

FUZZY C-MEANS UNTUK CLUSTERING DATA (STUDI KASUS : DATA PERFORMANCE MENGAJAR DOSEN)

Emha Taufiq Luthfi

STMIK AMIKOM Yogyakarta

e-mail : emha_tl@yahoo.com

Abstraksi

Clustering merupakan proses pengelompokan data dalam kelas-kelas atau cluster-cluster sehingga data dalam suatu cluster memiliki tingkat persamaan yang tinggi satu dengan lainnya tetapi sangat berbeda dengan data dalam cluster lain. Dalam tulisan ini dilakukan percobaan penggunaan metode Fuzzy C-Means untuk mengetahui kemungkinan adanya cluster-cluster dari data performance mengajar dosen.

Kata kunci : Clustering, Fuzzy C-Means, Data Performance Mengajar Dosen

1. PENDAHULUAN

Untuk evaluasi proses pengajaran oleh dosen selama satu semester, salah satu cara dengan meminta mahasiswa untuk memberikan penilaian terhadap performance dosen dalam mengampu suatu mata kuliah selalu satu semester dalam beberapa kriteria penilaian. Kriteria yang digunakan antara lain :

- N1 : Penguasaan dan kemampuan dalam menjelaskan materi
- N2 : Kemampuan dalam menjawab pertanyaan
- N3 : Kemampuan dalam memberi motivasi mahasiswa
- N4 : kemampuan membuat suasana kelas menyenangkan
- N5 : Kedisiplinan hadir dalam perkuliahan

Dari nilai kelima kriteria tersebut, didapat nilai indeks prestasi kumulatif dosen dalam mengampu suatu mata kuliah selama satu semester.

Seorang dosen dalam satu semester mungkin harus mengampu lebih dari satu mata kuliah. Atau satu mata kuliah mungkin harus diampu oleh lebih dari seorang dosen.

Proses clustering yang dilakukan, akan mencoba mengetahui kemungkinan adanya kelompok-kelompok yang mungkin belum diketahui dari data performance mengajar dosen terhadap suatu mata kuliah.

2. FUZZY C-MEANS

Fuzzy C-means Clustering (FCM), atau dikenal juga sebagai Fuzzy ISODATA, merupakan salah satu metode clustering yang merupakan bagian dari metode Hard K-Means. FCM menggunakan model pengelompokan fuzzy sehingga data dapat menjadi anggota dari semua kelas atau cluster terbentuk dengan derajat atau tingkat keanggotaan yang berbeda antara 0 hingga 1.

Tingkat keberadaan data dalam suatu kelas atau cluster ditentukan oleh derajat keanggotaannya. Teknik ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981.

Konsep dasar FCM, pertama kali adalah menentukan pusat cluster yang akan menandai lokasi rata-rata untuk tiap-tiap cluster. Pada kondisi awal, pusat cluster ini masih belum akurat. Tiap-tiap data memiliki derajat keanggotaan untuk tiap-tiap cluster. Dengan cara memperbaiki pusat cluster dan nilai keanggotaan tiap-tiap data secara berulang, maka dapat dilihat bahwa pusat cluster akan menuju lokasi yang tepat. Perulangan ini didasarkan pada minimasi fungsi obyektif (Gelley, 2000).

Fungsi obyektif yang digunakan FCM adalah (Ross, 2005) :

$$J(U, V; X) = \sum_{k=1}^n \sum_{i=1}^C (\mu_{ik})^w (d_{ik})^2$$

dengan $w \in [1, \infty]$,

$$d_{ik} = d(x_k - v_i) = \left[\sum_{j=1}^m (x_{kj} - v_{ij})^2 \right]^{1/2}$$

X adalah data yang dicluster :

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}$$

dan V adalah matriks pusat cluster :

$$V = \begin{bmatrix} v_{11} & \cdots & v_{1m} \\ \vdots & & \vdots \\ v_{c1} & \cdots & v_{cm} \end{bmatrix}$$

Nilai J_w terkecil adalah yang terbaik, sehingga :

$$J_w^*(U^*, V^*; X) = \min_{M_{fc}} J(U, V, X)$$

Algoritma FCM secara lengkap diberikan sebagai berikut (Zimmerman, 1991); (Yan, 1994); (Ross, 2005) :

1. Tentukan :
 - a. Matriks X berukuran $n \times m$, dengan n = jumlah data yang akan di cluster; dan m = jumlah variabel (kriteria).
 - b. Jumlah cluster yang akan dibentuk ($C \geq 2$)
 - c. Pangkat (pembobot $w > 1$)
 - d. Maksimum iterasi
 - e. Kriteria penghentian (ϵ = nilai positif yang sangat kecil)
2. Bentuk matriks partisi awal U (derajat keanggotaan dalam cluster); matriks partisi awal biasanya dibuat secara acak

$$U = \begin{bmatrix} \mu_{11}(x_1) & \mu_{12}(x_2) & \cdots & \mu_{1n}(x_n) \\ \mu_{21}(x_1) & \mu_{22}(x_2) & \cdots & \mu_{2n}(x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{c1}(x_1) & \mu_{c2}(x_2) & \cdots & \mu_{cn}(x_n) \end{bmatrix}$$

3. Hitung pusat cluster V untuk setiap cluster

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

4. Perbaiki derajat keanggotaan setiap data pada setiap cluster (perbaiki matriks partisi)

$$\mu_{ik} = \left[\sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}} \right)^{2/(w-1)} \right]^{-1}$$

dengan :

$$d_{ik} = d(x_k - v_i) = \left[\sum_{j=1}^m (x_{kj} - v_{ij})^2 \right]^{1/2}$$

5. Tentukan kriteria penghentian iterasi, yaitu perubahan matriks partisi pada iterasi sekarang dan iterasi sebelumnya

$$\Delta = \|U^t - U^{t-1}\|$$

Apabila $\Delta < \varepsilon$ maka iterasi dihentikan.

3. PERCOBAAN

Clustering dengan metode FCM terhadap data performance mengajar dosen di suatu semester dilakukan untuk mengetahui cluster-cluster yang mungkin ada dan memiliki pola tertentu dengan parameter sebagai berikut :

1. Matriks input satu dimensi berupa nilai indeks prestasi mengajar seorang dosen pada mata kuliah tertentu pada program studi tertentu.
2. Jumlah cluster yang dicoba dibentuk = $C = 4$
3. Pangkat (pembobot) = $w = 2$
4. Maksimum Iterasi = 1000
5. Kriteria penghentian = $\varepsilon = 10^{-6}$

Implementasi dilakukan dengan menggunakan software matlab.

4. HASIL

Penghitungan pusat cluster dan perbaikan nilai keanggotaan akan dihentikan pada iterasi ke 42. Pada iterasi tersebut, nilai $\Delta = 7.8161e-007$.

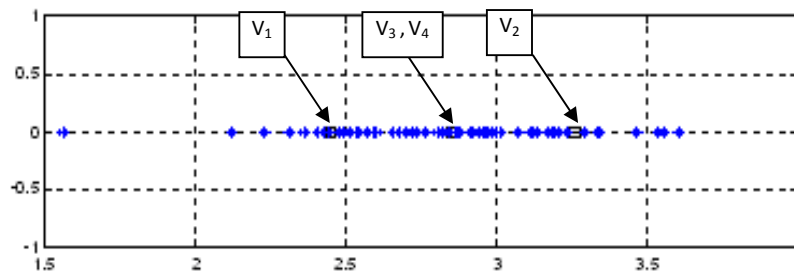
Nilai Pusat cluster adalah $V = [2.4485 \quad 3.2564 \quad 2.8508 \quad 2.8508]$

Nilai pusat cluster 3 dan 4 memiliki nilai sama, sehingga terbentuk hanya 3 cluster. Pada Percobaan lain mungkin didapat hasil yang sedikit berbeda, dikarenakan inisialisasi awal matriks partisi yang dilakukan secara acak.

Mata Kuliah	Dosen	Jenjang	IP_MK	Kecenderungan Cluster			
				c1	c2	c3	c4
Akuntansi I	I	D3	2.87	0.001	0.0012	0.4989	0.4989
Akuntansi I	SDMM	S1 REG	2.41	0.9829	0.002	0.0075	0.0075
Aljabar Linier	E	D3	2.68	0.2068	0.0333	0.3799	0.3799
Aljabar Linier	RJ	D3	3.20	0.0053	0.9454	0.0247	0.0247
Aljabar Linier	SM	S1 REG	2.76	0.0401	0.0158	0.4721	0.4721
Analisis Kinerja Sistem	MDI	S1 TR	1.56	0.45	0.1235	0.2133	0.2133
Analisis Kinerja Sistem	MDI	S1 REG	2.12	0.6721	0.0562	0.1359	0.1359
Analisis Laporan Keuangan	SH	S1 REG	2.80	0.0103	0.0061	0.4918	0.4918
Analisis Laporan Keuangan	SH	D3	2.87	0.001	0.0012	0.4989	0.4989
Analisis Sistem Informasi	HAF	D3	3.01	0.0322	0.1672	0.4003	0.4003
Analisis Sistem Informasi	MHT	S1 REG	3.11	0.029	0.5928	0.1891	0.1891
Analisis Sistem Informasi	SMT	S1 REG	2.59	0.6122	0.0276	0.1801	0.1801
Analisis Sistem Informasi	WW	D3	2.94	0.0156	0.0376	0.4734	0.4734
Bahasa Inggris II	TS	D3	3.55	0.0499	0.7024	0.1238	0.1238
Bahasa Inggris II	TS	S1 REG	3.60	0.059	0.6624	0.1393	0.1393
Broadcasting TV	MSMM	S1 TR	2.86	0.0003	0.0003	0.4997	0.4997
Broadcasting TV	MSMM	S1 REG	2.91	0.0081	0.0143	0.4888	0.4888
Broadcasting TV	MSMM	S1 REG	2.98	0.026	0.096	0.439	0.439
E-Business	MSMM	S1 TR	2.97	0.0235	0.0779	0.4493	0.4493
E-Commerce	J	S1 REG	2.49	0.9715	0.0028	0.0128	0.0128
E-Commerce	J	D3	2.72	0.1013	0.0259	0.4364	0.4364
Interaksi Manusia dan Komputer	PSM	S1 TR	3.12	0.0265	0.6432	0.1651	0.1651
Interaksi Manusia dan Komputer	PSM	S1 REG	3.29	0.0016	0.9869	0.0058	0.0058
Jaringan Komputer	AP	D3	2.51	0.933	0.0063	0.0303	0.0303
Jaringan Komputer	MS	S1 REG	3.46	0.0321	0.7912	0.0884	0.0884
Keamanan Komputer	MD	S1 TR	3.00	0.0304	0.1404	0.4146	0.4146
Kepemimpinan	ASSH	D3	2.66	0.2791	0.0351	0.3429	0.3429
Kepemimpinan	MI	D3	2.82	0.0034	0.0025	0.4971	0.4971
Kepemimpinan	SPDMM	S1 REG	2.23	0.7733	0.0351	0.0958	0.0958
Kewirausahaan	AMT	S1 REG	2.36	0.9304	0.0091	0.0303	0.0303
Kewirausahaan	AMT	S1 REG	2.73	0.0823	0.0235	0.4471	0.4471
Kewirausahaan	AMT	S1 REG	2.76	0.0401	0.0158	0.4721	0.4721

Mata Kuliah	Dosen	Jenjang	IP_MK	Kecenderungan Cluster			
				c1	c2	c3	c4
Kewirausahaan	KP	D3	2.66	0.2791	0.0351	0.3429	0.3429
Kewirausahaan	KP	D3	2.70	0.1478	0.0302	0.411	0.411
Komputer Grafis	AFS	S1 REG	2.70	0.1478	0.0302	0.411	0.411
Komputer Grafis	AY	D3	2.87	0.001	0.0012	0.4989	0.4989
Komunikasi Data	K	D3	2.88	0.0023	0.003	0.4974	0.4974
Komunikasi Data	K	S1 REG	2.95	0.0183	0.0489	0.4664	0.4664
Komunikasi Data	RMAKR	D3	2.76	0.0401	0.0158	0.4721	0.4721
Komunikasi Data	RMAKR	S1 REG	3.01	0.0322	0.1672	0.4003	0.4003
Manajemen Strategik	FA	S1 TR	3.19	0.0074	0.922	0.0353	0.0353
Matematika Diskret	SM	S1 TR	2.60	0.5609	0.0299	0.2046	0.2046
Metodelogi Penelitian	HS	D3	3.07	0.0355	0.3943	0.2851	0.2851
Metodelogi Penelitian	ES	S1 REG	2.43	0.9956	0.0005	0.0019	0.0019
Pemp. Basis Data	AA	D3	2.98	0.026	0.096	0.439	0.439
Pemp. Basis Data	AS	D3	2.76	0.0401	0.0158	0.4721	0.4721
Pemp. Basis Data	ETL	S1 REG	2.87	0.001	0.0012	0.4989	0.4989
Pemp. Basis Data	MRA	S1 REG	2.57	0.7115	0.0223	0.1331	0.1331
Pemp. Berorientasi Objek II	AB	D3	2.44	0.999	0.0001	0.0004	0.0004
Pemp. Berorientasi Objek II	ASMT	D3	2.53	0.876	0.011	0.0565	0.0565
Pemp. Berorientasi Objek II	WW	D3	3.00	0.0304	0.1404	0.4146	0.4146
Pemp. Berorientasi Obyek II	KMT	S1 REG	2.53	0.876	0.011	0.0565	0.0565
Pemp. Terstruktur	AA	S1 REG	2.92	0.0104	0.0205	0.4845	0.4845
Pemp. Terstruktur	EHS	D3	3.53	0.0461	0.7202	0.1169	0.1169
Pemp. Terstruktur	HP	S1 REG	2.54	0.8407	0.0137	0.0728	0.0728
Pengelolaan Instalasi Komputer	RW	D3	3.19	0.0074	0.922	0.0353	0.0353
Pengelolaan Instalasi Komputer	MHP	S1 REG	2.91	0.0081	0.0143	0.4888	0.4888
Pengelolaan Proyek SI	MHT	S1 TR	3.18	0.0098	0.894	0.0481	0.0481
Pengelolaan Proyek SI	MHT	S1 REG	3.07	0.0355	0.3943	0.2851	0.2851
Perancangan Multimedia	MSMM	D3	2.96	0.0209	0.0623	0.4584	0.4584
Perancangan Multimedia	MSMM	S1 REG	3.23	0.0011	0.9893	0.0048	0.0048
Perancangan Multimedia	TI	D3	2.84	0.0004	0.0003	0.4996	0.4996
Perancangan Multimedia	TI	S1 REG	2.95	0.0183	0.0489	0.4664	0.4664
Perilaku Organisasi	AR	S1 REG	3.18	0.0098	0.894	0.0481	0.0481
Perilaku Organisasi	BS	D3	3.34	0.0082	0.937	0.0274	0.0274
Rekayasa Perangkat Lunak	SP	S1 TR	2.48	0.9842	0.0016	0.0071	0.0071

Mata Kuliah	Dosen	Jenjang	IP_MK	Kecenderungan Cluster			
				c1	c2	c3	c4
Sistem Informasi Akuntansi	SH	S1 REG	2.70	0.1478	0.0302	0.411	0.411
Sistem Informasi Akuntansi	SH	D3	2.88	0.0023	0.003	0.4974	0.4974
Sistem Pakar	HAF	S1 REG	3.17	0.0124	0.8614	0.0631	0.0631
Sistem Pakar	HAF	S1 TR	3.33	0.0066	0.9486	0.0224	0.0224
Sistem Penunjang Keputusan	MD	S1 REG	3.18	0.0098	0.894	0.0481	0.0481
Sistem Penunjang Keputusan	SP	S1 TR	2.31	0.8675	0.0186	0.0569	0.0569
Sistem Penunjang Keputusan	SP	S1 REG	2.54	0.8407	0.0137	0.0728	0.0728
Statistik	MCM	D3	3.13	0.0238	0.6924	0.1419	0.1419
Statistik	S	S1 REG	2.96	0.0209	0.0623	0.4584	0.4584
Statistik	SS	D3	2.72	0.1013	0.0259	0.4364	0.4364
Statistik	ZM	S1 REG	2.57	0.7115	0.0223	0.1331	0.1331
Testing dan Implementasi Sistem	SR	S1 REG	2.66	0.2791	0.0351	0.3429	0.3429
Testing dan Implementasi Sistem	SR	S1 TR	2.82	0.0034	0.0025	0.4971	0.4971



Gambar 1 Hasil FCM pada Iterasi akhir

Cluster c1 didapat untuk nilai matriks input (IPK) antara 1.56 sampai dengan 2.60. Cluster c2 didapat untuk nilai matriks input (IPK) antara 3.07 sampai dengan 3.60. Sedangkan cluster c3 dan c4 untuk nilai matriks input (IPK) 2.66 sampai dengan 3.01.

Contoh hal yang dapat dilihat misal, untuk mata kuliah Pemrograman Basis Data yang diampu oleh 4 orang dosen memiliki kecenderungan berada pada cluster c3,c4 dengan derajat keanggotaan yang rendah serta 1 dosen berada pada cluster c1 dengan derajat keanggotaan cukup tinggi.

Begitu pula mata kuliah Pemrograman Berorientasi Obyek 2, untuk 4 orang dosen pengampu memiliki kecenderungan berada pada cluster c3,c4 serta cluster c1.

Atau mungkin dapat diamati mata kuliah atau dosen yang memiliki kecenderungan berada di cluster c2.

Untuk percobaan lebih lanjut, perlu dicoba clustering dengan lebih dari 1 variabel input. Semisal, untuk kasus data performance dosen dapat dilakukan clustering dengan kombinasi input kriteria penilaian N1, N2, N3, N4 dan N5.

5. KESIMPULAN

Clustering merupakan proses pengelompokan obyek atau data tidak berlabel kedalam suatu kelas atau cluster dengan obyek yang memiliki kesamaan. Clustering dengan menggunakan metode Fuzzy C-Means terhadap data performance mengajar dosen, dapat memunculkan beberapa cluster data yang dapat dianalisa lebih lanjut persamaan dan perbedaannya.

6. DAFTAR PUSTAKA

Lin, Ching-Teng; Lee, George. 1996. *Neural Fuzzy Systems : A Neuro-Fuzzy Synergism to Intelligent Systems*. United States of America. Prentice Hall International Inc.

Kusumadewi, Sri; Hartati, Sri. 2006. *Neuro-Fuzzy : Integrasi Sistem Fuzzy dan Jaringan Syaraf*. Yogyakarta. Graha Ilmu.

http://www.cs.bris.ac.uk/home/tr1690/documentation/fuzzy_clustering_initial_report/node11.html