dari nilai rasio sebelumnya.

Tabel 4. Posisi Cluster Pada Iterasi Kedua

NO	Kecamatan	Balita	Kanak-Kanak	Remaja Awal	Remaja Akhir	Dewasa Awal	Dewasa Akhir	Lansia Awal	Lansia Akhir	Manula Atas	C1	C2	C3	JTK
		(0 - 5 Thn)	(5 - 11 Thn)	(12 - 16 Thn)	(17 - 25 Thn)	(26 - 35 Thn)	(36 - 45 Thn)	(46 - 55 Thn)	(55 - 65 Thn)	(65 >= Thn)				
1	Siantar Selatan	1.334	1.574	1.637	3.196	2.205	2.535	2.346	1.805	1.378	16.843,28	999,33	53.324,62	C2
2	Siantar Barat	3.244	3.209	3.114	6.593	5.600	5.228	4.736	3.555	2.160	10.009,73	7.732,30	46.442,31	C2
3	Siantar Marimbun	1.468	1.583	1.611	2.820	1.861	2.061	2.057	1.388	890	17.609,34	0,00	54.077,54	C2
4	Siantar Marihat	1.741	1.930	1.969	3.559	2.396	2.466	2.328	1.743	1.125	16.377,84	1.254,66	52.839,37	C2
5	Siantar Martoba	4.378	4.347	4.021	7.364	6.221	5.788	4.545	2.664	1.481	8.864,66	9.145,03	45.085,93	C1
6	Siantar Sitalasari	2.717	2.826	2.625	5.229	4.106	4.161	3.594	2.255	1.246	12.932,35	4.756,82	49.362,73	C2
7	Siantar Timur	3.240	3.613	3.550	9.527	5.347	5.305	4.448	3.193	2.319	8.759,84	9.470,86	44.851,47	C1
8	Siantar Utara	4.599	4.648	4.485	9.387	7.094	6.552	5.776	3.907	2.502	5.900,77	11.870,59	42.214,99	C1

Dari Tabel 4 didapat keanggotaan *Cluster* sebagai berikut:

1. Kelompok 1 (atau C1) = { 5, 7, 8 }
2. Kelompok 2 (atau C2) = { 1, 2, 3, 4, 6 }
3. Kelompok 3 (atau C3) = { }

Pada tahap ini akan menghitung rasio antara besaran *Between Cluster Variation (BCV)* dengan *Within Cluster Variation (WCV)*. sebagai berikut:

$$BCV = D(C1.C2)+D(C1.C3)+D(C2.C3)$$

$$BCV = (17.609,34-0,00)+(0)+(0)$$

$$BCV = 17.609,34$$

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

$$WCV = 999,33^2 + 7.732,30^2 + 0,00^2 + 1.254,66^2$$

$$+ 8.864,66^2 + 4.756,82^2 + 8.759,84^2 + 5.900,77^2$$

$$WCV = 275.124.634,00$$

$$BCV/WCV = 17.609,34/275.124.634,00= 0.000064$$

Karena pada iterasi-2 nilai rasio antara besaran *Between Cluster Variation (BCV)* dengan *Within Cluster Variation (WCV)* tidak mengalami kenaikan dari nilai rasio sebelumnya (pada iterasi-1). maka iterasi di hentikan dan hasil akhir yang diperoleh ada 2 *cluster*:

- a) *Cluster* pertama memiliki pusat (4.072, 4.203, 4.019, 8.759, 6.221, 5.882, 4.923, 3.255, 2.101) yang dapat

diartikan sebagai kelompok kecamatan dimana jumlah penduduk setiap kategori usia tinggi.

- b) *Cluster* kedua memiliki pusat (2.101, 2.224, 2.191, 4.279, 3.234, 3.290, 3.012, 2.149, 1.360) yang dapat diartikan sebagai kelompok kecamatan dimana jumlah penduduk setiap kategori usia rendah.

4. KESIMPULAN

Dari hasil penelitian, dapat disimpulkan bahwa algoritma klasifikasi *K-Means* dapat digunakan untuk mengelompokkan setiap kecamatan kota Pematangsiantar berdasarkan usia penduduk berdasarkan kategori, yaitu: Balita, Kanak-Kanak, Remaja Awal, Remaja Akhir, Dewasa Awal, Dewasa Akhir, Lansia Awal, Lansia Akhir, Manula Atas. Dari data yang diuji, diperoleh 2 kelompok, yaitu :

- 1) Kelompok Kecamatan yang jumlah penduduknya tinggi untuk setiap kategori usia adalah Kecamatan siantar Martoba, Siantar Timur, dan Siantar utara dengan pusat Cluster (4.072, 4.203, 4.019, 8.759, 6.221, 5.882, 4.923, 3.255, 2.101).
- 2) Kelompok Kecamatan yang jumlah penduduknya rendah untuk setiap kategori usia adalah Kecamatan Siantar Selatan, Siantar Barat, Siantar Marimbun, Siantar Marihat, dan Siantar Sitalasari dengan pusat Cluster (2.101, 2.224, 2.191, 4.279, 3.234, 3.290, 3.012, 2.149, 1.360).

5. REFERENSI

- [1] S. Suryadi and D. Irmayani, "Penerapan Metode Clustering K-Means Untuk Pengelompokan Kelulusan Mahasiswa," *Inform. J. Ilm. AMIK Labuhan Batu*, vol. 6, no. 1, pp. 52–72, 2018.
- [2] F. S. Jumeilah, D. Pratama, and S. Penduduk, "MENGUNAKAN ALGORITMA K-MODES," *J. TAM (Technology Accept. Model.*, vol. 8, no. 2, pp. 85–89, 2017.
- [3] W. Purba, S. Tamba, and J. Saragih, "The effect of mining data k-means clustering toward students profile model drop out potential," *J. Phys. Conf. Ser.*, vol. 1007, no. 1, p. 12049, 2018.
- [4] S. P. Tamba, M. D. Batubara, W. Purba, M. Sihombing, V. M. Mulia Siregar, and J. Banjarnahor, "Book data grouping in libraries using the k-means clustering method," *J. Phys. Conf. Ser.*, vol. 1230, no. 1, 2019.
- [5] L. Rahmawati, S. W. Sihwi, and E. Suryani, "ANALISA CLUSTERING MENGGUNAKAN METODE K-MEANS DAN HIERARCHICAL CLUSTERING (STUDI KASUS : DOKUMEN SKRIPSI JURUSAN KIMIA , FMIPA , 2 . 3 Term Weighting dengan Term Frequency."