

PENERAPAN DATA MINING UNTUK *CLUSTERING* DATA PENDUDUK MISKIN MENGGUNAKAN ALGORITMA *K-MEANS*

*Dina Sunia*¹, *Kurniabudi*², *Pareza Alam Jusia*³

*dinasunia07@gmail.com*¹, Teknik Informatika, STIKOM Dinamika Bangsa, Jambi dan 36138, Indonesia
*kbudiz@yahoo.com*², Teknik Informatika, STIKOM Dinamika Bangsa, Jambi dan 36138, Indonesia
*parezaalam@gmail.com*³, Informatika, STIKOM Dinamika Bangsa, Jambi dan 36138, Indonesia

Abstrack

Source of data from BPS in Jambi City In March 2017 the number of poor people was 286.55 thousand people (8.19%), a decrease of 4.26 thousand people compared to the poor population in September 2016 which amounted to 290.81 thousand people (8.37 %). Based on these figures, there is still a high rate of poverty in the Jambi City area in general. Based on these conditions it is necessary to do clusters to help BKKBN in grouping poor families so that assistance can be channeled appropriately. By observing the above problems, Data Mining is needed to classify beneficiaries by using the K-Means method in the clustering of the poor. The calculation results show, from 512 data samples of poor population in South Jambi Subdistrict, with a total cluster of 5, with an iteration of 3 obtained 13 residents entered as cluster members 1, 153 residents were in cluster 2, 129 residents in cluster 3, 138 residents were in cluster 4, and 79 residents are in cluster 5.

Keywords: Data Mining, poor population, K-means

Abstrak

Sumber data dari BPS kota jambi Pada Maret 2017 jumlah penduduk miskin sebesar 286,55 ribu jiwa (8,19%), berkurang sebesar 4,26 ribu jiwa dibandingkan dengan penduduk miskin pada september 2016 yang sebesar 290,81 ribu jiwa (8,37%). Berdasarkan angka tersebut terlihat masih tingginya angka kemiskinan yang ada di wilayah Kota Jambi secara umum. Berdasarkan kondisi tersebut perlu dilakukan pengklasteran untuk membantu Dinas Sosial Kota Jambi dalam pengelompokan keluarga miskin sehingga bantuan dapat tersalurkan dengan tepat. Dengan mengamati persoalan diatas maka diperlukan *Data Mining* untuk mengelompokkan penerima bantuan dengan menggunakan metode *K-Means* dalam pengklasteran penduduk miskin. Hasil perhitungan menunjukkan, dari 512 sample data penduduk miskin di Kecamatan Jambi Selatan, dengan jumlah klaster sebanyak 5, dengan iterasi sebanyak 3 diperoleh 13 penduduk masuk sebagai anggota klaster 1, 153 penduduk berada di klaster 2, 129 penduduk di klaster 3, 138 penduduk berada di klaster 4, dan 79 penduduk berada di klaster 5.

Kata Kunci: *Data Mining*, penduduk miskin, *K-means*

1. LATAR BELAKANG

Bagi pemerintah Indonesia masalah kemiskinan merupakan masalah lama yang belum dan sulit untuk diselesaikan. Pemerintah sendiri telah melakukan beberapa upaya dalam melakukan pengentasan kemiskinan diantaranya melalui program bantuan sosial diantaranya Bantuan langsung tunai (BLT), Program Keluarga Harapan (PKH) dll. Berdasarkan hasil survey yang dilakukan bantuan yang diberikan pemerintah kepada penduduk tidak sesuai dengan apa yang mereka butuhkan. Satu kesulitan yang terkadang dihadapi oleh pemerintah dalam proses penanganan kemiskinan adalah proses pembagian bantuan sosial yang tidak merata dan tidak tepat sasaran. Ini disebabkan karena validasi data sering diabaikan sehingga menimbulkan data yang tidak akurat. Demi keadilan, penyaluran bantuan itu semestinya dibatalkan, dan harus di tunda, sampai data soal siapa yang berhak menerima bantuan itu benar-benar valid. Kecamatan jambi selatan digunakan sebagai sampel karena pada kecamatan ini dengan jumlah penduduk terbanyak setelah kota baru. Dengan data yang banyak maka Teknik data mining yang lebih tepat untuk mengolah data tersebut. *Clustering* bukanlah hal yang baru berdasarkan penelitian yang dilakukan Zainul Aras Z & Sarjono (2016) membahas tentang metode *Clustering* dan mereka menggunakan algoritma K-

Means untuk menentukan prioritas penerima bantuan bedah rumah, metode *Clustering K-Means* digunakan agar penelitian prioritas penduduk tidak mampu bisa lebih berkualitas dan efektif. Rizky Dwi Setiawan dkk (2016) mereka melakukan perhitungan pemetaan tingkat kemiskinan menggunakan algoritma *K-Means*, maka di dapat rentang dari setiap kategori tingkat kesejahteraan berdasarkan hasil *Clustering*. Adapun rentang dari tingkat kesejahteraan memperoleh 3 kelompok, kelompok pertama yaitu Hampir Miskin dengan rentang nilai <2.117 , kelompok kedua yaitu Miskin dengan rentang nilai $\geq 2.117 - <2.807$, dan kelompok ketiga Sangat Miskin ≥ 2.807 . Berdasarkan beberapa kajian penelitian, algoritma *K-Means* digunakan untuk mengklasterkan penilaian dosen berdasarkan indeks kepuasan mahasiswa, untuk cluster_baik 17.099 dan cluster_kurang 15.874. sehingga di peroleh penilaian dosen berdasarkan indek kepuasan mahasiswa dengan 5 dosen cluster_baik dan 7 dosen cluster_kurang

2. LANDASAN TEORI

2.1 Konsep *Data Mining*

Data adalah sebagai bahan keterangan tentang kejadian nyata atau fakta-fakta yang dirumuskan dalam sekelompok lambang tertentu yang tidak acak yang menunjukkan jumlah, tindakan, atau hal^[1]. Data adalah fakta dari suatu pernyataan yang berasal dari kenyataan, dimana pernyataan tersebut merupakan hasil pengukuran atau pgamatan^[2]. Jadi dapat disimpulkan bahwa data adalah sekumpulan fakta-fakta hasil pengukuran dan pengamatan yang belum memiliki arti dan dapat diolah menjadi suatu informasi tertentu. *Data Mining* atau penambangan data dapat didefinisikan sebagai proses seleksi, eksplorasi, dan pemodelan dari sejumlah besar data untuk menemukan pola atau kecenderungan yang biasanya tidak disadari keberadaannya^[3]. *Data Mining* bukanlah suatu bidang yang sama sekali baru. Salah satu kesulitan untuk mendefinisikan *Data Mining* adalah kenyataan bahwa *Data Mining* mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dahulu^[4]. *Data Mining* merupakan disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data^[5]. Jadi dapat disimpulkan bahwa *Data Mining* adalah proses penggalian data secara mendalam untuk mengetahui hal yang berarti dan tidak diketahui keberadaannya. *Clustering* merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data yang lain^[5]. *Clustering* adalah mengelompokkan item data ke dalam sejumlah kecil grup sedemikian sehingga masing-masing grup mempunyai sesuatu persamaan yang esensial^[6]. *Clustering* merupakan salah satu metode *Data Mining* yang melakukan pemisahan/pemecahan/segmentasi data ke dalam sejumlah kelompok (*Cluster*) menurut karakteristik tertentu yang diinginkan, dalam pekerjaan pengelompokan, label dari setiap data belum diketahui dan dengan pengelompokan diharapkan dapat diketahui kelompok data untuk kemudian diberi label sesuai keinginan^[7]. Jadi dapat disimpulkan bahwa *Clustering* adalah metode untuk mengelompokkan data yang memiliki kemiripan dan kemudian di beri label sesuai keinginan. Algoritma *K-Means Clustering* adalah algoritma sederhana dan efektif untuk menemukan *Cluster* dalam data dengan algoritma sebagai berikut:

Langkah 1 : Tentukan jumlah *Cluster*

Langkah 2 : Tentukan nilai yang menjadi lokasi pusat *Cluster* awal.

Langkah 3 : Hitung pusat *Cluster* terdekat untuk setiap record.

Langkah 4 : Untuk setiap *Cluster* k , hitung centroid *Cluster* dan memperbarui lokasi setiap pusat *Cluster*

Langkah 5 : Ulangi langkah 3 sampai 5 sampai konvergensi atau penghentian^[8]. Berdasarkan referensi dari penelitian-penelitian sebelumnya yang menggunakan algoritma *K-Means*, algoritma ini memiliki tingkat akurasi yang lebih tinggi. Penelitian yang dilakukan Gusti Ngurah Wisnu Paramartha, dkk (2017) mereka memperoleh tingkat akurasi algoritma *K-Means* sebesar 90,3%, sedangkan penelitian yang dilakukan oleh Andi Sri Irtawaty (2017) dia memperoleh tingkat keakuratan metode ini yaitu sebesar 90%.

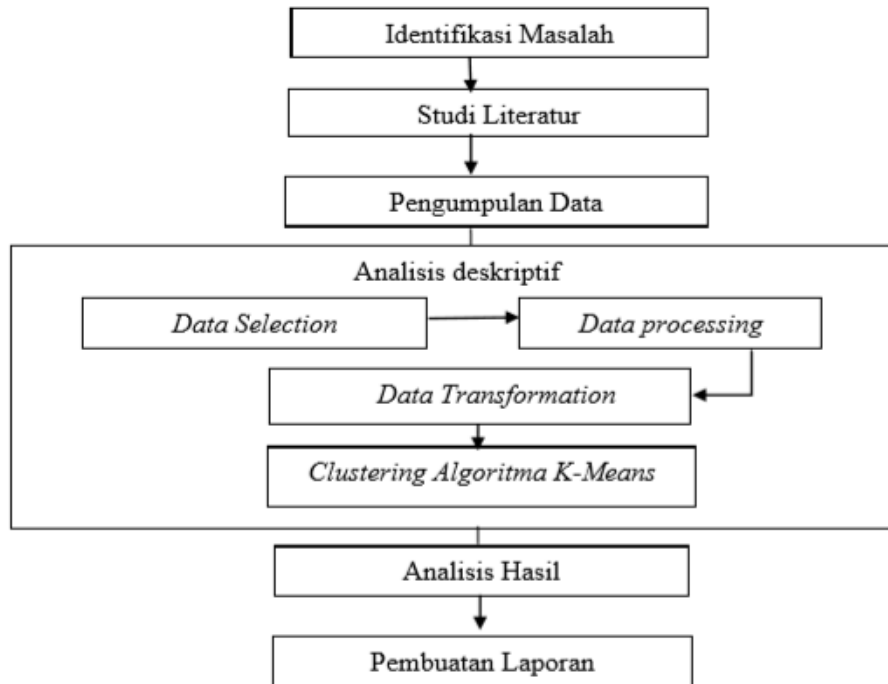
2.4 PENDUDUK MISKIN

Menurut Hasya Yanto mengatakan: “kemiskinan adalah keadaan dimana seseorang yang tidak bisa memenuhi kebutuhan hidupnya”. Kemiskinan adalah keadaan dimana terjadi ketidakmampuan untuk memenuhi kebutuhan dasar seperti makanan, pakaian, tempat berlindung, pendidikan, dan kesehatan^[9]. Kemiskinan secara konseptual dibedakan menurut kemiskinan relatif dan kemiskinan absolut, dimana perbedaannya terletak pada standar penilaiannya. Standar penilaian kemiskinan relatif merupakan standar kehidupan yang ditentukan dan ditetapkan secara subyektif oleh masyarakat setempat dan bersifat lokal serta mereka yang berada dibawah standar penilaian tersebut dikategorikan sebagai miskin secara relatif. Sedangkan standar penilaian kemiskinan secara absolut merupakan standar kehidupan minimum yang dibutuhkan untuk memenuhi kebutuhan dasar yang diperlukan, baik makanan maupun non makanan. Standar kehidupan minimum untuk memenuhi kebutuhan dasar ini disebut sebagai garis kemiskinan^[11].

Jadi dapat disimpulkan bahwa kemiskinan dimana seseorang tidak dapat memenuhi kebutuhan hidupnya seperti kebutuhan makanan, pakaian, tempat tinggal, pendidikan dan kesehatan.

3. METODE PENELITIAN

Kerangka kerja penelitian menggambarkan tahapan-tahapan yang dilakukan untuk mencapai tujuan penelitian. Kerangka kerja penelitian dibuat agar mempermudah pencapaian hasil penelitian, dapat menyelesaikan penelitian tepat waktu dan penelitian dapat berjalan sesuai dengan yang diharapkan. Adapun kerangka kerja penelitian yang digunakan dapat dilihat pada gambar 3.1.



Gambar 3.1 Kerangka Penelitian

Berikut ini merupakan penjelasan dari setiap bagian yang tercantum pada kerangka kerja penelitian dalam gambar 3.1.

3.1 Identifikasi Masalah

Pada tahap ini, penulis melakukan identifikasi dan merumuskan permasalahan pada penelitian, hal ini bertujuan untuk mengetahui masalah yang dialami di Dinas Sosial yaitu dalam memberikan bantuan kepada masyarakat miskin agar bantuan tersebut tepat sasaran.

3.2 Studi Literatur

Pada tahapan ini peneliti melakukan kajian pustaka, yaitu mempelajari buku-buku referensi, artikel-artikel, dan hasil penelitian sejenis yang relevan dengan permasalahan yang sedang diteliti. Studi literatur ini bertujuan untuk mendapatkan landasan teoritis mengenai permasalahan yang akan diteliti seperti teknik *data mining*, proses *Clustering* menggunakan algoritma *K-Means*, *tool data mining* yang akan digunakan, data penduduk miskin Kecamatan Jambi Selatan. Hal ini bertujuan agar peneliti dapat memahami permasalahan yang diteliti dengan benar dan sesuai dengan pembahasan yang dilakukan.

3.3 Pengumpulan Data

Pada tahap ini peneliti melakukan pengumpulan data yang dibutuhkan dalam penelitian yaitu data penduduk miskin Kecamatan Jambi Selatan.

Untuk mengumpulkan data dalam kegiatan penelitian ini penulis menggunakan beberapa metode pengumpulan data, yaitu :

- a. Pengamatan (Observasi)

Pengamatan (observasi) yang dimaksud dalam penelitian ini adalah melakukan penelitian secara langsung dengan tujuan untuk lebih memahami dan mengetahui langkah-langkah apa saja yang harus diambil dalam menyelesaikan permasalahan yang ditemukan.

b. Wawancara

Merupakan pengumpulan data dengan cara datang ke lokasi penelitian untuk melakukan tanya jawab langsung dengan pihak Dinas Sosial Kota Jambi, untuk memperoleh informasi yang dibutuhkan.

c. Penelitian Kepustakaan

Untuk menunjang penelitian yang akan dilakukan, penulis melakukan *Library research* yakni dengan mencari data-data dari buku maupun jurnal penelitian sejenis yang berhubungan dengan metode yang penulis gunakan dalam penelitian iniyaitu metode *Clustering K-Means*.

3.4 Analisis Deskriptif

Teknik *data mining* yang termasuk deskriptif mining adalah *Clustering*, *Asossiation* dan *sequential mining*. Proses analisis deskriptif dilakukan dengan beberapa tahapan, berikut ini penjelasan mengenai alur proses Analisis Deskriptif :

a. *Data Selection*

Pemilihan (Seleksi) data baru sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam *data mining* dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.

b. *Data Processing*

Pada tahap ini peneliti melakukan persiapan data penduduk miskin tahun 2015. Setelah itu peneliti melakukan seleksi atribut yang berpengaruh dalam *Clustering* untuk menentukan cluster penduduk miskin Kota Jambi.

c. *Data Transformation*

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*.

d. *Clustering Algoritma K-Means*

Pada tahap ini penulis menganalisis hasil dari perhitungan metode *k-means clusterin*, hasil analisis perhitungan penulis dengan hasil perhitungan menggunakan *tools WEKA*.

3.5 Analisis Hasil

Hasil dari analisis akan dinilai keakuratan data model yang sudah didapatkan dengan perhitungan *K-Means* yang populer dan akan mengelola data penduduk mengenai kemiskinan.

3.6 Penulisan Laporan

Pada tahap pembuatan laporan akhir penelitian, dilakukan berdasarkan kerangka yang telah dirancang yang terdiri dari Pendahuluan, Landasan Teori, Metodologi penelitian, analisis, Implementasi dan Pengujian Sistem, dan penutup serta lampiran bukti hasil penelitian.

4. ANALISIS DAN HASIL

4.1 PERHITUNGAN *K-MEANS* PADA *CLUSTERING* KEMISKINAN PROVINSI JAMBI

Pada perhitungan *K-Means Clustering* kemiskinan penduduk provinsi jambi akan membahas mulai dari representasi data, variable data yang digunakan, data penduduk kecamatan Jambi Selatan, dan Perhitungan secara manual mengenai *Clustering K-Means*.

1. Representasi Data

Berdasarkan hasil wawancara dan observasi yang sudah dilakukan, penulis memperoleh data-data penduduk kecamatan jambi selatan di Dinas Sosial Kota Jambi. Data penduduk keseluruhan dikecamatan Jambi Selatan adalah 2844 kepala keluarga terdiri dari 5 kelurahan yaitu Pasir Putih, Wijaya Pura, Tambak Sari, Thehok dan Pakuan Baru.

2. Variabel data

Pada tahap ini, variabel data harus ditentukan terlebih dahulu. Dimana variabel tersebut yang nantinya digunakan untuk mendapatkan informasi survey. Didalam buku pedoman penanganan fakir miskin perkotaan (2016) variabel yang digunakan ada pada Tabel 4.1, yang nanti akan digunakan untuk mengisi inisial nilai pada tabel 4.2

Tabel 4.1 Keterangan untuk mengisi variabel di tabel 4.2

No	Nama variabel/atribut	Keterangan
1	Anggota keluarga yang cacat	Jumlah anggota yang cacat
2	Lanjut usia	Jumlah anggota keluarga yang lanjut usia
3	Tanggungjawab dalam keluarga	Berapa orang yang menjadi tanggungan oleh kepala keluarga
4	Status Penguasaan Bangunan	0. tidak ada 1. sewa. 2. milik orang tua. 3. milik pemerintah. 4. milik bersama keluarga. 5. milik pribadi
5	Luas Bangunan (m ²)	Satuan meter
6	Jenis Lantai	0. kayu. 1. Semen. 2. kramik
7	Jenis Atap	0. ijuk. 1. Genteng. 2. Seng. 3. Asbes. 4. coran
8	Jenis Dinding	0. Tidak ada. 1. Kayu. 2. beton
9	Fasilitas Tempat Buang Air Besar	1. memiliki wc. 2. tidak memiliki wc
10	Tempat Pembuangan Akhir Tinja	1. tidak memiliki 2. memiliki
11	Sumber Air Minum	1. Sumur. 2. Pam
12	Sumber Penerangan Utama	1. Lilin. 2. listrik
13	Apakah mempunyai tempat tersendiri untuk dapur	1. Tidak 2. punya
14	Jenis Bahan Bakar untuk memasak	1. minyak tanah 2. gas
15	Lapangan Pekerjaan Utama Kepala Keluarga	0. tidak ada pekerjaan. 1. tidak tetap. 2. Pemulung. 3. tukang semir/penjahi sepatu keliling. 4. pedagang asongan jalan. 5. penjual makanan keliling. 6. penjaja mainan anak-anak. 7. tukang cukur jalan. 8. buruh lepas. 9. tukang ojek. 10. tukang bangunan. 11. pembantu rumah tangga. 12. Penjahit. 13. bekerja dengan orang. 14. usaha rumahan. 15. karyawan
16	Pendapatan Keluarga Per Bulan	Gaji perbulan

Tabel 4.2 TABEL INISIALISASI DATA

no	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	5	0	0	1	42	1	4	2	1	1	5	1	1	1	10	1400000
2	5	0	0	1	42	1	4	2	1	1	5	1	1	1	10	1400000
3	5	0	0	1	42	1	4	2	1	1	5	1	1	1	10	1400000
4	5	0	0	1	42	1	4	2	1	1	5	1	1	1	10	1400000
5	5	0	0	1	42	1	4	2	1	1	5	1	1	1	10	1400000
6	3	0	0	1	72	1	4	2	1	1	5	1	1	1	15	1000000
7	3	0	0	1	72	1	4	2	1	1	5	1	1	1	15	1000000
8	3	0	0	1	72	1	4	2	1	1	5	1	1	1	15	1000000
9	3	0	0	1	96	1	4	1	1	1	1	1	1	1	4	1200000
10	3	0	0	1	96	1	4	1	1	1	1	1	1	1	4	1200000
11	3	0	0	1	96	1	4	1	1	1	1	1	1	1	4	1200000
12	5	0	0	1	96	1	4	2	1	1	1	1	1	1	9	1600000
13	5	0	0	1	96	1	4	2	1	1	1	1	1	1	9	1600000
14	5	0	0	1	96	1	4	2	1	1	1	1	1	1	9	1600000
15	5	0	0	1	96	1	4	2	1	1	1	1	1	1	9	1600000
16	5	0	0	1	96	1	4	2	1	1	1	1	1	1	9	1600000
...
506	6	2	0	1	21	2	4	2	1	1	2	1	1	1	14	900000
507	6	2	0	1	21	2	4	2	1	1	2	1	1	1	14	900000
508	6	2	0	1	21	2	4	2	1	1	2	1	1	1	14	900000
509	6	2	0	1	21	2	4	2	1	1	2	1	1	1	14	900000
510	5	2	0	1	21	2	4	2	1	1	2	1	1	1	14	900000
511	5	2	0	1	21	2	4	2	1	1	2	1	1	1	14	900000
512	5	2	0	1	21	2	4	2	1	1	2	1	1	1	14	900000

3. Perhitungan Manual *K-Means Clustering*

Penulis melakukan perhitungan dengan menggunakan Persamaan untuk menghitung jarak antar data pada *K-Means* menggunakan rumus *Euclidian Distance (D)* yang ditunjukkan pada persamaan 2.3.

Berikut langkah-langkah *clustering* menggunakan algoritma *k-means* yaitu :

1. Langkah 1 : Tentukan jumlah *cluster* yang diinginkan (*cluster* = 5, karena mengacu pada metode *Lattern Class Cluster Analisis* menghasilkan 5 *cluster* untuk memperoleh banyak variasi).
2. Langkah 2 : Pilih *Centroid* awal secara acak. Pada langkah ini akan dipilih 5 buah data sebagai *Centroid*, data: 5,0,0,1,42,1,4,2,1,1,5,1,1,1,10,1400000
3. Langkah 3 : Hitung jarak dengan *centroid* (iterasi 1)

Pada langkah ini setiap data akan ditentukan *centroid* terdekatnya, dan data tersebut akan diterapkan sebagai anggota kelompok yang terdekat dengan *centroid*.

Untuk menghitung jarak ke *centroid* masing-masing *cluster* pada data penduduk No. 1 sebagai berikut :

Untuk menghitung jarak ke *centroid* masing-masing *Cluster* pada data penduduk sebagai berikut

Data: 5,0,0,1,42,1,4,2,1,1,5,1,1,1,10,1400000

Centroid m1=(5,0,0,1,96,1,4,2,1,1,1,1,1,1,9,1600000)

Centroid m2=(4,0,0,1,96,2,4,2,1,1,5,1,1,1,14,1250000)

Centroid m3=(5,0,0,1,48,1,0,0,1,1,0,1,1,1,0,800000)

Centroid m4=(5,0,0,1,35,1,0,1,1,1,0,1,1,1,0,900000)

Centroid m5=(7,0,0,2,40,1,4,2,1,1,4,1,1,1,9,1500000)

$$DM1 = \sqrt{\begin{matrix} (5-5)^2 + (0-0)^2 + (0-0)^2 + (1-1) + (42-96)^2 + \\ (1-1)^2 + (4-4)^2 + (2-2)^2 + (1-1)^2 + (1-1)^2 + \\ (5-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (10-9)^2 + \dots \end{matrix}} \quad (4.1)$$

$(1400000 - 1600000)^2 = 200000,0073$

$$DM2 = \sqrt{\begin{matrix} (5-4)^2 + (0-0)^2 + (0-0)^2 + (1-1) + (42-96)^2 + \\ (1-2)^2 + (4-4)^2 + (2-2)^2 + (1-1)^2 + (1-1)^2 + \\ (5-5)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (10-14)^2 + \dots \end{matrix}} \quad (4.2)$$

$(1400000 - 1250000)^2 = 150000,0098$

$$DM3 = \sqrt{\begin{matrix} (5-5)^2 + (0-0)^2 + (0-0)^2 + (1-1) + (42-48)^2 + \\ (1-1)^2 + (4-0)^2 + (2-0)^2 + (1-1)^2 + (1-1)^2 + \\ (5-0)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (10-0)^2 + \dots \end{matrix}} \quad (4.3)$$

$(1400000 - 800000)^2 = 600000,0002$

$$DM4 = \sqrt{\begin{matrix} (5-5)^2 + (0-0)^2 + (0-0)^2 + (1-1) + (42-35)^2 + \\ (1-1)^2 + (4-0)^2 + (2-1)^2 + (1-1)^2 + (1-1)^2 + \\ (5-0)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (10-0)^2 + \dots \end{matrix}} \quad (4.4)$$

$(1400000 - 900000)^2 = 500000,0002$

$$DM5 = \sqrt{\begin{matrix} (5-7)^2 + (0-0)^2 + (0-0)^2 + (1-2) + (42-40)^2 + \\ (1-1)^2 + (4-4)^2 + (2-2)^2 + (1-1)^2 + (1-1)^2 + \\ (5-4)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (10-9)^2 + \dots \end{matrix}} \quad (4.5)$$

$(1400000 - 1500000)^2 = 100000$

Dari hasil perhitungan diatas kemudian di tulis kedalam tabel 4.3

no	C1	C2	C3	C4	C5	jarak terkecil	e1	e2	e3	e4	e5
1	200000	150000	600000	500000	100000	100000,0001					v
2	200000	150000	600000	500000	100000	100000,0001					v
3	200000	150000	600000	500000	100000	100000,0001					v
4	200000	150000	600000	500000	100000	100000,0001					v
5	200000	150000	600000	500000	100000	100000,0001					v
6	600000	250000	200000	100000	500000	100000,0082					v
7	600000	250000	200000	100000	500000	100000,0082					v
8	600000	250000	200000	100000	500000	100000,0082					v
9	400000	50000	400000	300000	300000	50000,00119		v			
10	400000	50000	400000	300000	300000	50000,00119		v			
11	400000	50000	400000	300000	300000	50000,00119		v			
12	0	350000	800000	700000	100000	0	v				
13	0	350000	800000	700000	100000	0	v				
14	0	350000	800000	700000	100000	0	v				
15	0	350000	800000	700000	100000	0	v				
16	0	350000	800000	700000	100000	0	v				
...	
502	700000	350000	100000	20,4695	600000	20,46948949					v
503	700000	350000	100000	20,4695	600000	20,46948949					v
504	700000	350000	100000	20,4695	600000	20,46948949					v
505	700000	350000	100000	20,4695	600000	20,46948949					v
506	700000	350000	100000	20,4695	600000	20,46948949					v
507	700000	350000	100000	20,4695	600000	20,46948949					v
508	700000	350000	100000	20,4695	600000	20,46948949					v
509	700000	350000	100000	20,4695	600000	20,46948949					v
510	700000	350000	100000	20,4451	600000	20,4450483					v
511	700000	350000	100000	20,4451	600000	20,4450483					v
512	700000	350000	100000	20,4451	600000	20,4450483					v

Pada langkah ini dihitung pula rasio antara besaran BCV (*Between Cluster Variation*) dengan WCV (*within Cluster Variation*) :

Karena centroid m1,m2,m3,m4, dan m5 maka:

- Centroid m1=(5,0,0,1,96,1,4,2,1,1,1,1,1,9,1600000)
- Centroid m2=(4,0,0,1,96,2,4,2,1,1,5,1,1,14,1250000)
- Centroid m3=(5,0,0,1,48,1,0,0,1,1,0,1,1,1,0,800000)
- Centroid m4=(5,0,0,1,35,1,0,1,1,1,0,1,1,1,0,900000)
- Centroid m5=(7,0,0,2,40,1,4,2,1,1,4,1,1,1,9,1500000)

$$D(m1,m2)= \sqrt{\frac{(5-4)^2 + (0-0)^2 + (0-0)^2 + (1-1) + (96-96)^2 + (1-2)^2 + (4-4)^2 + (2-2)^2 + (1-1)^2 + (1-1)^2 + (1-5)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (9-14)^2 + \dots}{(1600000-1250000)^2 = 350000,0001}} \dots \dots \dots (4.6)$$

$$D(m1,m3)= \sqrt{\frac{(5-5)^2 + (0-0)^2 + (0-0)^2 + (1-1) + (96-48)^2 + (1-1)^2 + (4-0)^2 + (2-20) + (1-1)^2 + (1-1)^2 + (1-0)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (9-0)^2 + \dots}{(1600000-800000)^2 = 800000,0015}} \dots \dots \dots (4.7)$$

$$D(m1,m4)= \sqrt{\frac{(5-5)^2 + (0-0)^2 + (0-0)^2 + (1-1) + (96-35)^2 + (1-1)^2 + (4-0)^2 + (2-1) + (1-1)^2 + (1-1)^2 + (1-0)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (9-0)^2 + \dots}{(1600000-900000)^2 = 700000,0027}} \dots \dots \dots (4.8)$$

$$D(m1,m5)= \sqrt{\frac{(5-7)^2 + (0-0)^2 + (0-0)^2 + (1-2) + (96-40)^2 + (1-1)^2 + (4-4)^2 + (2-2) + (1-1)^2 + (1-1)^2 + (1-4)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (9-9)^2 + \dots}{(1600000-1500000)^2 = 100000,0157}} \dots \dots \dots (4.9)$$

$$D(m_2, m_3) = \sqrt{\begin{matrix} (4-5)^2 + (0-0)^2 + (0-0)^2 + (1-1) + (96-48)^2 + \\ (2-1)^2 + (4-0)^2 + (2-0) + (1-1)^2 + (1-1)^2 + \\ (5-0)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (14-0)^2 + \dots \end{matrix}} \quad (4.10)$$

$$(1600000 - 800000)^2 = 450000,0028$$

$$D(m_2, m_4) = \sqrt{\begin{matrix} (4-5)^2 + (0-0)^2 + (0-0)^2 + (1-1) + (96-35)^2 + \\ (2-1)^2 + (4-0)^2 + (2-1) + (1-1)^2 + (1-1)^2 + \\ (5-0)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (14-0)^2 + \dots \end{matrix}} \quad (4.11)$$

$$(1600000 - 900000)^2 = 350000,0057$$

$$D(m_2, m_5) = \sqrt{\begin{matrix} (4-7)^2 + (0-0)^2 + (0-0)^2 + (1-2) + (96-40)^2 + \\ (2-1)^2 + (4-4)^2 + (2-2) + (1-1)^2 + (1-1)^2 + \\ (5-4)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (14-9)^2 + \dots \end{matrix}} \quad (4.12)$$

$$(1600000 - 1250000)^2 = 250000,0063$$

$$D(m_3, m_4) = \sqrt{\begin{matrix} (5-5)^2 + (0-0)^2 + (0-0)^2 + (1-1) + (48-35)^2 + \\ (1-1)^2 + (0-0)^2 + (0-1) + (1-1)^2 + (1-1)^2 + \\ (0-0)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (0-0)^2 + \dots \end{matrix}} \quad (4.13)$$

$$(800000 - 900000)^2 = 100000,0009$$

$$D(m_3, m_5) = \sqrt{\begin{matrix} (5-7)^2 + (0-0)^2 + (0-0)^2 + (1-1) + (48-40)^2 + \\ (1-1)^2 + (0-4)^2 + (0-2) + (1-1)^2 + (1-1)^2 + \\ (0-4)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (0-9)^2 + \dots \end{matrix}} \quad (4.14)$$

$$(800000 - 1500000)^2 = 700000,0001$$

$$D(m_4, m_5) = \sqrt{\begin{matrix} (5-7)^2 + (0-0)^2 + (0-0)^2 + (1-1) + (35-40)^2 + \\ (1-1)^2 + (0-4)^2 + (1-2) + (1-1)^2 + (1-1)^2 + \\ (0-4)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (0-9)^2 + \dots \end{matrix}} \quad (4.15)$$

$$(900000 - 1500000)^2 = 600000,0001$$

$$\begin{aligned} BCV &= D(m_1, m_2) + D(m_1, m_3) + D(m_1, m_4) + D(m_1, m_5) + D(m_2, m_3) + D(m_2, m_4) + D(m_2, m_5) + D(m_3, m_4) + D(m_3, m_5) + D(m_4, m_5) \\ &= 350000,0001 + 800000,0015 + 700000,0027 + 100000,0157 + 450000,0028 + 350000,0057 + 250000,0063 + 100000,0009 + 700000,0001 + 600000,0001 = 4400000,036 \end{aligned}$$

Dalam hal ini $d(m_i, m_j)$ menyatakan jarak *includean* dari m ke m_j menyatakan jarak *Euclidean* dari m ke m_j . Menghitung WCV dengan memilih jarak terkecil antara data dengan *centroid* pada masing-masing *cluster*.

Menghitung MCV Yaitu dengan memilih jarak terkecil antara data dengan *centroid* pada masing-masing *cluster* dapat dilihat pada tabel 4.4

$$WCV = 100000,0001^2 + 100000,0001^2 + 100000,0001^2 + 100000,0001^2 + 100000,0001^2 + 100000,0082^2 + 100000,0082^2 + 100000,0082^2 + 50000,00119^2 + 50000,00119^2 + 50000,00119^2 + \dots + 20,46948949^2 + 20,46948949^2 + 20,4450483^2 + 20,4450483^2 + 20,4450483^2 = 12576900699857$$

$$\text{Sehingga besaran rasio} = BCV/WCV = 4400000,036/12576900699857 = 3,498$$

Karena langkah ini merupakan iterasi 1 maka lanjutkan ke langkah berikutnya

b. langkah 4: pembaruan *centroid* dengan menghitung rata-rata nilai pada masing-masing *cluster* dapat dilihat pada tabel 4.4

CLUSTER 1																
no	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
12	5	0	0	1	96	1	4	2	1	1	1	1	1	1	9	1600000
13	5	0	0	1	96	1	4	2	1	1	1	1	1	1	9	1600000
14	5	0	0	1	96	1	4	2	1	1	1	1	1	1	9	1600000
15	5	0	0	1	96	1	4	2	1	1	1	1	1	1	9	1600000
16	5	0	0	1	96	1	4	2	1	1	1	1	1	1	9	1600000
497	5	2	0	2	28	1	4	2	1	1	1	1	1	1	14	2000000
498	5	2	0	2	28	1	4	2	1	1	1	1	1	1	14	2000000
499	5	2	0	2	28	1	4	2	1	1	1	1	1	1	14	2000000
MEAN	5,1	0,9	0,1	2	44,1	1	3,4	1,6	1	1	2,6	0,9	1	1	12	1813793
CLUSTER 2																
9	3	0	0	1	96	1	4	1	1	1	1	1	1	1	4	1200000
10	3	0	0	1	96	1	4	1	1	1	1	1	1	1	4	1200000
11	3	0	0	1	96	1	4	1	1	1	1	1	1	1	4	1200000
17	4	0	0	1	96	2	4	2	1	1	5	1	1	1	14	1250000
18	4	0	0	1	96	2	4	2	1	1	5	1	1	1	14	1250000
19	4	0	0	1	96	2	4	2	1	1	5	1	1	1	14	1250000
408	4	0	0	1	72	1	2	1	1	1	5	1	1	1	9	1200000
449	2	0	0	2	28	1	4	1	1	1	5	1	1	1	0	1100000
452	2	0	0	2	28	1	4	1	1	1	5	1	1	1	0	1100000
MEAN	5,3	0,1	0	1,4	45,5	1	3,8	1,5	1	1	4,2	1	1	1	12	1206536
CLUSTER 3																
31	5	0	0	1	48	1	0	0	1	1	0	1	1	1	0	800000
32	5	0	0	1	48	1	0	0	1	1	0	1	1	1	0	800000
33	5	0	0	1	48	1	0	0	1	1	0	1	1	1	0	800000
34	5	0	0	1	48	1	0	0	1	1	0	1	1	1	0	800000
35	5	0	0	1	48	1	0	0	1	1	0	1	1	1	0	800000
475	6	2	0	1	20	2	4	2	1	1	2	1	1	1	14	800000
476	6	2	0	1	20	2	4	2	1	1	2	1	1	1	14	800000
477	4	0	0	1	48	1	1	1	1	1	1	1	1	1	1	500000
478	4	0	0	1	48	1	1	1	1	1	1	1	1	1	1	500000
479	4	0	0	1	48	1	1	1	1	1	1	1	1	1	1	500000
MEAN	4,1	0,3	0,1	1,6	40,5	1,1	2,5	1,1	1	1	2,9	1	1	1,1	8,2	657500
CLUSTER 4																
6	3	0	0	1	72	1	4	2	1	1	5	1	1	1	15	1000000
7	3	0	0	1	72	1	4	2	1	1	5	1	1	1	15	1000000
8	3	0	0	1	72	1	4	2	1	1	5	1	1	1	15	1000000
36	4	0	0	0	48	1	0	0	1	1	0	1	1	1	0	900000
37	4	0	0	0	48	1	0	0	1	1	0	1	1	1	0	900000
38	4	0	0	0	48	1	0	0	1	1	0	1	1	1	0	900000
511	5	2	0	1	21	2	4	2	1	1	2	1	1	1	14	900000
512	5	2	0	1	21	2	4	2	1	1	2	1	1	1	14	900000
MEAN	5,1	0,3	0,1	1,4	37,2	1,2	2,8	1,3	1	1	3,1	1	1	1,1	8,2	964814,8
CLUSTER 5																
1	5	0	0	1	42	1	4	2	1	1	5	1	1	1	10	1400000
2	5	0	0	1	42	1	4	2	1	1	5	1	1	1	10	1400000
3	5	0	0	1	42	1	4	2	1	1	5	1	1	1	10	1400000
4	5	0	0	1	42	1	4	2	1	1	5	1	1	1	10	1400000
5	5	0	0	1	42	1	4	2	1	1	5	1	1	1	10	1400000
21	5	0	0	1	96	1	4	2	1	1	1	1	1	1	3	1500000
493	3	2	0	2	42	2	3	2	1	1	1	1	1	1	14	1500000
494	3	2	0	2	42	2	3	2	1	1	1	1	1	1	14	1500000
MEAN	5,2	0,2	0	1,5	49,1	1	3,5	1,7	1	1	3,1	1	1	1	11	1476190

Sehingga di dapat centroid baru yaitu :

Centroid $m_1 = \{5.103448276, 0.896551724, 0.068965517, 1.965517241, 44.06896552, 1, 3.379310345, 1.620689655, 1, 1, 2.620689655, 0.931034483, 1, 1, 11.72413793, 1813793\}$

Centroid $m_2 = \{5.333333333, 0.117647059, 0.045751634, 1.444444444, 45.54666667, 1.032679739, 3.764705882, 1.496732026, 1, 0.921568627, 4.196078431, 1, 1, 1, 11.92810458, 1206536\}$

Centroid $m_3 = \{4.121212121, 0.257575758, 0.106060606, 1.583333333, 40.46212121, 1.106060606, 2.515151515, 1.083333333, 1.106060606, 1, 2.939393939, 0.96969697, 1, 1.106060606, 8.151515152, 657500\}$

Centroid $m_4 = \{5.059259259, 0.296296296, 0.066666667, 1.42962963, 37.17037037, 1.214814815, 2.82962963, 1.340740741, 1.059259259, 1.059259259, 3.074074074, 1, 1, 1.088888889, 8.177777778, 964815\}$

Centroid $m_5 = \{5.158730159, 0.222222222, 0, 1.53968254, 49.06451613, 1.047619048, 3.507936508, 1.650793651, 1, 1, 3.142857143, 1, 1, 1, 11.14285714, 1476190\}$

5. Langkah 5: (iterasi 2) kembali kelangkah 3, jika masih ada data yang berpindah cluster atau nilai centroid diatas ambang, atau jika nilai pada fungsi objektif yang akan digunakan masih diatas ambang. Selanjutnya langkah ini dilakukan penempatan lagi data kecentroid terdekat sama seperti

yang dilakukan langkah 3. Untuk menghitung jarak kecentroid masing-masing cluster pada data penduduk sebagai berikut:

Data: 5,0,0,1,42,1,4,2,1,1,5,1,1,1,10,1400000

Centroid m1={5.103448276, 0.896551724, 0.068965517, 1.965517241, 44.06896552, 1, 3.379310345, 1.620689655, 1, 1, 2.620689655, 0.931034483, 1, 1, 11.72413793, 1813793}

Centroid m2={5.333333333, 0.117647059, 0.045751634, 1.444444444, 45.54666667, 1.032679739, 3.764705882, 1.496732026, 1, 0.921568627, 4.196078431, 1, 1, 1, 11.92810458, 1206536}

Centroid m3={4.121212121, 0.257575758, 0.106060606, 1.583333333, 40.46212121, 1.106060606, 2.515151515, 1.083333333, 1.106060606, 1, 2.939393939, 0.96969697, 1, 1.106060606, 8.151515152, 657500}

Centroid m4={5.059259259, 0.296296296, 0.066666667, 1.42962963, 37.17037037, 1.214814815, 2.82962963, 1.340740741, 1.059259259, 1.059259259, 3.074074074, 1, 1, 1.088888889, 8.177777778, 964815}

Centroid m5={5.158730159, 0.222222222, 0, 1.53968254, 49.06451613, 1.047619048, 3.507936508, 1.650793651, 1, 1, 3.142857143, 1, 1, 1, 11.14285714, 1476190}

Adapun hasil perhitungan tersebut dapat dilihat pada tabel 4.6

Tabel 4.5 hasil perhitungan jarak antara masing-masing data dengan centroid (iterasi 2)

no	C1	C2	C3	C4	C5	jarak terkecil	c1	c2	c3	c4	c5
1	413793.1	193464.1	74250.0	435185.2	76190.48	76190.478					v
2	413793.1	193464.1	74250.0	435185.2	76190.48	76190.478					v
3	413793.1	193464.1	74250.0	435185.2	76190.48	76190.478					v
4	413793.1	193464.1	74250.0	435185.2	76190.48	76190.478					v
5	413793.1	193464.1	74250.0	435185.2	76190.48	76190.478					v
6	813793.1	206535.9	34250.0	35185.15	476190.5	35185.193					v
7	813793.1	206535.9	34250.0	35185.15	476190.5	35185.193					v
8	813793.1	206535.9	34250.0	35185.15	476190.5	35185.193					v
9	613793.1	6536.149	54250.0	235185.2	276190.5	6536.1488		v			
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
502	913793.1	306535.9	24250.0	64814.83	576190.5	64814.827					v
503	913793.1	306535.9	24250.0	64814.83	576190.5	64814.827					v
504	913793.1	306535.9	24250.0	64814.83	576190.5	64814.827					v
505	913793.1	306535.9	24250.0	64814.83	576190.5	64814.827					v
506	913793.1	306535.9	24250.0	64814.83	576190.5	64814.827					v
507	913793.1	306535.9	24250.0	64814.83	576190.5	64814.827					v
508	913793.1	306535.9	24250.0	64814.83	576190.5	64814.827					v
509	913793.1	306535.9	24250.0	64814.83	576190.5	64814.827					v
510	913793.1	306535.9	24250.0	64814.83	576190.5	64814.827					v
511	913793.1	306535.9	24250.0	64814.83	576190.5	64814.827					v
512	913793.1	306535.9	24250.0	64814.83	576190.5	64814.827					v

Dari data 4.6 didapatkan kenanggotaan penduduk sbegai berikut:

Cluster1={ 12,13,14,15,16,184,186,188,189,196,197,234,274,448,451,454,480,481,482,487,488,489,490, 491,495,496,497,498,499},

Cluster2={9,10,11,17,18,19,20,107,108,109,110,121,122,123,128,129,130,131,132,139,140,143,144,145, 146,147,165,166,167,168,169,171,172,174,175,178,179,180,181,182,183,185,187,190,191,192,19 3,194,195,199,202,203,204,206,207,209,210,211,212,213,214,216,219,220,221,222,224,225,226, 228,233,236,237,238,241,242,243,244,245,246,247,249,250,251,252,253,255,256,257,258,259,26 3,264,265,266,267,269,273,275,276,277,280,281,282,283,284,285,286,288,289,290,291,292,294, 295,296,297,298,299,300,301,302,303,304,305,306,307,308,309,310,313,314,317,318,319,320,32 1,322,323,324,325,326,327,328,329,330,335,398,401,407,408,449,452}

Cluster3={ 31,32,33,34,35,40,41,42,43,62,63,64,65,66,67,68,69,82,83,84,85,86,87,88,89,90,91,92,93,94,9 5,96,104,105,106,114,138,141,142,148,149,150,151,152,153,154,155,156,170,173,176,177,217,2 31,261,262,315,347,349,351,352,353,354,356,358,359,360,363,366,367,368,369,372,373,374,375 ,376,377,378,379,380,381,382,383,384,385,386,387,389,391,392,393,396,399,402,403,404,409,4

12,415,422,424,425,427,428,431,432,434,435,455,456,457,458,459,461,462,463,465,466,467,468,469,470,471,472,473,474,475,476,477,478,479}

Cluster4={6,7,8,36,37,38,39,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,97,98,99,100,101,102,103,111,112,113,115,116,117,118,119,120,124,125,126,127,157,158,159,160,198,200,201,205,208,218,223,240,248,279,287,316,331,332,333,334,339,340,341,342,343,344,345,346,348,350,355,357,361,362,364,365,370,371,388,390,394,395,397,400,405,406,410,411,413,414,416,417,418,419,420,421,423,426,429,430,433,436,437,438,439,440,441,442,443,444,445,446,447,450,453,460,464,500,501,502,503,504,505,506,507,508,509,510,511,512}

Cluster5={1,2,3,4,5,21,22,23,24,25,26,27,28,29,30,70,71,72,73,74,75,76,77,78,79,80,81,133,134,135,136,137,161,162,163,164,215,227,229,230,232,235,239,254,260,268,270,271,272,278,293,311,312,336,337,338,483,484,485,486,492,493,494}

Pada langkah ini dihitung pula rasio antar besaran BCV (*Between Cluster Variation*) dengan WCV (*Within Cluster Variation*):

$$BCV=d(m1,m2)+d(m1,m3)+d(m1,m4)+d(m1,m5)+d(m2,m3)+d(m2,m4)+d(m2,m5)+d(m3,m4)+d(m3,m5)+d(m4,m5)= 5647923,715$$

$$WCV=76190,47636^2+76190,47636^2+76190,47636^2+123809,5329^2+123809,5329^2+..+23809,5717^2+23809,5717^2+142500,0005^2+64814,82639^2+..+64814,82549^2+64814,82549^2+..+23809,52592^2+23809,52592^2+..+64814,82713^2+64814,82713^2= 7121397575383$$

Sehingga Besar Rasio = BCV/WCV=5647923,715/7121397575383=7,9309

Bila dibandingkan maka rasio sekarang (7,9309) lebih besar dari rasio sebelumnya (3,498) oleh karena itu algoritma dilanjutkan kelangkah berikutnya.

Langkah ke 4 iterasi ke 3

Pada langkah ini dilakukan pembaruan centroid lagi seperti terlihat pada tabel 4.6

Tabel 4.6 hasil perhitungan jarak antara masing-masing data dengan centroid (iterasi 3)

CLUSTER 1																
no	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
234	5	0	0	1	36	1	4	2	1	1	2	0	1	1	14	2500000
274	5	0	0	1	36	1	4	2	1	1	2	0	1	1	14	2500000
448	6	0	0	2	28	1	2	1	1	1	2	1	1	1	14	2000000
451	6	0	0	2	28	1	2	1	1	1	2	1	1	1	14	2000000
.....
497	5	2	0	2	28	1	4	2	1	1	1	1	1	1	14	2000000
498	5	2	0	2	28	1	4	2	1	1	1	1	1	1	14	2000000
499	5	2	0	2	28	1	4	2	1	1	1	1	1	1	14	2000000
mean	4,8	1,2	0	1,8	32,5	1	3,5	1,5	1	1	2,0	0,8	1	1	12,4	2076923
CLUSTER 2																
9	3	0	0	1	96	1	4	1	1	1	1	1	1	1	4	1200000
10	3	0	0	1	96	1	4	1	1	1	1	1	1	1	4	1200000
11	3	0	0	1	96	1	4	1	1	1	1	1	1	1	4	1200000
.....
449	2	0	0	2	28	1	4	1	1	1	5	1	1	1	0	1100000
452	2	0	0	2	28	1	4	1	1	1	5	1	1	1	0	1100000
MEAN	5,3	0,1	0,0	1,4	45,5	1	3,8	1,5	1	0,9	4,2	1	1	1	11,9	1206536
CLUSTER 3																
31	5	0	0	1	48	1	0	0	1	1	0	1	1	1	0	800000
32	5	0	0	1	48	1	0	0	1	1	0	1	1	1	0	800000
33	5	0	0	1	48	1	0	0	1	1	0	1	1	1	0	800000
.....
477	4	0	0	1	48	1	1	1	1	1	1	1	1	1	1	500000
478	4	0	0	1	48	1	1	1	1	1	1	1	1	1	1	500000
479	4	0	0	1	48	1	1	1	1	1	1	1	1	1	1	500000
MEAN	4,1	0,3	0,1	1,6	40,3	1,1	2,6	1,1	1,1	1	3,0	1	1	1,1	8,3	653023,3

CLUSTER 4																
6	3	0	0	1	72	1	4	2	1	1	5	1	1	1	15	1000000
7	3	0	0	1	72	1	4	2	1	1	5	1	1	1	15	1000000
8	3	0	0	1	72	1	4	2	1	1	5	1	1	1	15	1000000
....
511	5	2	0	1	21	2	4	2	1	1	2	1	1	1	14	900000
512	5	2	0	1	21	2	4	2	1	1	2	1	1	1	14	900000
MEAN	5	0,3	0,07	1,42	37,4	1,2	2,8	1,31	1,06	1,06	3,01	1	1	1	1,1	962318,8
CLUSTER 5																
1	5	0	0	1	42	1	4	2	1	1	5	1	1	1	10	1400000
2	5	0	0	1	42	1	4	2	1	1	5	1	1	1	10	1400000
3	5	0	0	1	42	1	4	2	1	1	5	1	1	1	10	1400000
....
493	3	2	0	2	42	2	3	2	1	1	1	1	1	1	14	1500000
494	3	2	0	2	42	2	3	2	1	1	1	1	1	1	14	1500000
MEAN	5,2	0,3	0,03	1,65	50	1	3,5	1,66	1	1	3,13	1	1	1	11,2	1501266

3. Langkah ketiga iterasi ketiga

Untuk menghitung jarak ke centroid masing-masing cluster pada data penduduk sebagai berikut:
Data mean ke

$$DM1 = \{4.769230769, 1.230769231, 0, 1.846153846, 32.46153846, 1, 3.538461538, 1.538461538, 1, 1, 2.076923077, 0.846153846, 1, 1, 12.38461538, 2076923.077\}$$

$$DM2 = \{5.333333333, 0.117647059, 0.045751634, 1.444444444, 45.54666667, 1.032679739, 3.764705882, 1.496732026, 1, 0.921568627, 4.196078431, 1, 1, 1, 11.92810458, 1206535.948\}$$

$$DM3 = \{4.147286822, 0, 0.108527132, 1.596899225, 40.28682171, 1.108527132, 2.573643411, 1.108527132, 1, 1, 3.007751938, 0.968992248, 1, 1.108527132, 8.341085271, 653023.2558\}$$

$$DM4 = \{5.014492754, 0, 0.065217391, 1.420289855, 37.4057971, 1.210144928, 2.768115942, 1.311594203, 1, 1.057971014, 3.007246377, 1, 1.086956522, 8, 962318.8406\}$$

$$DM5 = \{5.202531646, 0, 0.025316456, 1.64556962, 49.97435897, 1.037974684, 3.455696203, 1.658227848, 1, 1, 3.126582278, 1, 1, 1, 11.15189873, 1501265.823\}$$

Tabel 4.7 hasil perhitungan jarak antara masing-masing data dengan centroid (iterasi 3)

no	C1	C2	C3	C4	C5	jarak terkecil	c1	c2	c3	c4	c5
1	200000	150000	600000	500000	100000	100000,0001					v
2	200000	150000	600000	500000	100000	100000,0001					v
3	200000	150000	600000	500000	100000	100000,0001					v
4	200000	150000	600000	500000	100000	100000,0001					v
5	200000	150000	600000	500000	100000	100000,0001					v
6	600000	250000	200000	100000	500000	100000,0082				v	
7	600000	250000	200000	100000	500000	100000,0082				v	
8	600000	250000	200000	100000	500000	100000,0082				v	
9	400000	500000	400000	300000	300000	50000,00119		v			
10	400000	500000	400000	300000	300000	50000,00119		v			
11	400000	500000	400000	300000	300000	50000,00119		v			
....
507	700000	350000	100000	20,4695	600000	20,46948949				v	
508	700000	350000	100000	20,4695	600000	20,46948949				v	
509	700000	350000	100000	20,4695	600000	20,46948949				v	
510	700000	350000	100000	20,4451	600000	20,4450483				v	
511	700000	350000	100000	20,4451	600000	20,4450483				v	
512	700000	350000	100000	20,4451	600000	20,4450483				v	

Dari tabel 4.9 didapat keanggotaan data penduduk sebagai berikut:

- Cluster 1 = { 234,274,448,451,454,480,481,482,495,496,497,498,499 },
 cluster2 = { 9,10,11,17,18,19,20,107,108,109,110,121,122,123,128,129,130,131,132,139,140,143,144,145, 146,147,165,166,167,168,169,171,172,174,175,178,179,180,181,182,183,185,187,190,191,192,19 3,194,195,199,202,203,204,206,207,209,210,211,212,213,214,216,219,220,221,222,224,225,226,

228,233,236,237,238,241,242,243,244,245,246,247,249,250,251,252,253,255,256,257,258,259,263,264,265,266,267,269,273,275,276,277,280,281,282,283,284,285,286,288,289,290,291,292,294,295,296,297,298,299,300,301,302,303,304,305,306,307,308,309,310,313,314,317,318,319,320,321,322,323,324,325,326,327,328,329,330,335,398,401,407,408,449,452},

cluster3={31,32,33,34,35,40,41,42,43,62,63,64,65,66,67,68,69,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,114,138,141,142,148,149,150,151,152,153,154,155,156,170,173,176,177,217,231,261,262,315,347,349,351,352,353,354,356,358,359,360,363,366,367,368,369,372,373,374,375,376,377,378,379,380,381,382,383,384,385,386,387,389,391,392,393,396,399,402,403,404,409,412,415,422,424,425,427,428,431,432,434,435,455,456,457,458,459,461,462,463,465,466,467,468,469,470,471,472,473,474,475,476,477,478,479},

Cluster4={6,7,8,36,37,38,39,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,97,98,99,100,101,102,103,104,105,106,111,112,113,115,116,117,118,119,120,124,125,126,127,157,158,159,160,198,200,201,205,208,218,223,240,248,279,287,316,331,332,333,334,339,340,341,342,343,344,345,346,348,350,355,357,361,362,364,365,370,371,388,390,394,395,397,400,405,406,410,411,413,414,416,417,418,419,420,421,423,426,429,430,433,436,437,438,439,440,441,442,443,444,445,446,447,450,453,460,464,500,501,502,503,504,505,506,507,508,509,510,511,512},

Cluster5={1,2,3,4,5,12,13,14,15,16,21,22,23,24,25,26,27,28,29,30,70,71,72,73,74,75,76,77,78,79,80,81,133,134,135,136,137,161,162,163,164,184,186,188,189,196,197,215,227,229,230,232,235,239,254,260,268,270,271,272,278,293,311,312,336,337,338,483,484,485,486,487,488,489,490,491,492,493,494}

Pada langkah ini dihitung pula rasio antara besaran BCV (*Between Cluster Variation*) dengan WCV (*Within Cluster Variation*):

$$BCV = d(m1, m2) + d(m1, m3) + d(m1, m4) + d(m1, m5) + d(m2, m3) + d(m2, m4) + d(m2, m5) + d(m3, m4) + d(m3, m5) + d(m4, m5) = 6773493,251$$

$$WCV = 101265,8233^2 + 101265,8233^2 + 101265,8233^2 + \dots + 98734,18778^2 + 98734,18778^2 + \dots + 1266,687873^2 + 1266,687873^2 + \dots + 146976,7447^2 + 62318,84219^2 + \dots + 62318,8413^2 + 62318,8413^2 + \dots + 1265,865954^2 + 1265,865954^2 + \dots + 62318,8431^2 + 62318,8431^2 = 6168193472922$$

Sehingga besar rasio = $BCV/WCV = 6773493,251/6168193472922 = 1,09813$

Bila dibandingkan maka rasio sekarang (1,09813) sudah tidak lagi lebih besar dari rasio sebelumnya (7,9309) dan cluster tidak berpindah tempat oleh karena itu algoritma akan dihentikan.

5. KESIMPULAN

Kesimpulan yang dapat di ambil dari hasil yang penelitian yang telah dilakukan pada bab-bab sebelumnya yaitu

5.1 Simpulan

Berdasarkan hasil dari pengamatan dan penelitian yang telah dilakukan pada Dinas Sosial Kota Jambi, maka dapat diambil kesimpulan sebagai berikut :

1. Penelitian ini menggunakan data penduduk Jambi Selatan yang diambil di Dinas Sosial Kota Jambi dengan jumlah data yang digunakan 512 dan mempunyai 16 atribut yaitu : tanggungan dalam keluarga, anggota keluarga yang cacat, lanjut usia, penguasaan bangunan, luas bangunan, jenis lantai, jenis atap, jenis dinding, fasilitas tempat buang air besar, tempat pembuangan akhir tinja, sumber air minum, sumber penerangan utama, apakah memiliki dapur, jenis bahan bakar untuk memasak, pekerjaan utama kepala keluarga, dan pendapatan perbulan.
2. Metode yang digunakan dalam penelitian ini adalah *K-Means Clustering* dari perhitungan yang telah dilakukan, maka direkomendasikan penulis 5 cluster.
3. Hasil perbandingan perhitungan Hasil untuk cluster 1 13 penduduk , Cluster 2 sebanyak 153, Diharapkan dalam cluster 129 penduduk, Cluster 4 sebanyak 138 penduduk, Cluster 5 sebanyak 79 penduduk.

5.2 Saran

Analisa perhitungan penduduk miskin ini masih jauh dari sempurna sehingga perlu dilakukan perbaikan dan pengembangan, maka ada beberapa saran yang dapat diberikan untuk pengembangan lebih lanjut, yaitu :

1. penelitian selanjutnya peneliti dapat menggunakan metode yang lain dalam menganalisa data.
2. Diharapkan untuk analisa selanjutnya peneliti membuat system untuk perhitungan *Clustering K-Means*.

6. DAFTAR PUSTAKA

- [1] Hermansyah Sembiring, N. (2012). Sistem Informasi Jumlah Angkatan Kerja Menggunakan Visual Basic Pada Badan Pusat Statistik (Bps) Kabupaten Langkat, 5 (2), 13–19.
- [2] Sutarman. 2012. Pengantar Teknologi Informasi. Jakarta Pt. Bumi Aksara.
- [3] Aline Embun Pramadhani, T. S. (2014). Penerapan Data Mining Untuk Klasifikasi Prediksi Penyakit Ispa (Infeksi Saluran Pernapasan Akut) Dengan Algoritma Decision Tree (Id3), 2.
- [4] Alfi Fadliana, F. R. (N.D.). Penerapan Metode Agglomerative Hierarchical Clustering Untuk Klasifikasi Kabupaten/Kota Di Provinsi Jawa Timur Berdasarkan Kualitas Pelayanan Keluarga Berencana 1 Alfi.
- [5] Heni Sulastri, A. I. G. (2017). Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia, 02, 299–305.
- [6] P. A. Jusia, “Decision Support System for Supplier Selection using Analytical Hierarchy Process (AHP) Method,” *Sci. J. Informatics*, vol. 4, no. 2, pp. 1–6, 2017.
- [7] Sri Rahayu, Dodon T, Nugrahadi, Fatma Indriani. (2014) Clustering Penentuan Potensi Kejahatan Daerah Di Kota Banjar Baru Dengan Metode K-Means. Kumpulan Jurnal Ilmu Komputer. Volume 01 No 01 September 2014. Issn: 2406-7857.
- [8] Taslim, F. (2016). Penerapan Algorithma K-Mean Untuk Clustering Data Obat Pada Puskesmas Rumbai, X (X), 108–114.
- [9] Almira Qatrunnada Qurratu’ain, V. R. (2016). Analisis Indikator Tingkat Kemiskinan Di Jawa Timur Menggunakan Regresi Panel, 5 (2).
- [10] Rizki, Irtania Muthia, Septiadi Padmadisastra, B. T. (2017). Pengelompokan Rumah Tangga Miskin Di Kecamatan Tabir Barat Menggunakan Metode Latent Class Cluster Analysis, 9 (2), 63–74.
- [11] Risky Dwi Setiyawan, Dwi Sunaryono, R. J. A. (2016). Rancang Bangun Aplikasi Untuk Pemetaan Tingkat Kemiskinan Masyarakat Berbasis Perangkat Bergerak, 5 (2).