

Pemanfaatan Algoritma FCM Dalam Pengelompokan Kinerja Akademik Mahasiswa

Aidina Ristyawan¹⁾, Kusri²⁾, Andi Sunyoto³⁾

STMIK AMIKOM Yogyakarta

Jl. Ringroad Utara, Condongcatur, Sleman, Yogyakarta 55283

e-mail: ristyakdr@gmail.com¹⁾, kusri@mikom.ac.id²⁾, andi@amikom.ac.id³⁾

Abstrak

Clustering merupakan salah satu teknik data mining yang secara otomatis mengelompokkan data yang memiliki keterkaitan tertentu secara bersama – sama. Salah satu metode clustering adalah Fuzzy C-Means yang didasarkan pada logika fuzzy, yang diperkenalkan oleh Lotfi Zadeh (1965) dengan nama himpunan fuzzy (Fuzzy Set). Algoritma clustering Fuzzy C-Means mengizinkan suatu data dapat menjadi anggota lebih dari satu kelompok. Pada penelitian ini dilakukan pengelompokan kinerja mahasiswa pada program studi S-1 Sistem Informasi STMIK AMIKOM Yogyakarta angkatan 2002 hingga 2006 yang sudah lulus berdasar nilai akhir matakuliah dan keahlian. Pengelompokan tersebut menggunakan algoritma fuzzy clustering yaitu Fuzzy C-Means. Pengelompokan tersebut dilakukan pada masing – masing keahlian, sehingga masing – masing keahlian tersebut akan terdapat lima kelompok kinerja akademik mahasiswa. Penggunaan pengelompokan menggunakan Fuzzy C-Means ini terbukti dapat dilakukan dengan baik yang mampu memiliki nilai fungsi objektif di bawah ambang batas dan memiliki index Xie Beni terendah.

Kata kunci: Clustering, fuzzy, Fuzzy C-Means

1. Pendahuluan

Kinerja seorang pegawai atau individu merupakan hal penting dalam sebuah organisasi dalam menjalankan usahanya. Seperti yang diungkapkan oleh Stephen P. Robbins dalam [1], tingkat kinerja pegawai sangat tergantung kedua faktor yaitu kemampuan pegawai itu sendiri, seperti tingkat pendidikan, pengetahuan, pengalaman, dimana dengan tingkat kemampuan yang semakin tinggi akan mempunyai kinerja semakin tinggi pula. Oleh karena itu kompetensi yang tepat dibutuhkan dalam melakukan pekerjaan yang tepat pula sesuai dengan prinsip *Right Man on the Right Place*. Tingkat kesesuaian pendidikan formal dengan pekerjaan / posisi juga merupakan indikator dari kemampuan kerja pegawai [1]. Pentingnya kompetensi terhadap kinerja dalam bekerja tersebut telah dibuktikan oleh beberapa penelitian sebelumnya [1],[2]. Tingkat kemampuan dari calon pegawai biasanya dapat diperkirakan berdasar kinerja akademik ketika masih kuliah. Kinerja akademik dapat ditentukan oleh nilai yang diperoleh saat kuliah.

Terdapat penelitian sebelumnya yang evaluasi kinerja akademik mahasiswa dengan menggunakan data mining. Pada penelitian tersebut peneliti mengevaluasi kinerja akademik mahasiswa pada tahun ke – 2 [3] Sedangkan pada penelitian lain bahwa Algoritma FCM juga digunakan untuk mengelompokkan nilai siswa digunakan sebagai dasar peminatan siswa SMA dan terbukti memiliki tingkat akurasi yang tinggi memiliki tingkat akurasi yang lebih tinggi [4]

Oleh sebab itu peneliti mencoba menggunakan teknik pengelompokan untuk mengetahui tingkat keahlian antara mahasiswa satu dengan mahasiswa lain ke dalam lima kelompok keahlian. Pengelompokan dapat dilakukan menggunakan algoritma FCM. Diharapkan penggunaan algoritma FCM pada penelitian ini dapat membagi kelompok – kelompok keahlian dengan baik.

2. Metode Penelitian

Data yang digunakan pada penelitian ini adalah data nilai akhir matakuliah mahasiswa yang sudah lulus dari jurusan S-1 Sistem Informasi STMIK AMIKOM Yogyakarta yang sekaligus sebagai objek penelitian, dengan angkatan mahasiswa mulai tahun 2002 hingga 2006, data matakuliah, data keahlian dan data mahasiswa. Kemudian dihitung tingkat validitas *fuzzy clustering* menggunakan Indeks Xie dan Beni (XBI).

2.1. Algoritma FCM

Clustering dengan metode *Fuzzy C-Means* (FCM) didasarkan pada teori logika *fuzzy*, diperkenalkan oleh Lotfi Zadeh (1965) dengan nama himpunan *fuzzy* (Fuzzy Set). Dalam teori fuzzy, keanggotaan sebuah dinyatakan dengan derajat nilai keanggotaan yang jangkauan nilainya 0 sampai 1. Nilai keanggotaan suatu data adalah 0 ketika sama sekali tidak menjadi anggota, dan 1 jika menjadi anggota secara penuh dalam suatu himpunan. Semakin tinggi nilai keanggotaannya maka semakin tinggi derajat keanggotaannya, begitu pula sebaliknya. FCM merupakan versi *fuzzy* dari K-Means dengan beberapa modifikasi yang membedakannya dengan K-Means[5].

Asumsikan ada sejumlah data dalam set data X yang berisi n data yang dinotasikan $X = \{x_1, x_2, \dots, x_n\}$ dimana setiap data mempunyai fitur r dimensi : $x_{i1}, x_{i2}, \dots, x_{ir}$, dinotasikan $x_i = \{x_{i1}, x_{i2}, \dots, x_{ir}\}$. Ada sejumlah cluster C dengan *centroid* c_1, c_2, \dots, c_k , di mana k adalah jumlah *cluster*. Setiap data mempunyai derajat keanggotaan pada setiap cluster, dinyatakan dengan u_{ij} , dengan nilai diantara 0 dan 1, i menyatakan data x_i dan j menyatakan *cluster* c_j . Jumlah nilai derajat keanggotaan setiap data x_i selalu sama dengan 1, yang diformulasikan dengan persamaan 1 [5] :

$$\sum_{j=1}^k u_{ij} = 1 \dots\dots\dots(1)$$

Dalam FCM setiap data juga menjadi anggota pada setiap cluster dengan derajat keanggotaan u_{ij} . Nilai derajat keanggotaan pada x_i pada cluster c_j , diformulasikan pada persamaan 2

$$u_{ij} = \frac{D(x_i, c_j)^{-\frac{2}{w-1}}}{\sum_{l=1}^k D(x_i, c_l)^{-\frac{2}{w-1}}} \dots\dots\dots(2)$$

Parameter c_j adalah *centroid* kluster ke j , $D(x_i, c_j)$ adalah jarak antara data dengan *centroid*, sedangkan w adalah bobot pangkat (*weighting exponent*) yang diperkenalkan dalam FCM. w tidak memiliki ketetapan, biasanya $w > 1$ dan umumnya diberi nilai 2. Nilai keanggotaan tersebut disimpan dalam matriks *fuzzy pseudo-partition* berukuran $N \times k$, dimana baris merupakan data, sedangkan kolom adalah nilai keanggotaan setiap cluster. Untuk menghitung *centroid* pada *cluster* c_i pada fitur j , digunakan persamaan 3.

$$c_{ij} = \frac{\sum_{i=1}^N (u_{ij})^w x_{ij}}{\sum_{i=1}^N (u_{ij})^w} \dots\dots\dots(3)$$

Secara prinsip, algoritma FCM banyak kesamaan dengan K-Means[5]. Secara ringkas algoritma clustering dengan menggunakan Fuzzy C-Means adalah sebagai berikut [5] :

- 1) Inisialisasi : tentukan jumlah cluster ($k > 2$), tentukan bobot pangkat ($w > 1$), tentukan jumlah maksimal iterasi, tentukan ambang batas perubahan nilai fungsi objektif (jika perlu juga perubahan nilai centroid);
- 2) Berikan nilai awal pada matriks fuzzy *pseudo-partition* dengan syarat sama seperti pada persamaan (1);
- 3) Hitung nilai centroid dari masing – masing cluster menggunakan persamaan (4);
- 4) Hitung kembali matriks fuzzy *pseudo-partition* (derajat keanggotaan setiap data pada setiap cluster) menggunakan persamaan (3). Ulangi langkah nomor 4 dan 5 selama syarat masih terpenuhi : 1) apabila perubahan pada nilai fungsi objektif masih di atas nilai ambang batas yang ditentukan; atau 2) perubahan pada nilai centroid masih di atas nilai ambang batas yang ditentukan; atau 3) iterasi maksimal belum tercapai.

2.2. Validitas clustering

Evaluasi validitas pengclusteran dilakukan dengan menggunakan index Xie and Beni atau biasa disebut index XBI. Secara umum, nilai yang terbaik XBI adalah nilai indeks yang semakin kecil, semakin kecil nilai XBI maka hasil pengelompokan semakin baik (Wu dan Yang, 2005) [5]. Rumus perhitungan XBI dapat diunjukkan seperti pada formula 4

$$XBI = \frac{\sum_{j=1}^k \sum_{i=1}^N u_{ij}^m \times d(x_i, c_j)^2}{N \times \min_{i,j} (d(c_i, c_j))^2} = \frac{J_m(u, c)/N}{Sep(c)} \dots\dots\dots(4)$$

$J_m(u, c)$ adalah ukuran kohesi, sedangkan $Sep(c)$ adalah ukuran separasi.

3. Hasil dan Pembahasan

Dalam penelitian ini dilakukan tiga tahapan proses penelitian, 1) pemetaan keahlian matakuliah, 2) pemetaan keahlian mahasiswa, 3) adalah pengclusteran keahlian mahasiswa.

Pemetaan keahlian matakuliah dilakukan secara manual tanpa algoritma tertentu berdasar indikator capaian keahlian mahasiswa yang tertera pada SAP Matakuliah. Adapun contoh pemetaan keahlian matakuliah terdapat pada Tabel. 1.

Tabel 1. Pemetaan keahlian matakuliah

MKL	KEAHLIAN
Kepribadian managerial	Leadership
Kepribadian managerial	Manajemen
Bahasa inggris I	Bahasa asing
Pengenalan pengolahan data elektronik	Menguasai bidang it
Pengenalan pengolahan data elektronik	IT
.....

Pemetaan keahlian mahasiswa adalah perhitungan rata-rata nilai akhir matakuliah yang dikelompokkan berdasarkan keahlian matakuliah per mahasiswa. Adapun contoh pemetaan keahlian mahasiswa terdapat pada tabel 2.

Tabel 2. Pemetaan keahlian mahasiswa

MHS	Animasi	Bahasa Asing	Bahasa Inggris	Broadcasting Dan Perfilman	Cisco	...
1	3,00	2,00	2,00	2,50	3,00	...
2	0,00	0,00	0,00	4,00	3,00	...
3	0,00	0,00	0,00	0,00	0,00	...
...

Berdasarkan hasil dari pemetaan keahlian diatas maka data keahlian mahasiswa tersebut dapat diclusterkan menggunakan algoritma Fuzzy C-Means dengan parameter sebagai berikut :

Iterasi maksimal	: 100 iterasi	Nilai ambang batas	: 0,05
Jumlah cluster	: 5 Cluster	Nilai fungsi objectif awal	: 1000
Bobot perpangkatan	: 2	Pengukuran jarak	: Euclidean

Pada penelitian ini bobot perpangkatan menggunakan 2 karena menurut Saiful dan Fajriya [6] berdasarkan penelitian Klawonn dan Keller nilai bobot perpangkatan yang paling optimal dan sering dipakai adalah 2. Ambil contoh data pengclusteran keahlian mahasiswa pada keahlian animasi dengan data seperti ditunjukkan pada tabel 3.

Tabel 3. Data pengclusteran keahlian

MHS	1	2	3	...	301
ANIMASI	3,00	0,00	0,00	...	0,00

Kemudian bangkitkan bilangan acak Fuzzy antara 0 – 1 sebagai Fuzzy Pseudo Partition dengan komposisi $M \times N$, Dimana M adalah banyak cluster sebagai kolom, sedangkan N Adalah banyak data sebagai baris

Setelah mendapatkan matrik Fuzzy Pseudo Partition, maka dapat dihitung centroid masing – masing cluster menggunakan persamaan 3, perhitungan centroid pada C1 ditunjukkan pada tabel 5.

Tabel 4. Perhitungan centroid

MHS	$(u_{il})^w$	$(u_{il})^w x_{ij}$	Centroid C1 = $\frac{101,624719}{135,0908058} = 1,329310498$
1.	0,185796834	0,557390503	
2.	0,23755508	0	
3.	0,239447242	0	
...	
301.	0,09669088	0	
Σ	101,624719	135,0908058	

Untuk perhitungan centroid lainnya dapat dilakukan dengan cara yang sama, sehingga didapatkan tabel centroid ditunjukkan pada tabel 6.

Tabel 5. Centroid

	C1	C2	C3	C4	C5
Centroid	1,329310498	1,473386871	0,751698434	1,426379359	1,518507505

Kemudian hitung jarak setiap data ke masing – masing centroid dengan menggunakan fungsi jarak Euclidean. Sebagai contoh perhitungan jarak data mahasiswa pertama ke centroid C1.

$$D(X_i, C1) = \sqrt{(X_i - C1)^2} \quad D(X_i, C1) = 1,670689502$$

$$D(X_i, C1) = \sqrt{(3 - 1,329310498)^2}$$

Penghitungan jarak antara data dengan centroid cluster lain dapat dilakukan dengan cara yang sama, dan dilakukan pada setiap 301 data mahasiswa, sehingga didapatkan tabel jarak pada tabel 7.

Tabel 6. Jarak ke centroid

MHS	jarak				
	C1	C2	C3	C4	C5
1	1,67069	1,526613	2,248302	1,573621	1,481492
2	1,32931	1,473387	0,751698	1,426379	1,518508
3	1,32931	1,473387	0,751698	1,426379	1,518508
...
301	1,32931	1,473387	0,751698	1,426379	1,518508

Kemudian dihitung derajat keanggotaan dengan fungsi derajat keanggotaan menggunakan persamaan formula 2.

$$u_{i1} = \frac{1,670692^{-2}}{1,670692^{-2} + 1,5266132^{-2} + 2,2483022^{-2} + 1,5736212^{-2} + 1,4814922^{-2}}$$

$$u_{i1} = 0,194222264$$

Perhitungan derajat keanggotaan terhadap cluster lain dapat dilakukan dengan cara yang sama. Dan dihitung pada masing – masing data. Sehingga terbentuk tabel derajat keanggotaan. Yang mana tabel tersebut dapat digunakan matrik Fuzzy Pseudo Partition pada iterasi selanjutnya. Dari hasil derajat keanggotaan tersebut sudah dapat dilihat pad masing – masing data mengikuti cluster yang mana berdasar derajat keanggotaan tertinggi, namun hal tersebut belum bisa dikatakan sudah selesai proses pengclusterannya. Proses cluster akan benar – benar selesai apabila selisih nilai fungsi objektif sebelumnya dengan fungsi objektif saat ini berada di bawah nilai ambang batas. Adapun untuk menghitung nilai fungsi objektif seperti pada formula 5.

$$J = \sum_{i=1}^N \sum_{j=1}^k (u_{ij})^w D(x_i, c_j)^2 \dots\dots\dots(5)$$

Adapun hasil dari nilai fungsi objektif terdapat pada tabel 9.

Tabel 7. Perhitungan nilai fungsi objektif

NO	C1	C2	C3	C4	C5	JUMLAH
1	0,105291	0,126102	0,05814	0,118681	0,1339	0,542114
2	0,040861	0,033261	0,127785	0,035489	0,031314	0,268709
3	0,040861	0,033261	0,127785	0,035489	0,031314	0,268709
...
301	0,040861	0,033261	0,127785	0,035489	0,031314	0,268709
Nilai Fungsi Objektif						143,2604

Diketahui nilai fungsi objektifnya adalah 143,2604, sehingga dapat dihitung selisih nilai fungsi objektif dengan sebelumnya yaitu : 1000 - 143,2604 = 856,7396. Hasil tersebut masih jauh diatas nilai ambang batas, maka perhitungan cluster diulangi lagi secara rekursif hingga nilai selisih fungsi objektif berada di bawah ambang batas. Dari perhitungan cluster animasi dibutuhkan 9 kali iterasi perhitungan cluster denan nilai fungsi objektif akhir adalah -5,6053. Adapun untuk hasil cluster ditunjukkan pada tabel 10.

Tabel 8. Hasil akhir cluster

MHS	1	2	3	301
U	4	2	1	1

Dari hasil cluster tersebut peneliti mengelompokan data nilai keahlian mahasiswa per kelompok cluster, dan dilihat range nilai tertinggi dan terendah dari nilai keahlian percluster sebagai batasan penentuan label cluster. Batasan penentuan label cluster tersebut tersaji dalam tabel 11.

Tabel 9. Batasan label cluster

KH	U	MINIMAL	MAXIMAL	LABEL_ANGKA	LABEL_HURUF
ANIMASI	3	4	4	1	A
ANIMASI	2	3,5	3,5	2	B
ANIMASI	4	3	3	3	C
ANIMASI	5	1	1	4	D
ANIMASI	1	0	0	5	E

Sehingga berdasarkan batasan penentuan label cluster tersebut, hasil pelabelan cluster pada keahlian "ANIMASI" ditunjukkan pada tabel 12.

Tabel 10. Pelabelan cluster

MHS	U	MINIMAL	MAXIMAL	LABEL_HURUF
1	4	3	3	C
2	1	0	0	E
3	1	0	0	E
...
301	1	0	0	E

Dari hasil cluster tersebut peneliti mengukur validitas menggunakan Index XBI. Untuk menghitung ukuran kohesi maka yang perlu dilakukan pertama adalah memangkatkan derajat keanggotaan data terhadap masing – masing cluster dengan parameter bobot. Kemudian ditentukan juga kuadrat dari jarak ke centroid, dan pemangkatan bobot derajat keanggotaan tersebut dikalikan dengan kuadrat jarak ke centroid, dan dijumlahkan sehingga menghasilkan ukuran kohesi, seperti ditunjukkan pada tabel 15.

Tabel 11. Perhitungan kohesi.

MHS	$u_{i1}^m \times d(x_i, c_1)^2$	$u_{i2}^m \times d(x_i, c_2)^2$	$u_{i3}^m \times d(x_i, c_3)^2$	$u_{i4}^m \times d(x_i, c_4)^2$	$u_{i5}^m \times d(x_i, c_5)^2$
1	5,40E-35	1,95E-33	4,86E-34	2,21E-17	1,22E-34
2	1,62E-21	2,14E-43	1,64E-43	2,91E-43	2,62E-42
3	1,62E-21	2,14E-43	1,64E-43	2,91E-43	2,62E-42
...
301	1,62E-21	2,14E-43	1,64E-43	2,91E-43	2,62E-42
Σ	2,77E-19	2,29E-12	2,31E-18	2,10E-15	2,68E-16
$J_m(u, c)$					2,29437543086286E-12

Dari sini telah ditemukan bahwa ukuran kohesi = 2,29437543086286E-12. Kemudian dilanjutkan dengan menentukan ukuran separasi dengan cara menghitung nilai terkecil dari kuadrat jarak centroid antara dua pasangan cluster seperti pada tabel 16.

Tabel 12. Kuadrat jarak centroid terkecil

CLUSTER	C1	C2	C3	C4	C5
CENTROID	4,02E-11	3,50E+00	4,00E+00	3,00E+00	1,00E+00
C1	4,02E-11	12,25	16	9	1
C2	3,500000677	12,25000474	0,249999	0,250001	6,250003
C3	4	16	0,249999	1	9
C4	3,000000005	9,000000028	0,250001	1	4
C5	1,000000016	1,000000033	6,250003	9	4

Didapatkan nilai terkecil adalah : 0,249999323. Sehingga perhitungan XBI menjadi :

$$\frac{2,29437543086286E-12/301}{0,249999323} = 3,04901E-14$$

Contoh hasil pengclusteran ditunjukkan pada tabel 17.

Tabel 13. Hasil cluster

MHS	ANIMASI	BAHASA	BAHASA	BROADCASTING	CISCO	...
-----	---------	--------	--------	--------------	-------	-----

		ASING	INGGRIS	DAN PERFILMAN	
1	C	C	E	A	E ...
2	C	C	B	E	E ...
3	E	C	A	B	E ...
...
301	B	C	D	A	E ...

Proses pengclustering untuk keahlian lain dapat dilakukan dengan cara yang sama. Adapun untuk menentukan jumlah cluster yang paling optimal diberikan ketika nilai index Xie dan Beni minimum pada lembah pertama. Peneliti telah mencoba mengclustering dengan jumlah cluster 2 hingga 8, dan hasilnya ditunjukkan pada tabel 18.

C	Iterasi	Nilai fungsi objektif	XBI
2	8	0,00022981	0,02276493
3	10	0,00850076	0,00133477
4	7	0,00187030	0,00059104
5	9	-5,60536854	3,04901E-14
6	7	0,0502019	6,85555E-14
7	7	4,93813484E-5	3,42543E-23
8	7	2,45973987	2,08322E-14

4. Simpulan

Proses pengclustering keahlian yang dilakukan pada penelitian ini terbukti dapat dilakukan dengan baik. Hal ini bisa ditunjukkan pada saat bahwa nilai fungsi objektif sudah berada di bawah ambang batas yang ditetapkan 0,05 yaitu : -5,6053. Dan untuk perhitungan index XBI pada salah satu keahlian yaitu “ANIMASI” mencapai 3,04901E-14 pada saat cluster berjumlah 5. Sehingga pada penelitian ini susunan pengclustering paling optimal adalah dengan jumlah cluster 5. Hasil dari penelitian ini dapat digunakan untuk melihat posisi atau keadaan kinerja akademik mahasiswa, dan dapat pula dikembangkan sebagai dasar pertimbangan dalam memilih profesi dalam penelitian berikutnya. Adapun saran pada penelitian ini adalah perlu dilakukannya pengklasifikasian matakuliah kedalam kelompok – kelompok keahlian. Sehingga diharapkan hasil pengclustering keahlian dapat lebih aktual.

Daftar Pustaka

- [1] Wahyuningrum. Hubungan Kemampuan, Kepuasan Dan Disiplin Kerja Dengan Kinerja Pegawai Di Kecamatan Tanggunharjo Kabupaten Grobogan. Tesis. Semarang Universitas Diponegoro; 2008.
- [2] Suwardji, E., Hasbullah, R., & Albatross, E. Hubungan Kompetensi dan Disiplin Kerja Terhadap Kinerja Tenaga Kependidikan Univeritas Singaperbangsa Karawan. *Jurnal Manajemen*. 2012;10(1) : 955.
- [3] Ridwan, M., Suyono, H., & Sarosa, M. Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Jurnal EECCIS*. 2013; 7 (1) : 59.
- [4] Bahar. Penentuan Jurusan Sekolah Menengah Atas Dengan Algoritma Fuzzy C-Means. Tesis. Semarang Universitas Dian Nuswantoro; 2011.
- [5] Prasetyo, E. Data Mining Mengolah data menjadi informasi menggunakan matlab. I. Yogyakarta. Penerbit ANDI Yogyakarta. 2014 : 218.
- [6] Rizal, A.S., Hakim, R.B.F. Metode K-Means Cluster Dan Fuzzy C-Means Cluster (Studi Kasus: Indeks Pembangunan Manusia di Kawasan Indonesia Timur tahun 2012). *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika UMS*. 2015 : 643