

Analisis Kinerja *Fuzzy C-Means* (FCM) dan *Fuzzy Subtractive* (FS) dalam *Clustering* Data Alumni STMIK STIKOM Indonesia

I Kadek Dwi Gandika Supartha*, Adi Panca Saputra Iskandar**

STMIK STIKOM Indonesia

*gandika.supartha@stiki-indonesia.ac.id, **adipancaiskandar@gmail.com

ABSTRACT

In this study, clustering data on STMIK STIKOM Indonesia alumni using the Fuzzy C-Means and Fuzzy Subtractive methods. The method used to test the validity of the cluster is the Modified Partition Coefficient (MPC) and Classification Entropy (CE) index. Clustering is carried out with the aim of finding hidden patterns or information from a fairly large data set, considering that so far the alumni data at STMIK STIKOM Indonesia have not undergone a data mining process. The results of measuring cluster validity using the Modified Partition Coefficient (MPC) and Classification Entropy (CE) index, the Fuzzy C-Means Clustering algorithm has a higher level of validity than the Fuzzy Subtractive Clustering algorithm so it can be said that the Fuzzy C-Means algorithm performs the cluster process better than with the Fuzzy Subtractive method in clustering alumni data. The number of clusters that have the best fitness value / the most optimal number of clusters based on the CE and MPC validity index is 5 clusters. The cluster that has the best characteristics is the 1st cluster which has 514 members (36,82% of the total alumni). With the characteristics of having an average GPA of 3,3617, the average study period is 7,8102 semesters and an average TA work period of 4,9596 months.

Keyword: cluster, alumni, Fuzz C-Means, Fuzzy Subtractive

1. Introduction

Tingginya tingkat persaingan tenaga kerja saat ini menuntut kualitas individu yang semakin meningkat. Hal ini memiliki hubungan yang sangat erat dengan kualitas lulusan perguruan tinggi, dengan demikian kualitas suatu perguruan tinggi pun perlu ditinjau kembali untuk ditingkatkan. Selain masukan yang baik dari *tracer study* sebuah perguruan tinggi, kualitas sebuah perguruan tinggi dapat dilihat dari sebaran lulusan di dunia kerja. Lulusan pada perguruan tinggi disebut sebagai alumni atau merupakan mahasiswa/i yang telah mengikuti atau tamat dari perguruan tinggi.

STMIK STIKOM Indonesia (STIKI Indonesia) merupakan salah satu perguruan tinggi yang berlokasi di Denpasar Bali, berdiri sejak tahun 2008 dan sampai saat ini memiliki alumni kurang lebih 2000 orang. Mahasiswa yang telah menjadi alumni ditampung dalam sebuah organisasi dengan nama Ikatan Alumni STIKI Indonesia (IKA-STIKI). IKA-STIKI memiliki tugas untuk mengelola data alumni dan menjalin komunikasi dengan alumni. Secara berkala pendataan alumni dilakukan diantaranya dengan cara *tracer study*, informasi yang didapatkan yaitu tempat bekerja, bidang kerja, lama bekerja, kesesuaian bidang kerja, jumlah alumni, IPK (Indeks Prestasi Kumulatif), lama *study* dan lama pengerjaan tugas akhir. Data alumni yang sudah terkumpul cukup banyak kurang begitu dimaksimalkan dan kurang diolah menjadi informasi yang lebih berguna, untuk mendapatkan informasi tersembunyi dari kumpulan data alumni yang sangat banyak dapat menggunakan teknologi data *mining*, sehingga hasil dari data mining bisa dimanfaatkan untuk membantu dalam mengambil kebijakan di perguruan tinggi seperti perubahan silabus ataupun perbaikan kurikulum [1][2].

Data *mining* merupakan proses pencarian pola-pola yang menarik dan tersembunyi dari suatu kumpulan data yang berukuran besar yang tersimpan dalam suatu basis data [3]. Salah satu metode yang diterapkan dalam data *mining* adalah *clustering*. Menurut [4] teknik *clustering* digunakan untuk mengelompokkan objek-objek yang memiliki kemiripan dalam kelas atau segmen yang sama. Pada penelitian ini metode yang akan digunakan dalam melakukan *cluster* data alumni adalah metode *fuzzy clustering* yaitu dengan algoritma *Fuzzy C-Means Clustering* (FCM) dan *Fuzzy Subtractive Clustering* (FS). Algoritma *Fuzzy C-Means Clustering* merupakan algoritma *supervised clustering* (jumlah *cluster* ditentukan). FCM memiliki tingkat akurasi yang tinggi dan waktu komputasi yang cepat [5], sedangkan *Fuzzy Subtractive Clustering* (FS) termasuk algoritma *unsupervised clustering* dimana jumlah *cluster* tidak ditentukan. Hasil pengelompokan data

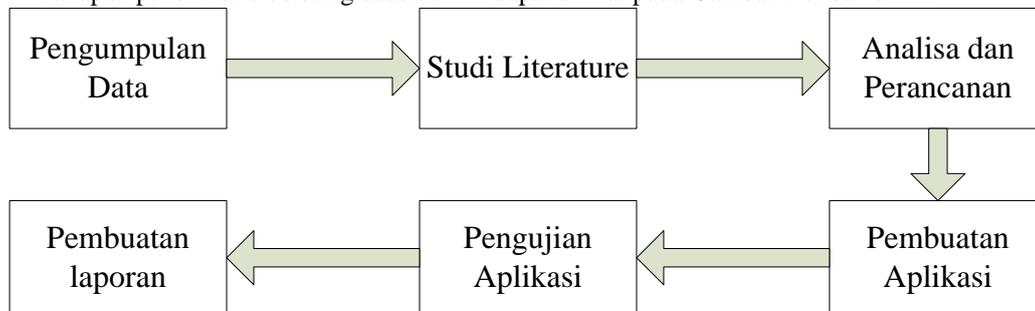
menggunakan kedua algoritma tersebut nantinya dapat dibandingkan. Atribut yang akan digunakan dalam melakukan *clustering* yaitu: Indeks Prestasi Kumulatif (IPK), lama *study* dan lama pengerjaan tugas akhir. Validitas partisi *fuzzy* akan diterapkan dalam proses *clustering*, metode yang akan digunakan untuk menguji validitas *cluster* pada penelitian ini adalah indeks *Modified Partition Coefficient* (MPC) dan *Classification Entropy* (CE).

Pemanfaatan data alumni dengan menggunakan teknik data *mining*, dimulai dari *selection* data pada system *existing* yang akan melewati proses *Extract-Transform-Load* (ETL) menjadi sebuah data yang siap untuk di *mining*, kemudian proses *mining* dilakukan dengan metode FCM dan FS. Hasil *clustering* kemudian dianalisa untuk menentukan karakteristik alumni. Data yang telah dianalisis disajikan pada stakeholder perguruan tinggi sehingga perguruan tinggi dapat mengambil keputusan dan langkah langkah untuk mengantisipasi hal-hal yang tidak harus terjadi. Dengan harapan perguruan tinggi STMIK STIKOM Indonesia dapat terus mengukur kualitas lulusan serta kualitas perguruan tinggi dari hasil pemanfaatan data alumni. Hasil penelitian ini diharapkan dapat membantu STMIK STIKOM Indonesia dalam melakukan pengambilan keputusan seperti peninjauan kurikulum ataupun perbaikan silabus dengan tujuan untuk meningkatkan kualitas lulusan.

2. Research Method

Objek penelitian adalah data alumni di STMIK STIKOM Indonesia, yang merupakan salah satu sekolah tinggi di bidang ilmu komputer di Bali. STMIK STIKOM Indonesia berdiri sejak tahun 2008 dengan mengelola dua (2) program studi, yaitu Teknik Informatika dan Sistem Komputer dimana keduanya merupakan pendidikan jenjang Sarjana (S1).

Tahapan penelitian clustering data alumni dapat dilihat pada Gambar 1 di bawah ini



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

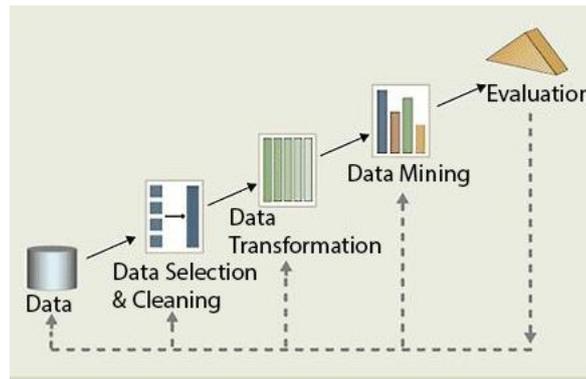
Teknik pengumpulan data yang digunakan yaitu observasi dan wawancara. Observasi dilakukan dengan cara menghubungi pihak yang menyediakan data alumni yang akan digunakan dalam penelitian dalam hal ini yaitu STMIK STIKOM Indonesia (STIKI Indonesia) yaitu dibagian alumni. Wawancara dilakukan secara langsung kepada bagian kemahasiswaan dan alumni. Observasi dan wawancara dilakukan untuk mengetahui pengumpulan dan cara pemrosesan data alumni yang selama ini dilakukan di STMIK STIKOM Indonesia. Jumlah data alumni yang digunakan dalam penelitian ini adalah sebanyak 1396 alumni

2.2. Studi Literature

Studi *Literature* dilakukan dengan membaca buku-buku *literatur* yang berkaitan dengan materi yang dipergunakan dalam penelitian. Materi yang digunakan dalam penelitian ini yaitu:

2.2.1 Data Mining

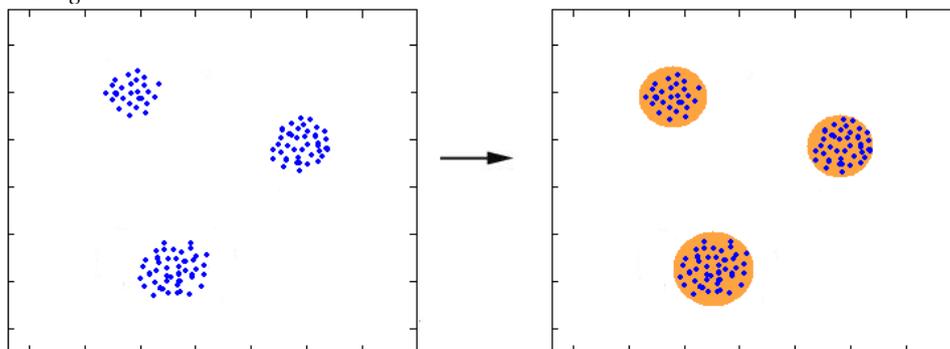
Data mining merupakan proses pencarian pola-pola yang menarik dan tersembunyi (*hidden pattern*) dari suatu kumpulan data yang berukuran besar yang tersimpan dalam suatu basis data, *data warehouse*, atau tempat penyimpanan data lainnya [6]. Proses data *mining* dapat dilihat pada Gambar 2 dibawah ini



Gambar 2 Proses *Data Mining*

2.2.2 Clustering

Pengelompokan (*clustering*) merupakan teknik yang sudah cukup dikenal dan banyak digunakan untuk mengelompokkan data/objek ke dalam kelompok data (*cluster*) sehingga setiap *cluster* memiliki data yang mirip dan berbeda dengan data yang berada dalam *cluster* lain. Dalam *conceptual clustering*, objek dikelompokkan berdasarkan kecocokannya menurut konsep deskriptif. Gambar 3 memperlihatkan contoh proses *clustering*.



Gambar 3 Proses *Clustering*

2.2.3 Fuzzy C-Means (FCM) dan Fuzzy Subtractive (FS)

Metode *Fuzzy C-Means Clustering* pertama kali dikenalkan oleh Jim Bezdek pada tahun 1981 [4]. *Fuzzy C-Means* adalah salah satu teknik pengelompokan data yang mana keberadaan tiap titik data dalam suatu kelompok (*cluster*) ditentukan oleh derajat keanggotaan. Metode *Fuzzy C-Means* termasuk metode *supervised clustering* dimana jumlah pusat *cluster* ditentukan di dalam proses *clustering*. Algoritma dari *fuzzy c-means* adalah sebagai berikut:

1. Input data yang akan dikelompokkan
2. Tentukan jumlah *cluster* (*c*), pangkat untuk matriks partisi (*w*), maksimum iterasi (*MaxIter*), error terkecil yang diharapkan (ξ), fungsi objektif awal ($P_0=0$), dan iterasi awal ($t=1$)
3. Bangkitkan bilangan random η_{ik} , $i=1,2,\dots,n$; $k=1,2,\dots,c$
4. Hitung pusat *cluster* dengan persamaan:

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w \cdot x_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \dots\dots\dots(1)$$

V_{kj} adalah pusat *cluster* ke-*k* untuk atribut ke-*j*, η_{ik} adalah derajat keanggotaan untuk data sampel ke-*i* pada *cluster* ke-*k*, x_{ij} adalah data ke-*i*, atribut ke-*j*

5. Hitung fungsi objektif pada iterasi ke-*t* dengan persamaan

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left(\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right) \dots\dots\dots(2)$$

V_{kj} adalah pusat *cluster* ke-*k* untuk atribut ke-*j*, η_{ik} adalah derajat keanggotaan, x_{ij} adalah data ke-*i*, atribut ke-*j*, P_t adalah fungsi objektif pada iterasi ke-*t*

6. Hitung perubahan matriks partisi menggunakan persamaan:

$$\mu_{ik} = \frac{[\sum_{j=1}^m (x_{ij} - v_{kj})^2]^{-1}}{\sum_{k=1}^c [\sum_{j=1}^m (x_{ij} - v_{kj})^2]^{-1}} \dots \dots \dots (3)$$

v_{kj} adalah pusat *cluster* ke- k untuk atribut ke- j , η_{ik} adalah derajat keanggotaan, x_{ij} adalah data ke- i , atribut ke- j

7. Cek Kondisi berhenti: Jika : $(|P_t - P_{t-1}| < \epsilon)$ atau $(t > MaxIter)$ maka berhenti. Jika tidak: $t = t + 1$, ulangi langkah ke-4

Dasar dari metode *Fuzzy Subtractive Clustering* adalah ukuran densitas (potensi) titik-titik data dalam suatu ruang (variabel). Konsep dasar dari metode *Fuzzy Subtractive Clustering* adalah menentukan daerah-daerah dalam suatu variabel yang memiliki densitas tinggi terhadap titik-titik di sekitarnya. Titik dengan jumlah tetangga terbanyak akan dipilih untuk menjadi pusat kelompok. Titik yang sudah dipilih menjadi pusat kelompok ini kemudian akan dikurangi densitasnya. Selanjutnya akan dipilih titik lain yang menjadi tetangga terbanyak untuk dijadikan pusat kelompok yang lain. densitas suatu titik dapat dihitung dengan persamaan:

$$D_k = \sum_{j=1}^n e^{-\left(\frac{\|x_k - x_j\|}{\left(\frac{r_a}{2}\right)^2}\right)} \dots \dots \dots (4)$$

D_k = Densitas titik ke- k , x_k = titik ke- k , r_a = konstanta positif.

Misalkan x_{c1} adalah titik yang terpilih menjadi pusat kelompok dan D_{c1} adalah ukuran densitasnya, selanjutnya densitas dari titik-titik di sekitarnya akan dikurangi dengan persamaan:

$$D_k = D_k - D_{c1} * e^{-\left(\frac{\|x_k - x_{c1}\|}{\left(\frac{r_b}{2}\right)^2}\right)} \dots \dots \dots (5)$$

Nilai r_b menunjukkan suatu lingkungan yang mengakibatkan titik-titik berkurang ukuran densitasnya. Nilai r_b diperoleh dari persamaan :

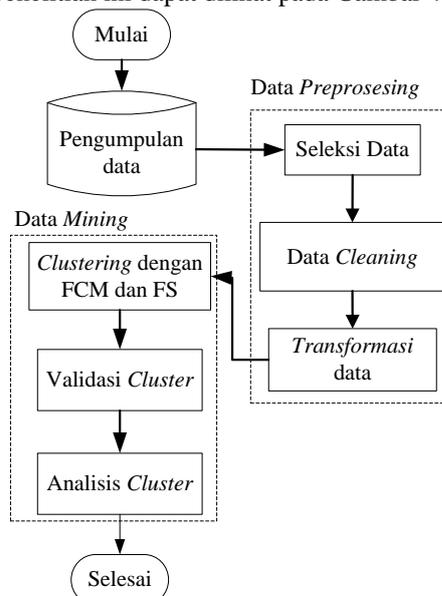
$$r_b = squashfactor * r_a \dots \dots \dots (6)$$

Biasanya *squashfactor* bernilai 1,5. Dengan demikian r_b bernilai lebih besar dibandingkan r_a .

2.3 Analisa dan Perancangan

Berdasarkan pengumpulan data yang telah didapatkan maka kebutuhan dari aplikasi *clustering* data alumni dengan *Fuzzy C-Means* dan *Fuzzy Subtractive* yaitu user memerlukan sebuah aplikasi yang mampu melakukan *clustering* data alumni dengan metode *Fuzzy C-Means* dan *Fuzzy Subtractive* serta dapat memberikan informasi yang dapat digunakan untuk melakukan analisa *clustering*.

Perancangan gambaran umum dari penelitian ini dapat dilihat pada Gambar 4 di bawah ini



Gambar 4 Gambaran Umum Sistem

Berdasarkan Gambar 4 dapat dijelaskan bahwa gambaran umum penelitian secara garis besar dibagi menjadi 3 tahapan yaitu pengumpulan data, data *preprocessing* dan tahapan data *mining*.

2.3.1 Pengumpulan Data

Pada tahapan ini yang dilakukan yaitu mengumpulkan data yang akan digunakan dalam penelitian ini. Data tersebut berjenis data sekunder yaitu data alumni STIKI Indonesia. Atribut data alumni yang akan digunakan dalam penelitian ini yaitu :Indeks Prestasi Kumulatif (IPK), lama studi (dalam semester), lama pengerjaan tugas akhir (dalam bulan). Pemilihan atribut ini dilakukan karena tidak semua data alumni dapat digunakan dalam proses *data mining* misalkan seperti alamat dan nomor telpon.

2.3.2 Data Preprocessing

Dalam tahapan ini terdapat tiga langkah yaitu seleksi data, data *cleaning* dan *transformasi* data. Tujuan dari tahapan ini adalah untuk mengubah data mentah menjadi data yang berkualitas, seperti memperbaiki data yang *incomplete* (tidak lengkap), *noisy* (data berisi kesalahan nilai) dan data tidak konsisten.

1. Seleksi Data

Pada tahapan seleksi data yang akan dilakukan yaitu melakukan seleksi atribut yang akan digunakan. Tentunya tidak semua atribut dimasukkan dalam set data yang digunakan dalam proses *data mining* karena hanya yang berperan sebagai *referensi* identifikasi yang akan dipilih. Dari semua atribut yang disebutkan diatas maka yang digunakan dalam data mining yaitu atribut Indeks Prestasi Kumulatif (IPK), Lama masa studi (dalam semester), Lama pengerjaan Tugas Akhir (TA) dalam bulan. Tabel 1 menunjukkan data yang telah mengalami seleksi data.

Tabel 1 Seleksi Data

NO	NIM	IPK	Lama Studi	Pengerjaan TA
1	08101011	2.99	10	3
2	08101014	3.14	10	4
3	08101018	3.46	10	10
4	08101023	3.26	10	11
5	08101024	2.59	10	11
6	08101026	3.1	8	6

2. Data Cleaning

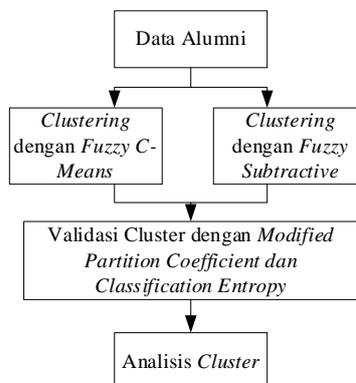
Tujuan dari proses ini adalah untuk memastikan kualitas data yang telah dipilih pada tahapan seleksi data. Proses data *cleaning* meliputi menghilangkan duplikasi data, memeriksa inkonsistensi data dan memperbaiki kesalahan pada data

3. Transformasi Data

Pada tahapan ini data yang akan dilakukan yaitu mengubah data yang telah dipilih sehingga data sesuai untuk proses data mining, misalkan atribut lama studi yang didapatkan yaitu dalam bentuk bulan maka diubah dijadikan dalam satuan semester hal ini dilakukan untuk memudahkan dalam melakukan analisa hasil.

2.3.3 Data Mining

Pada tahapan ini terdapat tiga proses yaitu melakukan *clustering* dengan *Fuzzy C-Means* (FCM) dan *Fuzzy Subtractive* (FS), *Validasi Cluster*, *Analisis Cluster*. Berikut ini merupakan proses *data mining* data alumni :



Gambar 5. Proses Data Mining

1. Clustering dengan *Fuzzy C-Means* dan *Fuzzy Subtractive*

Setelah data melewati *preprocessing* maka selanjutnya yaitu melakukan *clustering* dengan *Fuzzy C-Mean* dan *Fuzzy Subtractive*.

2. Validasi Cluster

Tujuan dari validasi *cluster* untuk menentukan jumlah *cluster* yang paling optimal dan memvalidasi apakah metode yang diterapkan sesuai dengan data. Proses *clustering* diuji coba dengan berbagai nilai parameter. Dalam penelitian ini menggunakan validasi *cluster* indeks *Modified Partition Coefficient* (MPC) (untuk mengukur *cluster* yang mengalami *overlap*) dan *Classification Entropy* (CE) (untuk mengukur tingkat kekaburan/*fuzziness* dari partisi *cluster*).

3. Analisis Cluster

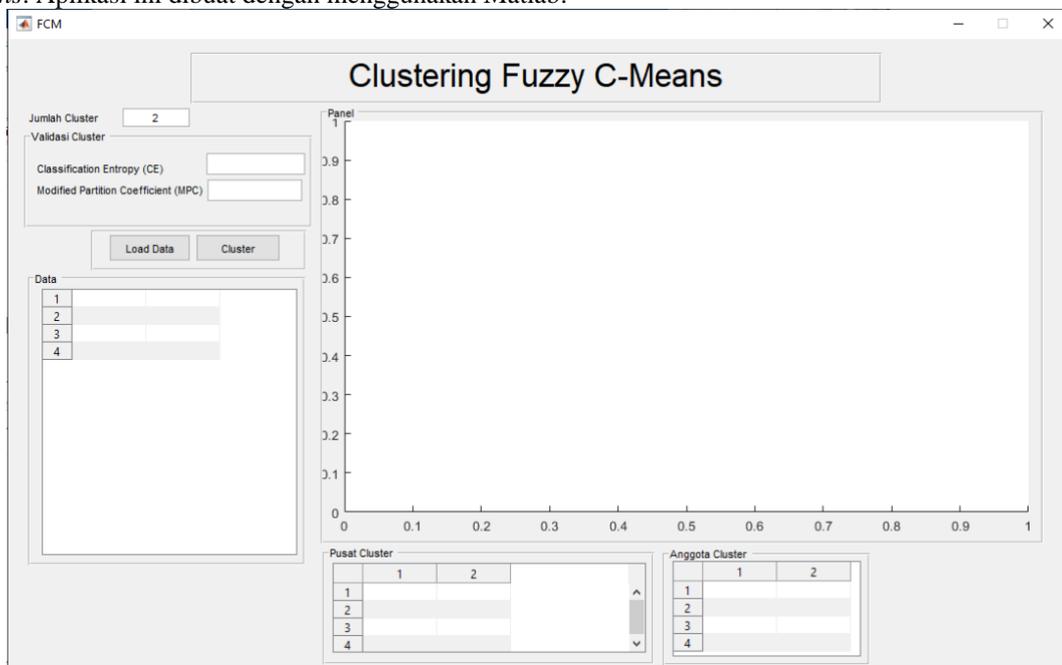
Setelah melewati proses data *mining* dan sudah divalidasi maka tahapan selanjutnya yaitu analisis *cluster*. Dalam tahapan ini yaitu melakukan analisis terhadap *cluster-cluster* yang telah terbentuk, hasilnya berupa karakteristik dari *profile* alumni STIKI Indonesia. Misalkan pada *cluster* pertama cenderung ditempati oleh IPK diatas 3,6 memiliki lama studi kurang dari 8 semester dan lama pengerjaan tugas akhir selama satu bulan.

3. Result and Analysis

Untuk mengetahui tingkat keberhasilan dari penelitian ini, dapat dilihat dari hasil kerja sistem yang telah dibuat. Berikut ini merupakan hasil dan pembahasan yang telah dilakukan.

3.1. Antar Muka Sistem Clustering

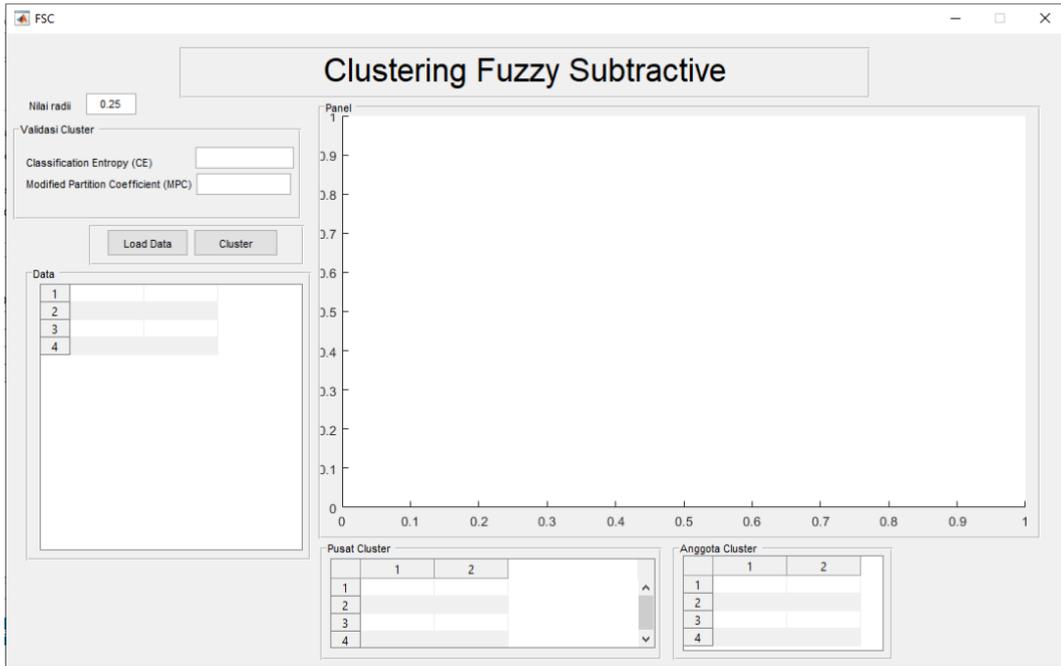
Pada Gambar 6 dapat dilihat tampilan dari aplikasi analisa clustering data alumni dengan *Fuzzy C-Means*. Aplikasi ini dibuat dengan menggunakan Matlab.



Gambar 6. Antar Muka Sistem Clustering

Pada aplikasi *clustering* dengan *Fuzzy C-Means* terdapat beberapa komponen yaitu Inputan jumlah cluster, dimana user bisa menentukan jumlah cluster yang akan terbentuk. Tombol *Load data* yang berfungsi untuk mengambil data alumni dan menampilkannya ditabel. Tombol *Cluster* yang berfungsi untuk melakukan clustering data alumni. *Output* dalam bentuk tabel data, tabel jumlah data dan tabel anggota cluster yang berfungsi untuk menampilkan data alumni, anggota cluster dan pusat *cluster*. *Output* dalam bentuk *Chart* data yang berfungsi untuk menampilkan persebaran data sebelum dan sesudah mengalami proses *clustering*. *Output* dalam bentuk *textboxt* untuk menampilkan nilai MPC dan CE.

Pada Gambar 7 dibawah ini dapat dilihat tampilan dari aplikasi *clustering* data alumni dengan *Fuzzy Subtractive*.



Gambar 7. Tampilan *Clustering dengan Fuzzy Subtractive*

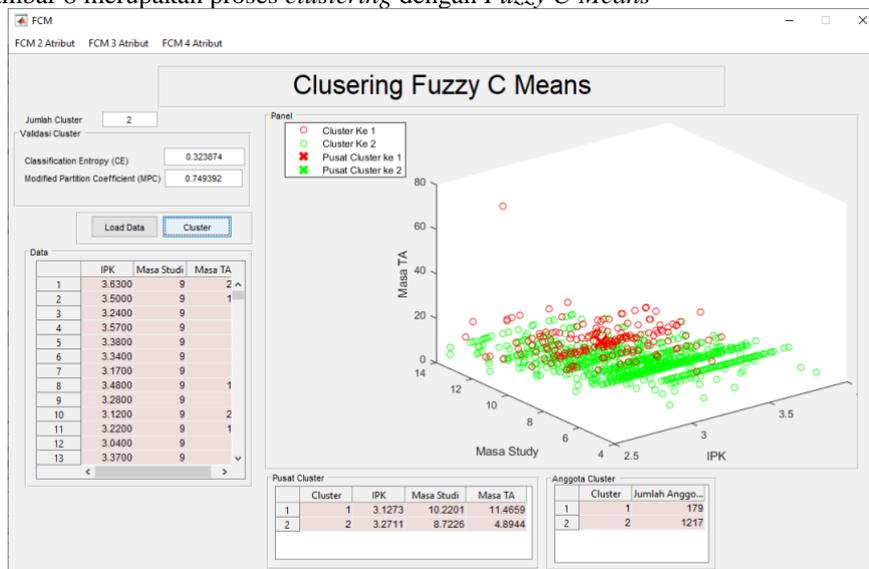
Pada aplikasi *clustering dengan Fuzzy Subtractive* komponen yang terdapat didalamnya kurang lebih sama dengan *clustering Fuzzy C Means*, perbedaannya yaitu kalau di *clustering dengan Fuzzy C-Means* jumlah cluster ditentukan sedangkan kalau di *clustering dengan Fuzzy Subtractive* jumlah *cluster* akan otomatis yang menjadi inputan yaitu nilai radii.

3.2. Clustering Data Alumni dengan Fuzzy C-Means

Proses *clustering* data alumni menggunakan algoritma *Fuzzy C-Means* diuji coba dengan berbagai jumlah *cluster*. Proses *clustering* dilakukan dengan menetapkan nilai awal sebagai berikut [7]:

1. Jumlah *cluster* = ditentukan *user*
2. Pangkat = 2
3. Maksimum iterasi = 100
4. *Error* terkecil yang diharapkan = 10^{-5}
5. Fungsi *objektif* awal = 0
6. Iterasi awal = 1

Pada Gambar 8 merupakan proses *clustering dengan Fuzzy C-Means*



Gambar 8. *Clustering dengan Fuzzy C-Means*

Berikut ini merupakan hasil dari proses *clustering* yang telah dilakukan yaitu:

1. *Clustering Alumni Jumlah Cluster 2*

Proses *clustering* dihentikan pada iterasi ke 22 dengan nilai fungsi objektif 8143,613437. Pada Tabel 2 dapat dilihat no *cluster*, jumlah anggota tiap *cluster* dan koordinat titik pusat setiap *cluster* yang terbentuk.

Tabel 2. Koordinat Titik Pusat *Cluster* (Jumlah *Cluster* 2)

No Cluster	Jumlah anggota	(IPK)	(Masa Studi)	(Masa TA)
1	179	3,1273	10,2201	11,4659
2	1217	3,2711	8,7226	4,8944

2. *Clustering Alumni Jumlah Cluster 3*

Proses *clustering* dihentikan pada iterasi ke 31 dengan nilai fungsi objektif 3363,726197. Pada Tabel 3 dapat dilihat no *cluster*, jumlah anggota tiap *cluster* dan koordinat titik pusat setiap *cluster* yang terbentuk.

Tabel 3. Koordinat Titik Pusat *Cluster* (Jumlah *Cluster* 3)

No Cluster	Jumlah anggota	(IPK)	(Masa Studi)	(Masa TA)
1	168	3,1354	10,0583	11,9801
2	394	3,0767	10,9853	4,4481
3	834	3,3379	7,9680	5,1642

3. *Clustering Alumni Jumlah Cluster 4*

Proses *clustering* dihentikan pada iterasi ke 23 dengan nilai fungsi objektif 4542,811561. Pada Tabel 4 dapat dilihat no *cluster*, jumlah anggota tiap *cluster* dan koordinat titik pusat setiap *cluster* yang terbentuk.

Tabel 4. Koordinat Titik Pusat *Cluster* (Jumlah *Cluster* 4)

No Cluster	Jumlah anggota	(IPK)	(Masa Studi)	(Masa TA)
1	196	3,0039	12,0157	4,5095
2	371	3,1708	9,5580	4,9587
3	661	3,3602	7,7876	5,1384
4	168	3,0039	12,0157	4,5095

4. *Clustering Alumni Jumlah Cluster 5*

Proses *clustering* dihentikan pada iterasi ke 26 dengan nilai fungsi objektif 2701.080562. Pada Tabel 5 dapat dilihat no *cluster*, jumlah anggota tiap *cluster* dan koordinat titik pusat setiap *cluster* yang terbentuk.

Tabel 5. Koordinat Titik Pusat *Cluster* (Jumlah *Cluster* 5)

No Cluster	Jumlah anggota	(IPK)	(Masa Studi)	(Masa TA)
1	514	3,3617	7,8102	4,9596
2	168	3,1359	10,0236	11,8562
3	127	2,9799	12,3572	4,6856
4	223	3,2937	8,2293	5,9303
5	364	3,1386	9,9882	4,7231

5. *Clustering Alumni Jumlah Cluster 6*

Proses *clustering* dihentikan pada iterasi ke 38 dengan nilai fungsi objektif 2170,806330. Pada Tabel 6 dapat dilihat no *cluster*, jumlah anggota tiap *cluster* dan koordinat titik pusat setiap *cluster* yang terbentuk.

Tabel 6. Koordinat Titik Pusat *Cluster* (Jumlah *Cluster* 5)

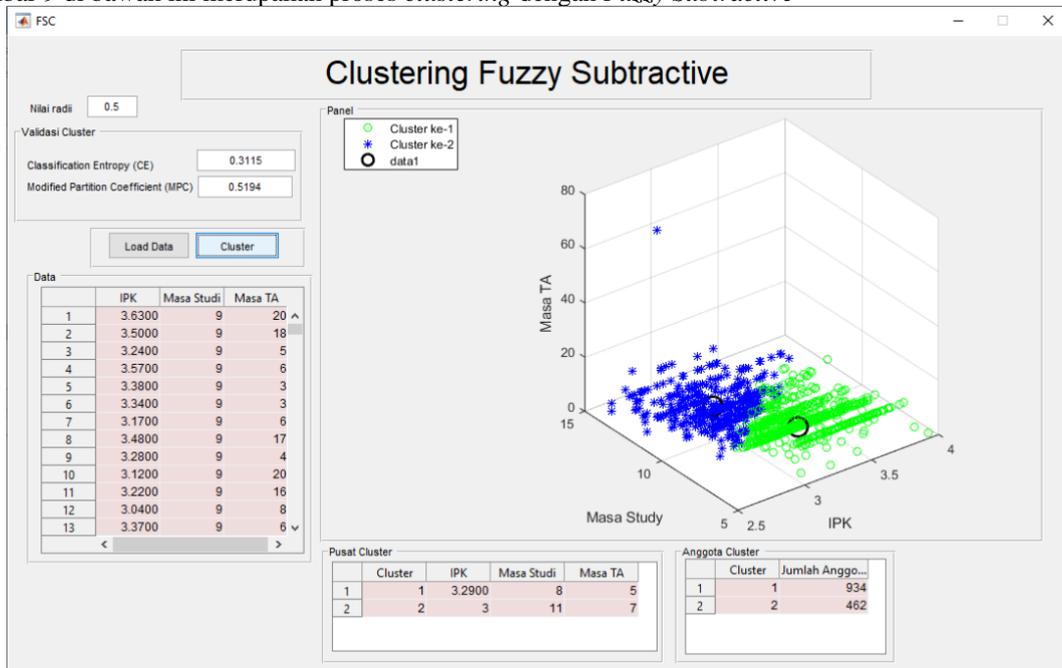
No Cluster	Jumlah anggota	(IPK)	(Masa Studi)	(Masa TA)
1	214	3,1250	10,2051	4,8680
2	514	3,3654	7,8001	4,9775
3	127	2,9702	12,5779	4,7937
4	159	3,1769	9,0652	4,8631
5	214	3,2981	8,0773	5,9811
6	168	3,1356	10,0151	11,7736

3.3. *Clustering Data Alumni dengan Fuzzy Subtractive*

Proses *clustering* menggunakan algoritma *Fuzzy Subtractive* diuji coba dengan berbagai nilai jari-jari atau *radii*. Proses *clustering* dilakukan dengan menetapkan nilai awal sebagai berikut[8]:

1. Jari-jari atau *radii*, biasanya nilai yang terbaik untuk *radii* berada antara 0,2 dan 0,5.
2. *Accept ratio* = 0,5
3. *Reject ratio* = 0,15
4. *Squash factor* = 1,25

Gambar 9 di bawah ini merupakan proses *clustering* dengan *Fuzzy Subtractive*



Gambar 9. *Clustering* dengan *Fuzzy Subtractive*

Berikut ini merupakan hasil dari proses *clustering* yang telah dilakukan yaitu:

1. *Clustering* dengan nilai *radii* 0,5

Uji coba *clustering* dengan nilai *radii* 0,5 menghasilkan 2 *cluster*. Proses *clustering* dengan nilai *radii* = 0,5 menghasilkan nilai *sigma* untuk ketiga dimensi data yaitu *sigma* IPK = 0,2440, *sigma* masa studi = 1,5910, *sigma* masa TA= 13,0815. Pada Tabel 7 dapat dilihat no *cluster*, jumlah anggota tiap *cluster* dan koordinat titik pusat setiap *cluster* yang terbentuk.

Tabel 7. Titik Koordinat Pusat *Cluster* (*Radii*=0,5)

No <i>Cluster</i>	Jumlah anggota	(IPK)	(Masa Studi)	(Masa TA)
1	934	3,2900	8	5
2	462	3	11	7

2. *Clustering* dengan nilai *radii* 0,4

Uji coba *clustering* dengan nilai *radii* 0,4 menghasilkan 3 *cluster*. Proses *clustering* dengan nilai *radii* = 0,4 menghasilkan nilai *sigma* untuk ketiga dimensi data yaitu *sigma* IPK = 0,1952, *sigma* masa studi = 1,2728 *sigma* masa TA= 10,4652. Pada Tabel 8 dapat dilihat no *cluster*, jumlah anggota tiap *cluster* dan koordinat titik pusat setiap *cluster* yang terbentuk.

Tabel 8. Titik Koordinat Pusat *Cluster* (*Radii*=0,4)

No <i>Cluster</i>	Jumlah anggota	(IPK)	(Masa Studi)	(Masa TA)
1	675	3,3000	8	5
2	549	3,0300	10	7
3	172	3,7400	7	5

3. *Clustering* dengan nilai *radii* 0,3

Uji coba *clustering* dengan nilai *radii* 0,3 menghasilkan 4 *cluster*. Proses *clustering* dengan nilai *radii* = 0,3 menghasilkan nilai *sigma* untuk ketiga dimensi data yaitu *sigma* IPK = 0.1464, *sigma* masa studi =

0.9546, σ masa TA= 7.8489. Pada Tabel 9 dapat dilihat no *cluster*, jumlah anggota tiap *cluster* dan koordinat titik pusat setiap *cluster* yang terbentuk.

Tabel 9. Titik Koordinat Pusat *Cluster* (*Radii*=0,3)

No <i>Cluster</i>	Jumlah anggota	(IPK)	(Masa Studi)	(Masa TA)
1	565	3,3000	8	5
2	391	3,0600	10	6
3	267	3,6200	8	5
4	173	2,8700	12	5

4. *Clustering* dengan nilai *radii* 0,25

Uji coba *clustering* dengan nilai *radii* 0,25 menghasilkan 5 *cluster*. Proses *clustering* dengan nilai *radii* = 0,25 menghasilkan nilai σ untuk ketiga dimensi data yaitu σ IPK= 0,1220, σ masa studi = 0,7955, σ masa TA = 6,5407. Pada Tabel 10 dapat dilihat no *cluster*, jumlah anggota tiap *cluster* dan koordinat titik pusat setiap *cluster* yang terbentuk.

Tabel 10. Titik Koordinat Pusat *Cluster* (*Radii*=0,25)

No <i>Cluster</i>	Jumlah anggota	(IPK)	(Masa Studi)	(Masa TA)
1	419	3,3000	8	5
2	333	3,0900	10	6
3	286	3,5900	8	5
4	179	3,0300	8	5
5	179	2,9000	12	5

5. *Clustering* dengan nilai *radii* 0,205

Uji coba *clustering* dengan nilai *radii* 0,205 menghasilkan 6 *cluster*. Proses *clustering* dengan nilai *radii* = 0,205 menghasilkan nilai σ untuk ketiga dimensi data yaitu σ IPK = 0,1000, σ masa studi = 0,6523, σ masa TA = 5,3634. Pada Tabel 11 dapat dilihat no *cluster*, jumlah anggota tiap *cluster* dan koordinat titik pusat setiap *cluster* yang terbentuk.

Tabel 11. Titik Koordinat Pusat *Cluster* (*Radii*=0,205)

No <i>Cluster</i>	Jumlah anggota	(IPK)	(Masa Studi)	(Masa TA)
1	382	3,3100	8	5
2	266	3,1000	9	5
3	211	3,5600	8	5
4	182	3,2400	10	7
5	277	2,9600	11	5
6	78	3,7500	7	5

3.3. Analisis *Cluster*

Dari pengujian yang telah dilakukan dengan menggunakan data alumni dari wisuda pertama sampai pada wisuda angkatan ke 10 dengan jumlah data 1396 maka dapat dianalisis beberapa hal. Untuk menentukan jumlah *cluster* yang memiliki nilai *fitness* terbaik digunakan indeks pengukuran validitas *Modified Partition Coeficient* (MPC) dan *Classification Entropy* (CE). Tabel 12. menunjukkan nilai indeks MPC dan CE untuk setiap jumlah *cluster* yang diujikan pada algoritma *Fuzzy C-Means* (FCM) dan *Fuzzy Subtractive* (FS) *Clustering*.

Tabel 12. Indeks MPC dan CE pada Algoritma FCM dan FS

Jumlah <i>Cluster</i>		2	3	4	5	6
Indeks	FCM	0,749392	0,6453	0,565366	0,789592	0,501937
MPC	FS	0,5194	0,5019	0,6014	0,7324	0,6346
Indeks	FCM	0,323874	0,614129	0,896933	1,11924	1,263
CE	FS	0,3315	0,6829	0,9432	0,2142	0,2832

Nilai indeks MPC untuk masing-masing jumlah *cluster* pada algoritma *Fuzzy C-Means* lebih besar dibandingkan *Fuzzy Subtractive Clustering*. Sedangkan nilai indeks CE untuk masing-masing jumlah *cluster* pada algoritma *Fuzzy C-Means* lebih kecil dibandingkan *Fuzzy Subtractive Clustering*. Dari hasil perbandingan tersebut algoritma *Fuzzy C-Means* memiliki kinerja yang lebih baik dibandingkan algoritma *Fuzzy Subtractive Clustering*.

Berdasarkan Tabel 12 juga dapat dilihat bahwa nilai indeks MPC yang terbesar untuk clustering dengan *Fuzzy C-Means* dan clustering dengan *Fuzzy Subtractive* berada pada jumlah *cluster* = 5. Untuk nilai indeks CE yang terkecil untuk clustering dengan *Fuzzy C-Means* dan clustering dengan *Fuzzy Subtractive* berada pada jumlah *cluster* = 5. Sehingga berdasarkan indeks CE (nilai terkecil) dan indeks MPC (nilai terbesar), jumlah *cluster* yang paling optimal untuk data alumni dengan 3 buah atribut (IPK, masa *study* dan masa TA) adalah 5 *cluster*. Karakteristik dari data berdasarkan Tabel 5 yang terdapat pada clustering dengan jumlah *cluster* 5 dengan metode *Fuzzy C-Means* yaitu *cluster* 1 yang jumlah anggotanya 514 orang memiliki rata-rata IPK 3,3617, masa studi rata-rata 7.8102 semester dan masa pengerjaan TA rata-rata 4,9596 bulan. *cluster* 2 yang jumlah anggotanya 168 orang memiliki rata-rata IPK 3,1359, masa studi rata-rata 10,0236 semester dan masa pengerjaan TA rata-rata 11,8562 bulan. *cluster* 3 yang jumlah anggotanya 127 orang memiliki rata-rata IPK 2,9799, masa studi rata-rata 12,3572 semester dan masa pengerjaan TA rata-rata 4,6856 bulan. *cluster* 4 yang jumlah anggotanya 223 orang memiliki rata-rata IPK 3,2937, masa studi rata-rata 8,2293 semester dan masa pengerjaan TA rata-rata 5,9303 bulan. *cluster* 5 yang jumlah anggotanya 364 orang memiliki IPK rata-rata 3,1386, masa studi rata-rata 9,9882 semester dan masa pengerjaan TA rata-rata 4,7231 bulan.

Jadi untuk clustering dengan 3 atribut (IPK, Masa Studi, Masa TA) jumlah *cluster* yang memiliki nilai *fitnes* terbaik/jumlah *cluster* yang paling optimal berdasarkan indeks validitas CE dan MPC yaitu adalah 5 *cluster*. *Cluster* yang memiliki karakteristik terbaik adalah *cluster* ke 1 yang memiliki anggota sebanyak 514 orang (36,82% dari total alumni).

4. Conclusion

Berdasarkan pengujian dan analisis yang telah dilakukan maka dapat ditarik simpulan sebagai berikut:

1. Hasil pengukuran validitas *cluster* menggunakan indeks Modified *Partition Coefficient* (MPC) dan *Classification Entropy* (CE), algoritma *Fuzzy C-Means Clustering* memiliki tingkat validitas lebih tinggi dibandingkan algoritma *Fuzzy Subtractive Clustering* sehingga dapat dikatakan bahwa algoritma *Fuzzy C-Means* melakukan proses *cluster* lebih baik dibandingkan dengan metode *Fuzzy Subtractive* dalam melakukan clustering data alumni. Jumlah *cluster* yang terbaik berdasarkan indeks MPC dan CE adalah 5.
2. Untuk clustering dengan 3 atribut (IPK, Masa Studi, Masa TA) jumlah *cluster* yang memiliki nilai *fitnes* terbaik/jumlah *cluster* yang paling optimal berdasarkan indeks validitas CE dan MPC yaitu adalah 5 *cluster*. *Cluster* yang memiliki karakteristik terbaik adalah *cluster* ke 1 yang dengan karakteristik memiliki rata-rata IPK 3,3617, masa studi rata-rata 7,8102 semester dan masa pengerjaan TA rata-rata 4,9596 bulan memiliki anggota sebanyak 514 orang (36,82% dari total 1396 alumni).

References

- [1]. Supartha, I Kadek Dwi Gandika; Sudarma, Made; Wiharta, Dewa Made. **Sistem Informasi Geografis Pemetaan Persebaran Alumni dengan Analisa Clustering**. Majalah Ilmiah *Teknologi Elektro*, [S.l.], v. 17, n. 3, p. 377-384, dec. 2018. ISSN 2503-2372.
- [2]. Iskandar, Adi Panca Saputra; Supartha I Kadek Dwi Gandika. **Sistem Informasi Geografis Sebaran Alumni Pada STMIK STIKOM Indoensia**. J-COSINE, Vol. 3, No.2 Desember 2019
- [3]. Tan, P.N., Steinbach, M. & Kumar, V. 2006. **Introduction to Data Mining**. Pearson Education, Inc.
- [4]. Jain, A. K., Murthy, M. N. & Flynn, P. J. (1999). **Data Clustering: A Review**. *ACM Computing Surveys*, Vol. 31, No. 3.
- [5]. Hammouda, K., Prof. Karaay, F. 2000. **A Comparative Study of Data Clustering Techniques**. University of Waterloo, Ontario, Canada.
- [6]. Tan, P.N., Steinbach, M. & Kumar, V. 2006. **Introduction to Data Mining**. Pearson Education, Inc.
- [7]. Yan, J., Power, J., 1994. **Using Fuzzy Logic (Toward Intelligent System)**. Prentice Hall, New York.
- [8]. Gelley, N., Jang, R. 2000. **Fuzzy Logic Toolbox**. Mathwork, Inc., USA.