

Penerapan Algoritma Fuzzy Clustering Untuk Penentuan Jurusan Sekolah Menengah Kejuruan (SMK Negeri 1 Rambah)

Basorudin

Program Studi Teknik Informatika, Fakultas Ilmu Komputer
Universitas Pasir Pengaraian
Telp. 085263974690
Email : basorudin09@gmail.com

Abstrak : Algoritma Fuzzy C-Means merupakan satu algoritma yang mudah dan sering digunakan di dalam teknik pengelompokan data kerana membuat suatu perkiraan yang efisien dan tidak memerlukan banyak parameter. Beberapa penelitian telah menghasilkan kesimpulan bahwa algoritma Fuzzy C-Means dapat dipergunakan untuk mengelompokkan data berdasarkan atribut-atribut tertentu. Pada penelitian ini akan digunakan algoritma Fuzzy C-Means untuk mengelompokkan data siswa Sekolah Menengah Pertama berdasarkan Nilai mata pelajaran inti untuk proses penjurusan. Penelitian ini juga menguji tingkat akurasi algoritma Fuzzy C-Means dalam penentuan jurusan pada Sekolah Menengah Kejuruan. Dalam penelitian ini data yang akan di Cluster yaitu 4 Cluster yang terdiri dari 4 Jurusan di SMK Negeri 1 Rambah yakni Teknik Komputer Jaringan (TKJ), Teknik Audio Video (TAV), Akuntansi (AK) dan Mekanisasi Pertanian (MP). Penerapan algoritma Fuzzy C-Means dalam penentuan jurusan di Sekolah Menengah Kejuruan pada 20 sampel data siswa yang diuji dalam penelitian ini menunjukkan bahwa Algoritma Fuzzy C-Means memiliki tingkat akurasi yang lebih tinggi yaitu mencapai 55.00%. Dalam penelitian ini pengujian data sampel dilakukan dengan menggunakan bantuan software Matlab 6.1.

Kata Kunci: Klastering, Penjurusan Siswa, Kriteria Ketuntasan Minimum Fuzzy C-Means, Matlab 6.1

Abstract : Fuzzy C-Means algorithm is an algorithm that is easy and is often used in the technique of grouping the data as it makes an estimate efisien and does not require a lot of parameters. Several studies have concluded that the Fuzzy C-Means algorithm can be used to classify data based on certain attributes. In this study will be used Fuzzy C-Means algorithm to classify the junior high school student data based Value core subjects for the majors. The study also examined the level of accuracy of Fuzzy C-Means algorithm in determining the majors in Vocational Education. In this penelitian data to be in Cluster 4 Cluster is comprised of 4 courses in SMK Negeri 1 Rambah namely Computer Engineering Network (TKJ), Engineering Audio Video (TAV), Accounting (AK) and Mechanization of Agriculture (MP). Application of Fuzzy C-Means algorithm in determining the majors in Vocational High School student data at 20 samples tested in this study showed that the Fuzzy C-Means algorithm has a higher accuracy rate, reaching 55.00%. In this study testing the sample data is done by using the help of Matlab 6.1.

Keywords: Clustering, Majors Student, completeness Minimum Criteria Fuzzy C-Means, Matlab 6.1.

PENDAHULUAN

Data mining dengan teknik Clustering, berbeda dengan teknik Assosiation rule mining dan clasification di mana kelas data telah ditentukan sebelumnya. Clustering melakukan pengelompokan data tanpa berdasarkan kelas data tertentu. Bahkan clustering dapat dipakai untuk memberikan label pada kelas data yang belum diketahui. Karena itu clustering sering digolongkan sebagai metode *unsupervised learning*. Prinsip dari clustering adalah memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar kelas atau cluster.

TINJAUAN PUSTAKA

Data mining merupakan bagian dari *Knowledge Discovery in Databases* (KDD). *Knowledge Discovery in Databases* atau biasa disingkat dengan KDD berisi serangkaian proses-proses yang

harus dilakukan sebelum dan sesudah menganalisis dengan metode data mining, (Ira Melissa et al., 2013).

Data Mining *Data Mining* (DM) adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual.

Konsep *Clustering*, *Clustering* adalah suatu teknik untuk mengklasifikasikan data ke dalam kelompok-kelompok tertentu yang didasarkan atas kesamaan sifat atau karakter (Valente J:2010). *Clustering* digunakan untuk mengelompokkan data secara alamiah berdasarkan kemiripan pada objek data dan sebaliknya meminimalkan kemiripan terhadap cluster lain.

Algoritma *Fuzzy C-Means* (FCM), *Fuzzy C-means cluster* pertama kali dikemukakan oleh Dunn (1973) dan kemudian dikembangkan oleh Bezdek (1981) yang banyak digunakan dalam pattern recognition. Metode ini merupakan pengembangan dari metode non hierarki *K-means Cluster*, karena pada awalnya ditentukan dulu jumlah kelompok atau cluster yang akan dibentuk. Kemudian dilakukan iterasi sampai mendapatkan keanggotaan kelompok tersebut (Anindya_et.al. 2009) .

Sistem Penilaian dan Penentuan Jurusan di Sekolah Menengah Kejuruan (SMK), Sesuai kurikulum yang berlaku di seluruh Indonesia, siswa kelas IX SMP yang lulus dan akan melanjutkan ke jenjang berikutnya akan mengalami pemilihan jurusan (penjurusan). Penjurusan yang tersedia di SMK itu sendiri tergantung dari SMK masing-masing, karena setiap SMK tidaklah sama. Penjurusan akan disesuaikan dengan kemampuan dan minat siswa. Tujuannya agar kelak di kemudian hari, pelajaran yang akan diberikan kepada siswa menjadi lebih terarah karena telah sesuai dengan minatnya.

Ernawati dan Susanto dari Program Studi Teknik Informatika Universitas Atma Jaya Yogyakarta, meneliti tentang penerapan *Fuzzy Clustering* untuk pembagian kelas peserta kuliah. Penelitian ini menggunakan sampel 121 orang mahasiswa Jurusan Teknik Informatika Universitas Atma Jaya Yogyakarta yang menempuh perkuliahan Struktur Data Lanjut. Setiap mahasiswa dicirikan oleh dua atribut, yaitu nilai yang pernah mereka peroleh untuk dua mata kuliah prasyarat (Algoritma dan Pemrograman dan Struktur Data). Dalam pembahasan untuk menilai anggota kelompok yang sah dan *valid*, menggunakan ukuran *Fukuyama-Sugeno's Fuzzy Cluster Validity Index*. Hasil dari penelitian ini adalah pembagian kelas para peserta kuliah sebagai hasil penerapan algoritma *Fuzzy Clustering* terhadap data ke- 121 peserta mata kuliah Struktur Data Lanjut untuk pelbagai kemungkinan jumlah kelas, dan rekomendasi peserta kelas yang *valid* (Bahar 2011).

METODE PENELITIAN

Analisa Data

Dalam menganalisa data dibutuhkan data yang dapat dipertanggung jawabkan kebenarannya, hal ini tentunya bertujuan untuk memaksimalkan keaslian hasil pengujian yang dilakukan pada penelitian ini. Dalam hal ini penelitian dilakukan di SMK Negeri 1 Rambah dengan mengambil data dari database akademik SMK Negeri 1 Rambah, database akademik di SMK Negeri 1 rambah ini sendiri diperoleh dari siswa SMP sederajat mulai dari semester I s/d VI yang mendaftar di SMK Negeri 1 Rambah. Data yang diambil berupa nilai siswa ketika berada di SMP sederajat dan data yang relevan dengan penelitian, data tersebut masih dalam bentuk *Hardcopy*.

Metode Pengumpulan Data

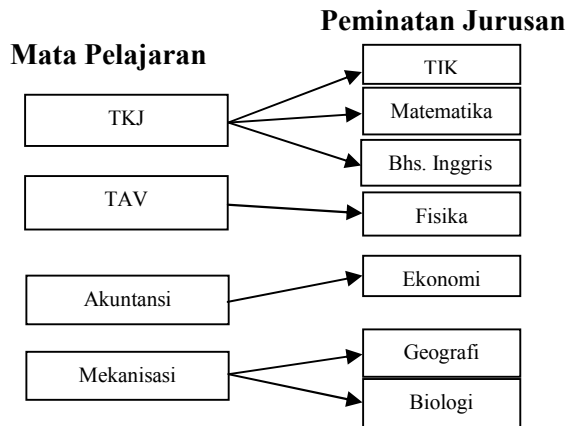
Jenis Data

Data yang digunakan dalam penelitian ini berupa data primer dan data sekunder :

- a. Data Primer : Berupa data tentang mekanisme pelaksanaan peminatan yang dilaksanakan di Sekolah Menengah Kejuruan (SMK) Negeri 1 Rambah, Pasir Pengaraian Kab. Rokan Hulu, yang diperoleh dengan wawancara.
- b. Data Sekunder : Berupa data siswa dan nilai mata pelajaran peminatan sebelum pelaksanaan peminatan, yang diperoleh dari database akademik SMK Negeri 1 Rambah, database akademik di SMK N 1 rambah ini sendiri diperoleh dari siswa SMP sederajat mulai dari semester I s/d VI yang mendaftar di SMK N 1 Rambah.

Sampel Data

Sampel data yang digunakan dalam penelitian ini adalah data nilai (sebelum dan setelah peminatan/penjurusan) siswa SMK Negeri 1 Rambah (angkatan tahun 2013) sebanyak 20 siswa dari 45 siswa. Penetapan jumlah sampel ini di dasarkan pada metode penentuan sampel yang dikembangkan oleh *Isaac* dan *Michael* untuk menentukan ukuran sampel dari populasi mulai dari 10 sampai dengan 1.000.000. Tingkat kesalahan untuk menetapkan jumlah sampel yang dipilih dalam metode ini adalah sekitar 10%. Pada tahap awal dilakukan pemetaan korelasi antara permintaan dengan mata pelajaran peminatan, hasilnya ditunjukkan pada gambar 4.1 berikut :



Gambar 4.1 Diagram Korelasi Antara Mata Pelajaran Dengan Peminatan

Tabel 4.2 Sampel Data Nilai Rata-rata Siswa pada Bidang Minat Tertentu Sebelum Peminatan

Siswa	Nilai Rata-rata Peminatan			
	TKJ	TAV	AK	MP
1	75,2	74,8	76,5	78,6
2	74,9	76,5	67,3	72,8
3	77,5	70,6	74,6	70,6
4	68,1	77,1	77,8	64,4
5	76,6	71,3	76,1	72,9
6	78,2	76,4	64,6	77,8
7	70,4	74,5	72,9	74,5
8	70,4	74,5	72,9	78,2
9	75,3	75,2	68,7	77,1
10	77,1	73,7	80,5	75,5
11	74,0	69,8	80,3	70,4
12	72,8	78,4	69,8	73,9
13	73,7	73,9	71,4	73,9
14	73,1	73,9	72,5	72,8
15	71,9	73,2	74,9	73,7
16	74,3	73,2	70,3	78,3
17	78,3	79,2	79,0	71,8
18	71,8	74,3	77,9	74,3
19	74,1	75,3	80,3	75,3
20	74,7	72,4	72,1	78,5

Sumber : Database Akademik SMK Negeri 1 Rambah, Tahun 2013

Tabel 4.3 Sampel Data Nilai Rata-rata Bidang Peminatan Siswa Angkatan 2013 SMK Negeri 1 Rambah Setelah Peminatan

Siswa	Jurusan Yang Dipilih	Nilai Rata-rata Mata Pelajaran Peminatan Siswa Setelah Peminatan / Penjurusan	
		Kelas XI	Kelas XII
1	TKJ	76,50	73,20
2	TAV	77,00	75,50
3	Akuntansi	75,50	78,00
4	MP	74,00	72,50
5	TKJ	72,80	71,50
6	TKJ	71,50	74,50
7	Akuntansi	69,00	65,50
8	Akuntansi	70,50	72,50
9	MP	76,25	75,50
10	MP	74,50	76,50
11	TKJ	78,00	76,50
12	TAV	73,00	74,80
13	Akuntansi	71,50	74,00
14	MP	68,50	80,00
15	TKJ	75,50	69,50
16	TKJ	84,30	81,30
17	Akuntansi	80,70	82,50
18	Akuntansi	76,00	80,50
19	MP	71,50	82,50
20	TAV	77,00	79,00

Sumber : Database Akademik SMK Negeri 1 Rambah, Tahun 2013

Penerapan Fuzzy Clustering

Menentukan nilai bidang minat tertentu, yang diperoleh dari hasil rata-rata mata pelajaran peminatan yang berada dalam kelompok bidang minat tersebut sebelum dilakukan peminatan. Data ini akan digunakan sebagai data parameter uji coba peminatan menggunakan Algoritma *Fuzzy Clustering*.

Pemetaan/klastering data mengikuti Algoritma *Fuzzy Clustering* yaitu sebagai berikut :

- a. Menetapkan matriks partisi awal U berupa matriks berukuran $n \times m$ (n adalah jumlah sampel data, dan m adalah parameter/atribut setiap data, yaitu = 4). X_{ij} = data sampel ke- i ($i=1,2,3,4,...,n$), atribut ke- j ($j=1,2,3,4,...,m$). Data untuk matriks partisi awal yang digunakan adalah data pada tabel 4.4.
- b. Menentukan nilai parameter awal :
 - Jumlah Cluster (c) = 4
 - Pangkat = 2
 - Maksimum iterasi ($MaxIter$) = 50
 - Error terkecil yang diharapkan (ξ) = 10^{-5}
 - Fungsi objektif awal (P_0) = 0
 - Iterasi awal (t) = 1
- c. Membangkitkan bilangan random μ_{ik} , $i=1,2,3,4,...,n$; $k=1,2,3,4...c$; sebagai elemen-elemen matriks partisi awal (U).
- d. Menentukan Pusat Cluster (V)
 Untuk menentukan Pusat Cluster Pada iterasi pertama Cluster 1, 2, 3 dan 4 menggunakan persamaan sebagai berikut :

$$V_{kj} = \frac{\sum_{i=1}^{20} ((\mu_{ik})^2 * |X_{ij})}{\sum_{i=1}^{20} (\mu_{ik})^2}$$

Menghitung Fungsi Objektif (P)

Untuk menghitung Fungsi objektif pada iterasi pertama ($p1$) dapat dihitung dengan menggunakan rumus :

$$P_1 = \sum_{i=1}^{20} \sum_{k=1}^4 ([\sum_{j=1}^4 (X_{ij} - V_{ij})^2] (\mu_{ik})^2)$$

Menghitung Perubahan Matriks Partisi (U)

HASIL DAN PEMBAHASAN

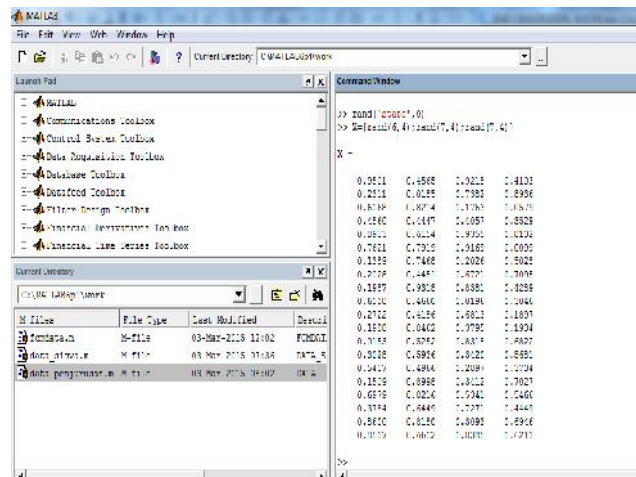
Hasil Pengujian

Untuk membuktikan kebenaran hasil analisa diperlukan sebuah proses pengujian untuk menguji kebenaran dari hasil pengolahan data yang dikerjakan secara manual pada BAB IV yang telah dikerjakan sebelumnya, untuk proses pengujian tersebut disini penulis menggunakan salah satu *software* aplikasi yaitu *MATLAB 6.1*.

Membangkitkan Bilangan Random

Untuk membangkitkan bilangan random μ_{ik} , $i=1,2,3,4,\dots,n$; $k=1,2,3,4,\dots,c$; sebagai elemen-elemen matriks partisi awal (U) dengan menggunakan *Matlab* dapat dilakukan dengan mengetikkan Screenshot atau syntax sebagai berikut :

```
>> rand('state',0)
>> X=[rand(6,4);rand(7,4);rand(7,4)]
X = .....
```

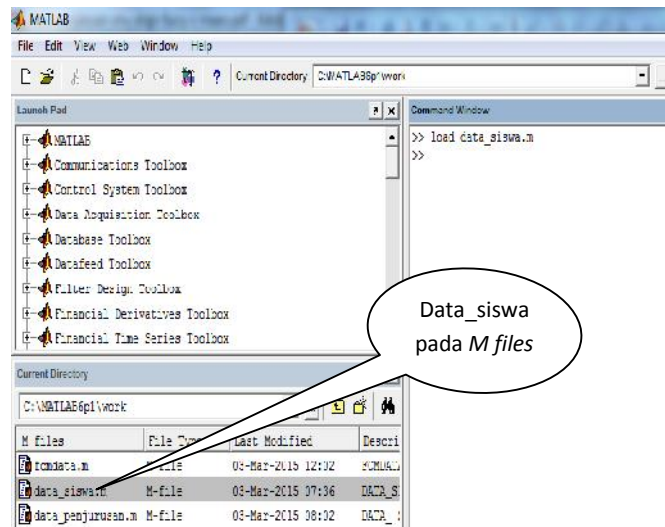


Gambar 4.1 Membangkitkan Bilangan Random dengan Software *MATLAB 6.1*

Implementasi *Fuzzy C-Means* Menggunakan *MATLAB*

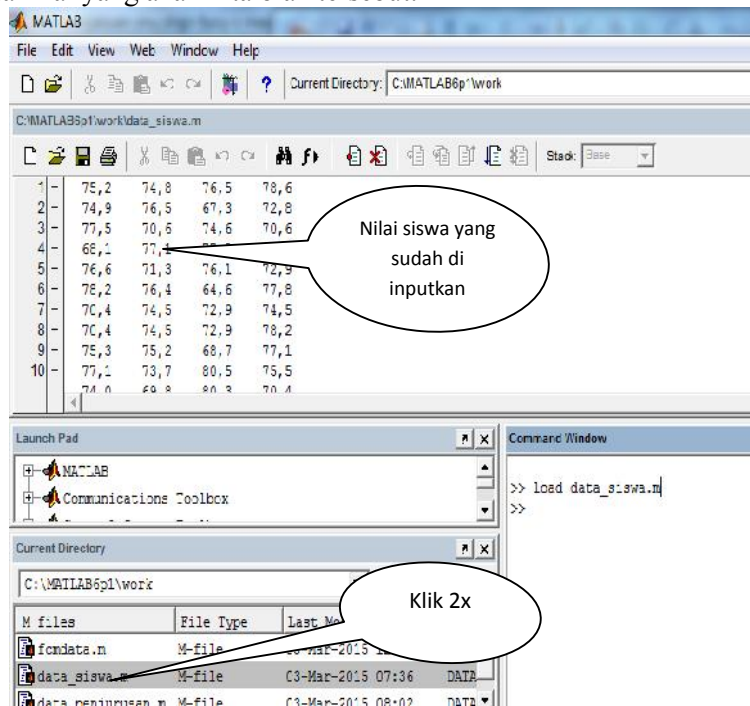
Algoritma *Fuzzy C-Means* dapat diselesaikan dengan menggunakan program *MATLAB*, dengan menjalankan fungsi-fungsi yang sudah tersedia dalam program *MATLAB* itu sendiri, seperti mencari iterasi, menampilkan penyebaran *Clustering* dll. Dalam BAB ini penulis akan mengimplementasikan bagaimana cara mencari iterasi dan menampilkan penyebaran *Clustering* dengan menggunakan program *Matlab*. Untuk mencari *iterasi* langkah-langkah nya adalah sebagai berikut :

1. Siapkan dan tuliskan data yang akan kita olah melalui fasilitas m file, kemudian berinama file misalnya : *data_siswa.m*



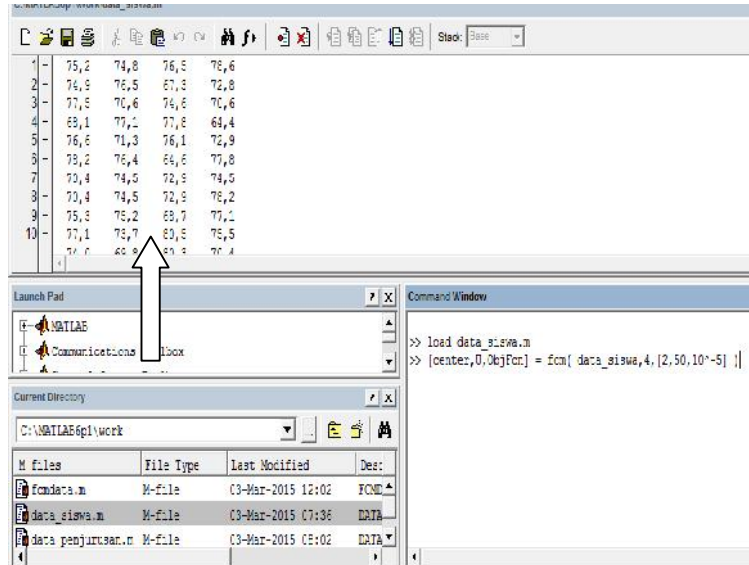
Gambar 5.2 Data Siswa pada M Files

2. Kemudian klik 2x pada data_siswa tersebut maka akan tampil form seperti gambar di bawah ini dan isikan data nilai yang akan kita olah tersebut.



Gambar 5.2 Nilai Siswa yang Sudah di Inputkan Pada M Files

3. Melalui *Command Windows* panggil file data tersebut dengan perintah : `load data_siswa.m` > enter
4. Setelah di enter kemudian ketikkan perintah berikut ini : `center,U,ObjFcn] = fcm(data_siswa,4,[2,50,10^-5])` > enter



Gambar 5.3 Load data_siswa.m dan Iterasi

Keterangan :

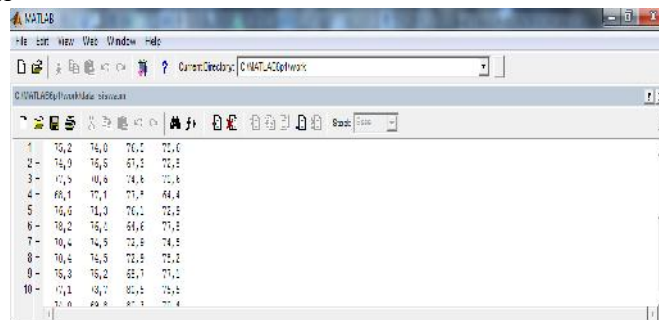
- *Center* berisi pusat *cluster*
- *U* adalah banyaknya kluster yang terbentuk x jumlah data dan berisi derajat keanggotaan tiap data pada masing-masing kluster.
- *ObjFcn* berisi nilai fungsi objektif dari tiap iterasi.

Pada iterasi terakhir ini, pusat kelaster V_{kj} yang dihasilkan (*Software Matlab*) dengan $k=1,2,3,4$; dan $j=1,2,3,4$ adalah :

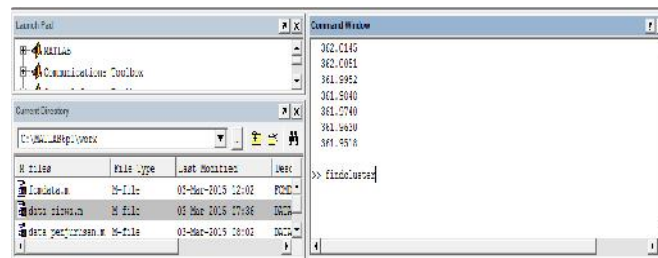
$$V = \begin{pmatrix} 75.1282 & 73.5018 & 78.4217 \\ 73.8746 & 73.6511 & 72.2349 \\ 74.2183 & 74.0958 & 70.1998 \\ 71.6332 & 74.1806 & 73.2296 \end{pmatrix}$$

Untuk menampilkan Penyebaran *Clustering* langkah-langkah nya adalah sebagai berikut :

1. Buka aplikasi *Matlab*, selanjutnya pada *command windows* ketikkan perintah `>> Findcluster`
> enter

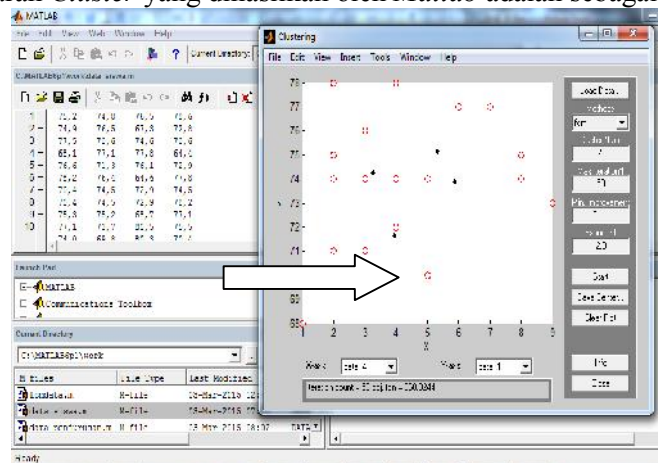


Gambar 5.4 Tampilan Findcluster Pada Matlab



Gambar 5.5 Tampilan Findcluster Pada Matlab

2. Maka penyebaran Cluster yang dihasilkan oleh Matlab adalah sebagai berikut :



Gambar 5.5 Tampilan Cluster Pada Matlab

KESIMPULAN

Dari uraian pada bab-bab yang sudah dibahas sebelumnya maka dapat ditarik kesimpulan, Dari hasil pengujian algoritma Fuzzy C-Means (FCM) dalam penentuan jurusan di Sekolah Menengah Kejuruan dari 20 sampel data siswa yang diuji dalam penelitian ini, menunjukkan bahwa Algoritma FCM memiliki tingkat akurasi yang lebih tinggi (yaitu rata-rata 52,5%). Dari data yang dilatih, diperoleh empat kelompok berdasarkan nilai rata-rata mata pelajaran peminatan, yaitu: Kelompok pertama, terdiri atas siswa yang memiliki nilai rata-rata mata pelajaran peminatan TKJ sekitar 75.0868; nilai rata-rata mata pelajaran peminatan TAV sekitar 73.4915; nilai rata-rata mata pelajaran peminatan AK sekitar 78.3237; dan nilai rata-rata mata pelajaran peminatan MP sekitar 73.5496, Kelompok kedua, terdiri atas siswa yang memiliki nilai rata-rata mata pelajaran peminatan TKJ sekitar 73.9027; nilai rata-rata mata pelajaran peminatan TAV sekitar 74.1341; nilai rata-rata mata pelajaran peminatan AK sekitar 71.3612; dan nilai rata-rata mata pelajaran peminatan MP sekitar 74.4878, Kelompok ketiga, terdiri atas siswa yang memiliki nilai rata-rata mata pelajaran peminatan TKJ sekitar 73.8519; nilai rata-rata mata pelajaran peminatan TAV sekitar 73.8752; nilai rata-rata mata pelajaran peminatan AK sekitar 72.1414; dan nilai rata-rata mata pelajaran peminatan MP sekitar 73.4130, Kelompok keempat, terdiri atas siswa yang memiliki nilai rata-rata mata pelajaran peminatan TKJ sekitar 71.9978; nilai rata-rata mata pelajaran peminatan TAV sekitar 74.1774; nilai rata-rata mata pelajaran peminatan AK sekitar 72.6931; dan nilai rata-rata mata pelajaran peminatan MP sekitar 73.9216.

DAFTAR PUSTAKA

- Anindya, et al. (2009). "Penentuan Banyak Kelompok dalam Fuzzy C-Means Cluster Berdasarkan Proporsi Eigen Value Dari Matriks Similarity dan Indeks XB (Xie dan Beni)."
- Bahar (2011). "Penentuan Jurusan Sekolah Menengah Atas Dengan Algoritma Fuzzy CMeans."
- Ira Melissa, et al (2013). Analisis Data Pembayaran Kredit Nasabah Bank Menggunakan Metode Data Mining, 1 Mei 2013, ISSN: 2085-4579
- Angga Ginanjar Mabur et al. (2012). Penerapan Data Mining Untuk Memprediksi Kriteria Nasabah Kredit, Edisi. I Vol. 1, Maret 2012