

KLASTERISASI DATA PENDUDUK BERDASARKAN STATUS EKONOMI DI DESA GEBANG MENGGUNAKAN METODE K-MEANS CLUSTERING

Ibnu Badrus Syamsi (085234269499)

Daryanto, S.Kom,M.Kom

Lutfi Ali Muharrom, S.Si

Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember

Email : Ibnuadrussyamsi@gmail.com

ABSTRAK

Dalam penelitian ini, peneliti menemukan sebuah permasalahan yang terdapat di Desa Gebang Kabupaten Jember. Permasalahannya adalah ingin mengetahui status ekonomi masyarakat yang sebelumnya memang sulit untuk mengetahui status ekonomi di desa Gebang tersebut. Peneliti menggunakan kuisioner sebanyak 50 responden, karena peneliti tidak memperoleh data dari dinas terkait mengenai data yang diharapkan. Diharapkan dengan menggunakan kuisioner, data yang diperoleh adalah data yang valid sehingga dapat dikelompokkan sesuai dengan maksud dan tujuan penelitian. Data yang dihasilkan dari kuisioner adalah data yang acak dan belum dikelompokkan berdasarkan status ekonomi masyarakat, maka peneliti menggunakan metode *K-means Clustering*. Metode tersebut mampu mengimplementasikan *clustering* (suatu alat untuk analisa data yang memecahkan permasalahan penggolongan) dengan melakukan pengelompokan data tanpa berdasarkan kelas data tertentu. Dari *clustering* itulah dapat diketahui pengelompokan data menjadi tiga cluster yaitu masyarakat dengan status ekonomi kaya, status ekonomi menengah, status ekonomi miskin. Hasil penelitian yang dihasilkan adalah keberhasilan menggunakan metode clustering untuk mengklasifikasikan kelas yang mampu membentuk 3 kelompok status ekonomi.

Kata kunci : *sistem informasi, status ekonomi, k-means, clustering, penduduk.*

PENDAHULUAN

1.1 Latar Belakang

Pada saat ini masyarakat sebenarnya telah mengenal pembagian atau pelapisan sosial sejak dahulu. Pada zaman dahulu, Aristoteles menyatakan bahwa di dalam setiap Negara selalu terdapat tiga unsur, yakni orang-orang kaya sekali, orang-orang miskin dan orang-orang menengah. Menurut Aristoteles orang-orang kaya sekali ditempatkan dalam lapisan atas oleh masyarakat, sedangkan orang-orang miskin ditempatkan dalam lapisan bawah, dan

orang-orang di tengah ditempatkan dalam lapisan menengah. Status sosial merupakan kedudukan atau posisi sosial seseorang dalam masyarakat. Status terbagi atas *ascribed status* yaitu status yang di dapat secara otomatis melalui kelahiran, *achieved status* yaitu status yang di dapat melalui usaha sendiri, *assigned status* yaitu status yang diberikan kepada seseorang karena jasanya bagi masyarakat. Jika dalam suatu masyarakat, faktor ekonomi merupakan salah satu hal yang dihargai maka memungkinkan terjadinya pelapisan atau stratifikasi sosial di bidang ekonomi. Orang-

orang yang mampu memperoleh kekayaan akan menduduki lapisan atas. Pelapisan ekonomi dapat dilihat dari segi pendapatan, kekayaan dan pekerjaan. Kemampuan ekonomi yang berbeda-beda dapat menyebabkan terjadinya stratifikasi ekonomi. Orang-orang yang berpendapatan sangat kecil dan tidak memiliki harta benda akan menduduki lapisan bawah. Suatu hal yang perlu diperhatikan adalah bahwa stratifikasi sosial dalam bidang ekonomi ini bersifat terbuka, jadi perpindahan antar kelas dapat terjadi secara bebas sesuai dengan kemampuan seseorang.

Berdasarkan adanya stratifikasi ekonomi yang telah ada, maka peneliti akan menggolongkan status ekonomi penduduk Desa Gebang Kecamatan Patrang Kabupaten Jember sesuai dengan pekerjaan penduduk tersebut. Status ekonomi adalah kedudukan seseorang atau keluarga di masyarakat berdasarkan pendapatan perbulan. Status ekonomi dapat dilihat dari pendapatan yang disesuaikan dengan harga barang pokok. Karena kurangnya pemahaman, maka dibutuhkan suatu metode pengelompokan data penduduk menurut status ekonomi berdasarkan pekerjaan dengan cara *K-means Clustering*.

Berdasarkan latar belakang di atas, pengelompokan adat dengan status ekonomi menurut pekerjaan sangat diperlukan, agar muntuk mempermudah dalam pengelompokan data dengan menggunakan metode *K-means Clustering* sehingga status ekonomi penduduk Desa Gebang Kecamatan Patrang Kabupaten Jember dapat diidentifikasi dengan mudah. Oleh karena itu, peneliti bermaksud

melakukan penelitian dengan judul "*Klasterisasi Data Penduduk Berdasarkan Status Ekonomi Di Desa Gebang Menggunakan Metode K-means Clustering*".

1.2 Perumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan di atas, maka perumusan masalah yang dikemukakan dala penelitian ini adalah.

1. Bagaimanakah proses pengumpulan data kuesioner ?
2. Bagaimanakah cara pengelompokan data dengan menggunakan metode *K-means* ?

1.3 Tujuan Penelitian

Sejalan dengan perumusan masalah di atas, maka yang menjadi tujuan penelitian ini adalah sebagai berikut.

1. Untuk mengelompokkan data berdasarkan status ekonomi dengan menggunakan *K-means*.
2. Untuk mengevaluasi kinerja *K-means* dalam melakukan *Clustering*.

1.4 Manfaat Penelitian

Penelitian ini diharapkan bermanfaat dan berguna dalam beberapa hal antara lain.

1. Untuk meningkatkan dan mengembangkan kemampuan penulis melalui karya ilmiah yang dilakukan.
2. Sebagai bahan informasi bagi masyarakat Desa Gebang mengenai kondisi status ekonomi masyarakatnya.
3. Sebagai bahan pertimbangan dan sumbangan pemikiran bagi pihak-pihak yang ingin membahas dan memperdalam tentang tingkat sosial ekonomi masyarakat.

TINJAUAN PUSTAKA

2.1 Pengertian Klasterisasi

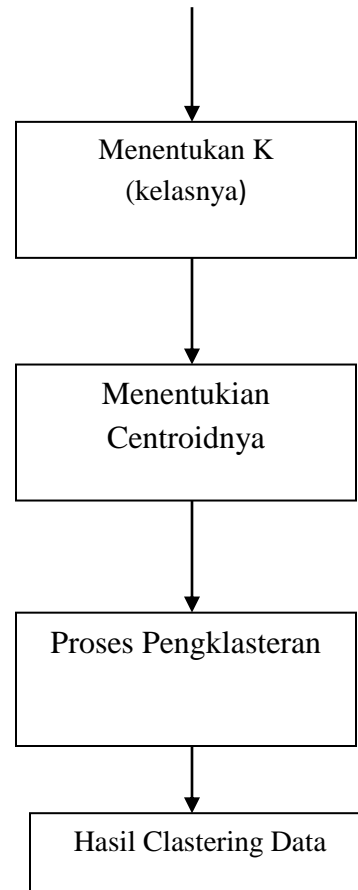
Klasterisasi merupakan salah satu metode untuk mengidentifikasi kelas berdasarkan obyek-obyek. Representasi kesamaan berbeda-beda pada model klasterisasi satu dengan yang lain. Di dalam banyak model, konsep kesamaan dihitung berdasarkan fungsi jarak, meliputi *Manhattan Distance*, *Euclidean distance* dan lain-lain. Tetapi fungsi jarak tidak selalu cukup untuk mengambil korelasi-korelasi antara obyek-obyek yang jauh. Dalam kenyataannya, korelasi-korelasi yang kuat masih ada antra himpunan obyek-obyek yang sama jika himpunann obyek-obyek terpisah jauh dari satu sama lain yang dihitung berdasarakan fungsi jarak.

Metode klasterisasi secara khusus dipelajari dalam berbagai bidang, meliputi *statistic*, *machine learning*, *pattern recognition*, dan *image processing*. Banyak riset tercurah pada berbagai masalah dalam klasterisasi, seperti scalability, the curve of dimensionality, dan lain-lain.

Dalam kebanyakan model klasterisasi, kesamaan antar obyek –obyek yang berbeda pada semua, sebagian, atau hanya subset dari dimensi dihitung oleh jarak. Beberapa fungsi jarak yang terkenal meliputi *Euclidean distance*, *Manhattan distance*, dan *Cosine Distance*. Tetapi fungsi jarak tidak mampu untuk menangkap korelasi-korelasi antar obyek-obyek. Dalam kenyataannya, korelasi-korelais yang kuat masih ada himpunan obyek-obyek yang sama jika himpunan obyek-obyek terpisah jauh dari satu sama lain yang dihitung berdasarkan fungsi jarak (Wijayanti Dan Soelaiman, 2008).

METODOLOGI PENELITIAN

3.1 Gambaran Sistem



Gambar 3.1 Blok Diagram

Langakah – langkah dalam pengklasteran data, antara lain :

1. Input Data
Dalam menginput data, peneliti menyusun kuesioner tentang pengclustering status ekonomi masyarakat. Kemudian kuesioner tersebut di bagikan kepada 48 responden yang ada di Desa Gebang. Setelah selesai responden mengisi kuesioner, peneliti mengambil kuesioner setiap responden sebagai data.
2. Menentukan K.
Dari data yang terkumpul, maka peneliti ingin mengetahui rata-rata

setiap data yang diperoleh dari responden.

3. Menentukan Centroid
Setelah data ditentukan rata-ratanya maka akan ditentukan *centroid* setiap clusternya
4. Proses Pengklasteran
Setelah titik *centroid* ditentukan, barulah proses pengklasteran dilakukan. Yang pada awalnya banyaknya K sudah diketahui sebanyak 3 klaster, terdiri dari dari miskin, sedang, dan kaya.
5. Hasil Clustering Data
Setelah proses pengclustering selesai, maka akan diketahui hasil clustering data.

3.2 Analisis Algoritma K-means

K-means termasuk dalam *Partitioning Clustering* yaitu, setiap data harus masuk dalam cluster tertentu dan memungkinkan bagi setiap data yang termasuk dalam cluster tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke cluster yang lain. K-means memisahkan data ke k daerah bagian yang terpisah, dimana k adalah bilangan interger positif. *Algoritma C-means* sangat terkenal dengan kemudahannya dan kemampuannya untuk mengklasifikasi data besar outlier dengan sangat cepat.

Berikut adalah langkah-langkah algoritma K-means :

- a. Penentuan pusat cluster awal
Dalam menentukan n buah pusat *cluster* awal dilakukan pembangkitan bilangan random yang mempresentasikan urutan data input. Pusat awal cluster didapatkan dari data sendiri bukan dengan menentukan titik baru, yaitu dengan merandom pusat awal dari data.
- b. Perhitungan jarak dengan pusat *cluster*

Untuk mengukur jarak antara data dengan pusat *cluster* digunakan Euclidean Distance.

Algoritma perhitungan jarak data dengan pusat cluster, yaitu sebagai berikut :

- 1) Ambil nilai data dan nilai pusat cluster
- 2) Hitung *Euclidean Distance* data dengan tiap pusat *cluster*
- 3) Pengelompokan data.

Jarak hasil hitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat cluster., jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat cluster terdekat. Algoritma pengelompokan data, yaitu sebagai berikut:

- 1) Ambil nilai jarak tiap pusat cluster dengan data
- 2) Cari nilai jarak terkecil
- 3) Kelompokkan data dengan pusat data cluster yang memiliki jarak terkecil
- 4) Penentuan pusat cluster baru.

Untuk mendapatkan pusat data cluster baru bisa dihitung dari rata-rata nilai anggota cluster dan pusat *cluster*. Pusat cluster yang baru digunakan untuk melakukan iterasi selanjutnya, jika hasil yang didapatkan belum *konvergen*. Proses iterasi akan berhenti jika telah memenuhi maksimum iterasi yang dimasukkan oleh User atau hasil yang dicapai sudah konvergen (pusat cluster baru dengan pusat cluster lama). Algoritma penentuan pusat cluster baru, yaitu sebagai berikut :

- 1) Cari jumlah tiap anggota cluster
- 2) Hitung pusat baru dengan rumus.

HASIL DAN PEMBAHASAN

Pada bab ini akan dibahas hasil yang diperoleh dari pengujian yang telah dilakukan. Data-data yang telah diperoleh, akan diolah untuk mendapatkan hasil klasifikasi yaitu kelas miskin, kelas sedang, kelas kaya. Klasterisasi dilakukan menggunakan algoritma K-Means dengan menghitung probabilitas (penghasilan) dari tiap-tiap atribut.

4.1 Skenario Uji Coba

Seluruh data yang digunakan pada tugas akhir ini akan dibagi menjadi 3 cluster dengan ukuran yang kurang lebih sama berdasarkan metode k-means yang digunakan.

Tabel 4.1 Proporsi jumlah data dari setiap kelas pada masing-masing kelas

Kelas	Centroid	Atribut	Data	Jumlah
K1	983333, 753030	Miskin	33	48
K2	2475000, 1625000	Sedang	8	
K3	3622857, 2742857	Kaya	7	

4.2 Hasil dari Percobaan untuk menentukan cluster

Diketahui angka perkiraan penghasilan dan angka perkiraan pengeluaran 10 masyarakat seperti pada tabel 4.2 masyarakat tersebut akan dikelompokkan berdasarkan penghasilan dan pengeluarannya menjadi tiga kelompok. Proses pengelompokan menggunakan metode -Means.

Tabel 4.2 Penghasilan dan Pengeluaran 10 masyarakat

No	Nama	Penghasilan	Pengeluaran
1.	Sunardi	750.000	550.000
2.	Basar Sujono	1.500.000	700.000

3.	Achmad Qurbani Yusuf	3.000.000	2.000.000
4.	Suwono	3.000.000	2.000.000
5.	Mohammad Rusbandi	3.000.000	2.000.000
6.	Purwadi	3.000.000	2.000.000
7.	Supriyanto	3.000.000	2.000.000
8.	Sutikno	3.000.000	2.000.000
9.	Kusniadi	3.500.000	3.500.000
10.	Moch. Soeh	2.500.000	2.250.000

1. Penulis akan mengelompokkan data tersebut menjadi 3 *cluster*, $K = 3$. Misalkan pusat *cluster* kita tetapkan secara random, $C1 = (1.000.000, 750.000)$; $C2 = (2.000.000, 1.200.000)$; dan $C3 = (4.000.000, 3.000.000)$.
2. Hitung jarak setiap data terhadap setiap pusat *cluster*. Misalkan untuk menghitung jarak data pertama (Sunardi) dengan pusat *cluster* pertama adalah:

Suatu data akan menjadi anggota dari suatu *cluster* yang memiliki jarak terkecil dari pusat clusternya. Misalkan untuk data pertama, jarak terkecil diperoleh pada *cluster* kesatu, sehingga data pertama akan menjadi anggota dari cluster pertama. Demikian juga untuk data yang kedua (Basar Sujono), jarak terkecil ada pada *cluster* pertama, maka data tersebut akan masuk pada cluster pertama. Posisi *cluster* selengkapnya adalah :

Tabel 4.4 Data yang sudah terklasterisasi pada iterasi pertama

No	Nama	Penghasilan	Pengeluaran	C1	C2	C3
1.	Sunardi	750.000	550.000	*		
2.	Basar	1.500.000	700.000	*		

	Sujono	00	0			
3.	Achmad Qurbani Yusuf	3.000.00	2.000.00		*	
4.	Suwono	3.000.00	2.000.00		*	
5.	Mohammad Rusbandi	3.000.00	2.000.00		*	
6.	Purwadi	3.000.00	2.000.00		*	
7.	Supriyanto	3.000.00	2.000.00		*	
8.	Sutikno	3.000.00	2.000.00		*	
9.	Kusniadi	3.500.00	3.500.00			*
10.	Moch. Sokeh	2.500.00	2.250.00		*	

Pembentukan Iterasi kedua dengan pembangkitan centroid atau nilai tengah tiap-tiap data yang sudah terkalsterisasi pada iterasi pertama. Berdasarkan hasil pengolongan tersebut, diperoleh anggota cluster pertama ada 2, cluster kedua ada 7, cluster ketiga ada 1.

Hitung pusat cluster baru. Untuk cluster kesatu, ada 2 data yaitu data ke-1 dan ke-2, sehingga :

$$c = \frac{750.000 + 1.500.000}{2} = 1.125.000$$

$$c = \frac{550.000 + 750.000}{2} = 650.000$$

Untuk cluster kedua, ada 7 data yaitu data ke-3, data ke-4, data ke-5, data ke-6, data ke-7, data ke-8 dan data ke-10, sehingga :

$$c = \frac{3.000.000 + 3.000.000 + 3.000.000 + 3.000.000 + 3.000.000 + 3.000.000 + 2.500.000}{7} = 2.928.571,4$$

$$c = \frac{2.000.000 + 2.000.000 + 2.000.000 + 2.000.000 + 2.000.000 + 2.000.000 + 2.250.000}{7} = 2.035714,2$$

Untuk cluster ketiga, ada 1 data yaitu data ke-9, sehingga :

$$c = \frac{3.500.000}{1} = 3.500.000$$

$$c = \frac{3.500.000}{1} = 3.500.000$$

Ulangi jarak menghitung jarak setiap data terhadap setiap pusat cluster yang baru. Hasil perhitungan jarak selengkapnya terlihat pada tabel 4.5

Suatu data akan menjadi anggota dari suatu cluster yang memiliki jarak terkecil dari pusat clusternya. Misalkan untuk data ketiga, jarak terkecil diperoleh pada cluster ketiga, sehingga data ketiga akan menjadi anggota dari cluster ketiga. Demikian juga untuk data kedua (Achmad Yusuf Qurbani), jarak terkecil ada pada cluster ketiga, maka data tersebut akan masuk pada cluster pertama. Posisi cluster selengkapnya :

Tabel 4.6 Data yang sudah terklasterisasi pada iterasi kedua

N o	Nama	Penghasilan	Pengeluaran	C 1	C 2	C 3
1.	Sunardi	750.000	550.000	*		
2.	Basar Sujono	1.500.000	700.000	*		
3.	Achmad Qurbani Yusuf	3.000.000	2.000.000			*
4.	Suwono	3.000.000	2.000.000			*
5.	Mohammad Rusbandi	3.000.000	2.000.000			*
6.	Purwadi	3.000.000	2.000.000			*
7.	Supriyanto	3.000.000	2.000.000			*

8.	Sutikno	3.000.000	2.000.000			*
9.	Kusniadi	3.500.000	3.500.000			*
10.	Moch. Soceh	2.500.000	2.250.000		*	

Pembentukan iterasi ketiga dengan pembangkitan *centroid* atau nilai tengah tiap-tiap data yang sudah terklusterisasi pada iterasi kedua. Terlihat masih ada 6 data yang berubah posisi dari kondisi semula, yaitu data ke-3, data ke-4, data ke-5, dan data ke-6, data ke-7 dan data ke-8. Sehingga perlu dihitung pusat cluster baru.

Hitung cluster baru sebagaimana pada langkah ke-4, sehingga diperoleh :

Untuk cluster pertama, ada 2 data yaitu data ke-1 dan data ke-2, sehingga :

$$c = \frac{750.000 + 1.500.000}{2} = 1.1250.000$$

$$c = \frac{550.000 + 700.000}{2} = 625.000$$

Untuk cluster ketiga, ada 1 data yaitu data ke-10, sehingga :

$$c = \frac{2.500.000}{1} = 2.500.000$$

$$c = \frac{2.250.000}{1} = 2.250.000$$

Untuk cluster ketiga, ada 7 data yaitu data ke-3 sampai dengan data ke-9, sehingga :

$$c = \frac{3.000.000 + 3.000.000 + 3.000.000 + 3.000.000 + 3.000.000 + 3.000.000 + 3.500.000}{7}$$

$$= 3.071.428,5$$

$$c = \frac{2.000.000 + 2.000.000 + 2.000.000 + 2.000.000 + 2.000.000 + 2.000.000 + 3.500.000}{7}$$

$$= 2.214.285,7$$

Ulangi menghitung jarak setiap data terhadap setiap pusat *cluster* yang baru. Hasil perhitungan jarak selengkapnya terlihat pada tabel 4.7

4.4 Implementasi Aplikasi

Aplikasi klusterisasi data penduduk berdasarkan status ekonomi ini dirancang dengan menggunakan pemrograman Java. Dalam tahap pembuatannya telah disesuaikan dengan perancangan sistem sebelumnya. Dimana sebuah sistem dapat bergerak melalui tahap inputan dari data yang akan diproses sesuai dengan tujuan masing-masing pemrosesan yang nantinya akan menghasilkan output. Dalam implementasi ini menerangkan tampilan desain aplikasi yang telah dibuat untuk dibahas agar dapat mengetahui alur-alur sistem dari aplikasi klusterisasi data penduduk berdasarkan status ekonomi.

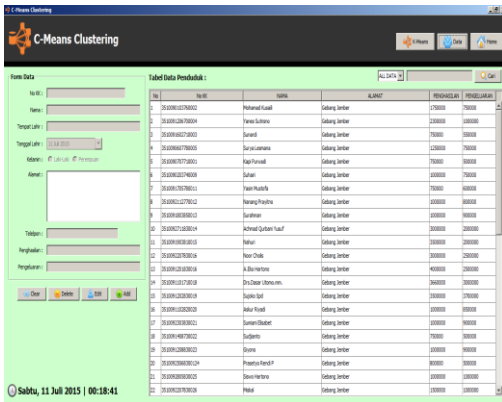
4.4.1 Desain Aplikasi

Rancangan aplikasi klusterisasi data penduduk berdasarkan status ekonomi menggunakan metode K-means dapat diimplementasikan seperti di bawah ini :



Gambar 4.1 Tampilan Home

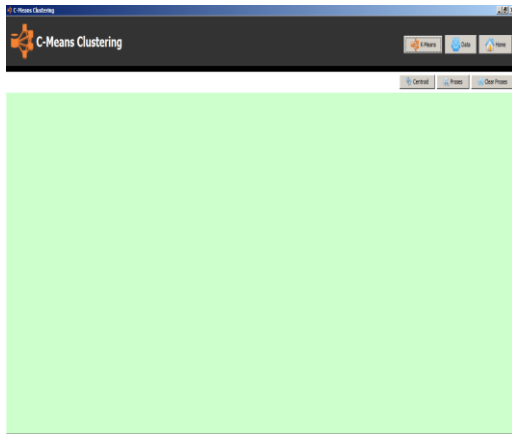
4.4.2 Form Data Set



Gambar 4.2 Form Data Set.

Form data set ini berfungsi untuk menginput data penduduk yang diperoleh dari hasil survey lapangan.

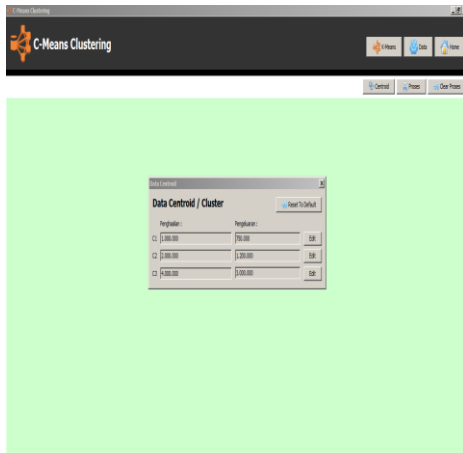
4.4.3 Panel K-means



Gambar 4.3 Panel K-means

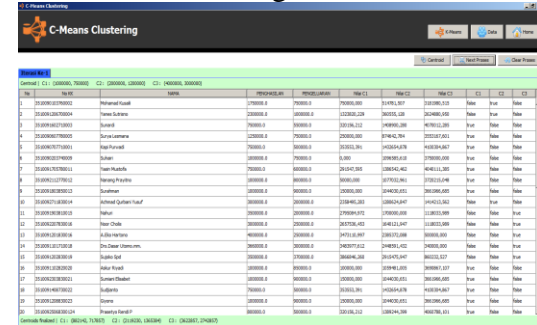
Pada panel ini akan digunakan untuk pemrosesan perhitungan data, yang di dalamnya terdiri dari setting centroid dan proses perhitungan

4.4.4 Form Penentuan Data Centroid



Gambar 4.4 Form Penentuan Data Centroid Form penentuan data centroid ini berfungsi untuk menentukan centroid tiap-tiap kelas.

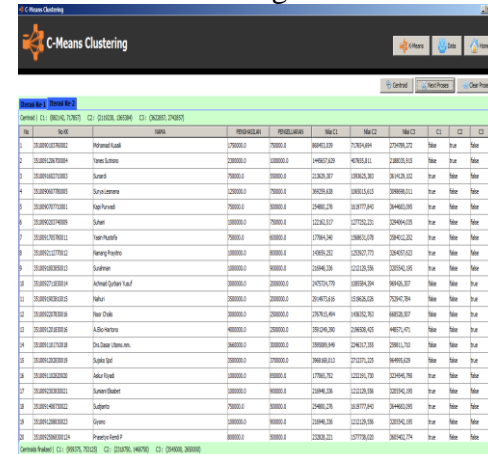
4.4.5 Form Perhitungan Iterasi Pertama



Gambar 4.5 Form Iterasi Pertama

Pada form ini berfungsi untuk memulai proses perhitungan pertama, dimana user hanya akan menentukan pengelompokan tiap-tiap kelas yang pertama.

4.4.6 Form Perhitungan Iterasi Kedua



Gambar 4.6 Form Iterasi Kedua

Pada form iterasi kedua ini fungsinya juga sama merupakan lanjutan dari proses perhitungan iterasi yang pertama, dimana masih terlihat ada 1 data yang berubah posisi dari kondisi semula, sehingga perlu dihitung pusat cluster baru.

metode akurasi untuk menghitung tingkat keberhasilannya.

2. Dalam penelitian ini dataset yang digunakan hanya berupa kuisioner sehingga untuk menambah tingkat kevalidan dataset diharapkan pada pengembangan yang akan datang menggunakan dataset dari Badan Pusat Statistik (BPS) ataupun Instansi atau Lembaga Negara lainnya.

DAFTAR PUSTAKA

Wijayanti dan Soelaiman, 2008. *Klasterisasi Berdasarkan Kesamaan Pola Dengan Menggunakan Algoritma Pcluster*. Institut Teknologi Sepuluh November.

Organisasi.Org. *Klasifikasi, Jenis dan Macam Data – Pembagian Data*

Dalam Ilmu Ekstrak Sains Statistik / Statistika. Komunitas & Perpustakaan Online Indonesia.

Soetjningsih,2004. *Tumbuh Kembang Anak*. Jakarta EGC.

Kartono, 2006. *Perilaku Manusia*. ISBN. Jakarta.

Friedman, 2004. *Keperawatan Keluarga*. Jakarta:EGC

Rismawan dan Kusumadewi, 2008. *Aplikasi K-Means Untuk Pengelompokan Mahasiswa Berdasarkan Nilai Body Mass Index (BMI) & Ukuran Kerangka*. Yogyakarta : Seminar Nasional Aplikasi Teknologi Infrmasi 2008 (SNATI 2008).