

CLUSTERING LULUSAN MAHASISWA MATEMATIKA FMIPA UNTAN PONTIANAK MENGGUNAKAN ALGORITMA FUZZY C-MEANS

Cary Lineker Simbolon, Nilamsari Kusumastuti, Beni Irawan

INTISARI

Clustering adalah proses pengelompokan data ke dalam cluster berdasarkan parameter tertentu sehingga obyek-obyek dalam sebuah cluster memiliki tingkat kemiripan yang tinggi satu sama lain dan sangat tidak mirip dengan obyek yang lain pada cluster yang berbeda. Fuzzy C-Means termasuk dalam salah satu teknik clustering. Seperti teknik clustering lainnya. Fuzzy C-Means juga mengelompokkan sejumlah obyek. Pada jurnal ini teknik Fuzzy C-Means digunakan untuk mengelompokkan lulusan jurusan Matematika FMIPA Universitas Tangjungpura (UNTAN). Lulusan dibagi kedalam empat cluster berdasarkan IPK dan lama studi. Hasil dari penelitian ini adalah diperoleh empat pusat cluster atau center. Untuk cluster 1 terdiri dari lulusan dengan kisaran IPK 3,15 dan lama studi 5,09 tahun, cluster 2 terdiri dari lulusan dengan kisaran IPK 2,88 dan lama studi 7,32 tahun, cluster 3 terdiri dari lulusan dengan kisaran IPK 3,48 dan lama studi 4,37 tahun serta cluster 4 terdiri dari lulusan dengan kisaran IPK 2,89 dan lama studi 5,91 tahun. Lulusan yang paling banyak anggotanya ada pada cluster 4. Hasil yang diperoleh menunjukkan bahwa masih banyak lulusan jurusan Matematika di Fakultas MIPA UNTAN yang menempuh lama studi lebih dari 5 tahun.

Kata Kunci : *Fuzzy, Fuzzy C-Means, clustering*

PENDAHULUAN

Proses *clustering* adalah proses pengelompokan data ke dalam *cluster* berdasarkan parameter tertentu sehingga obyek-obyek dalam sebuah *cluster* memiliki tingkat kemiripan yang tinggi satu sama lain dan sangat tidak mirip dengan obyek yang lain pada *cluster* yang berbeda [1]. Pembentukan *cluster* data merupakan salah satu teknik yang digunakan dalam mengetahui pola kecenderungan suatu data. Analisis *cluster* merupakan proses membagi data dalam suatu himpunan ke dalam beberapa kelompok yang kesamaan karakteristik datanya dalam suatu kelompok lebih besar daripada kesamaan karakteristik data tersebut dengan data dalam kelompok lain [2].

Fuzzy C-Means merupakan suatu teknik *clustering* data yang mana keberadaan tiap-tiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaannya. Teknik ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981 [3]. Kelebihan dari *Fuzzy C-Means* adalah dapat melakukan *clustering* lebih dari satu variabel secara sekaligus. Penelitian dengan metode *Fuzzy C-Means* telah dilakukan Hadi untuk pembagian kelas kuliah mahasiswa. Hasilnya dibandingkan dengan hasil menggunakan algoritma *subtraktif*. Kesimpulan yang diperoleh bahwa algoritma *Fuzzy C-Means* lebih cocok untuk alokasi mahasiswa menjadi beberapa kelas karena masukan berupa jumlah *cluster* yang diinginkan [4]. Metode *Fuzzy C-Means* digunakan untuk menentukan nilai akhir kuliah [5]. Dengan menggunakan metode ini, nilai akhir ditentukan secara alami karena berdasarkan pada kecenderungan masing-masing data pada kelompoknya. *Fuzzy C-Means* digunakan untuk *clustering* data *performance* mengajar dosen [6]. Dari 4 pusat *cluster* yang di set di awal iterasi, ada 2 *cluster* yang mempunyai nilai sama, yaitu *cluster* 3 dan *cluster* 4. Dari hasil *clustering*, dihasilkan 3 *cluster performance* mengajar dosen. Aplikasi algoritma *Fuzzy C-Means* diterapkan untuk pengelompokan lulusan Universitas Islam Negeri (UIN) Malang semester ganjil tahun akademik 2005/2006 [7]. *Fuzzy C-Means* digunakan untuk meneliti tentang bahan pangan berdasarkan kandungan nutrisinya [8].

Dari hasil penelitian tersebut, diperoleh 4 *cluster* yaitu golongan hijau, golongan kuning, golongan jingga, dan golongan merah

Penelitian ini menggunakan metode *clustering* dengan *Fuzzy C-Means*. Permasalahan yang dibahas dalam penelitian ini adalah bagaimana cara kerja algoritma *Fuzzy C-Means* untuk menyelesaikan masalah *clustering* lulusan mahasiswa Matematika FMIPA UNTAN Pontianak. Tujuan dari penelitian adalah melakukan pengelompokan lulusan mahasiswa Matematika FMIPA UNTAN Pontianak yang dibatasi dari tanggal 24 Februari 2007 sampai 31 Maret 2012 dan mencari titik pusat *cluster* dengan *Fuzzy C-Means clustering* berdasarkan nilai IPK dan lama studi.

FUZZY C-MEANS CLUSTERING

Fuzzy C-Means clustering adalah suatu teknik *clustering* yang banyak digunakan dalam aplikasi *clustering*. *Fuzzy C-Means* menerapkan pengelompokan *fuzzy*, dimana setiap data dapat menjadi anggota dari beberapa *cluster* dengan derajat keanggotaan yang berbeda-beda pada setiap *cluster*. *Fuzzy C-Means* merupakan algoritma iteratif, yang menerapkan iterasi pada proses *clustering* data. Tujuan dari *Fuzzy C-Means* adalah untuk mendapatkan pusat *cluster* yang nantinya akan digunakan untuk mengetahui data yang masuk ke dalam sebuah *cluster*.

Algoritma *Fuzzy C-Means* adalah sebagai berikut [3]:

1. Menentukan data yang akan dicluster X , berupa matriks berukuran $n \times m$ (n =jumlah sampel data, m =atribut setiap data). X_{ij} =data sampel ke- i ($i=1,2,\dots,n$), atribut ke- j ($j=1,2,\dots,m$).
2. Tentukan jumlah *cluster* (c), pangkat (w), maksimum iterasi (MaxIter), *error* terkecil yang diharapkan (ζ), fungsi obyektif awal ($P_0=0$), iterasi awal ($t=1$).
3. Bangkitkan bilangan *random* μ_{ik} , $i=1,2,\dots,n$; $k=1,2,\dots,c$; sebagai elemen-elemen matriks partisi awal U . Matriks partisi (U) pada pengelompokan *fuzzy* memenuhi kondisi sebagai berikut [9]:

$$\mu_{ik} \in [0,1] \quad 1 \leq i \leq n; \quad 1 \leq k \leq c$$

μ_{ik} adalah derajat keanggotaan yang merujuk pada seberapa besar kemungkinan suatu data bisa menjadi anggota ke dalam suatu *cluster*.

Hitung jumlah setiap kolom (atribut):

$$Q_i = \sum_{k=1}^c \mu_{ik}$$

$$Q_i = \mu_{i1} + \mu_{i2} + \dots + \mu_{ic}$$

dengan $i=1,2,\dots,n$

4. Hitung pusat *cluster* ke- k : V_{kj} , dengan $k=1,2,\dots,c$; dan $j=1,2,\dots,m$

$$V_{kj} = \frac{\sum_{i=1}^n \left((\mu_{ik})^w X_{ij} \right)}{\sum_{i=1}^n (\mu_{ik})^w} \quad (1)$$

5. Hitung fungsi obyektif pada iterasi ke- t , P_t :

Fungsi obyektif digunakan sebagai syarat perulangan untuk mendapatkan pusat *cluster* yang tepat. Sehingga diperoleh kecenderungan data untuk masuk ke *cluster* mana pada step akhir. Untuk iterasi awal nilai $t=1$.

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

6. Hitung perubahan matriks partisi:

$$\mu_{ik} = \frac{\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{-\frac{1}{w-1}}}{\sum_{k=1}^c \left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{-\frac{1}{w-1}}} \quad (3)$$

7. Cek kondisi berhenti:

- $|P_t - P_{t-1}| < \zeta$ atau $(t > \text{MaxIter})$ maka berhenti;
- Jika tidak, iterasi dinaikkan $t=t+1$, ulangi langkah ke-4.

HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini adalah data lulusan mahasiswa Matematika dari tanggal 24 Februari 2007 sampai 31 Maret 2012. Data berupa nilai IPK dan lama studi sebanyak 93 lulusan. Dalam proses *clustering*, IPK dijadikan sebagai X_{i1} dan lama studi dijadikan X_{i2} . Pengolahan data pada penelitian ini menggunakan bantuan program yaitu Matlab. Dari *clustering* yang dilakukan diperoleh hasil yaitu nilai fungsi obyektif selama iterasi, pusat *cluster* atau *center* serta derajat keanggotaan lulusan untuk setiap *cluster* pada iterasi terakhir. Dalam penelitian ini, proses iterasinya berhenti pada iterasi ke-27 karena nilai $|P_t - P_{t-1}| < \zeta$. Nilai fungsi obyektif pada iterasi terakhir yang diperoleh adalah 9,355243. Hasil dari *clustering* yang dilakukan dapat dilihat di bawah ini:

- Nilai fungsi obyektif selama 27 iterasi dapat dilihat pada Tabel 1 di bawah ini:

Tabel 1. Nilai Fungsi Obyektif Selama 27 Iterasi

Iterasi ke-	Fungsi Obyektif
1	28,832597
2	21,933190
3	18,872350
4	14,825657
5	13,204085
6	12,404592
7	12,076182
8	11,634414
9	10,799451
10	9,994893
⋮	⋮
27	9,355243

2. Nilai pusat *cluster* atau *center*

Pada iterasi terakhir (iterasi ke-27), pusat *cluster* V_{kj} yang dihasilkan dengan $k=1,2,3,4$ dan $j=1,2$ adalah:

$$V_{kj} = \begin{pmatrix} 3,15 & 5,09 \\ 2,88 & 7,32 \\ 3,48 & 4,37 \\ 2,89 & 5,91 \end{pmatrix}$$

Nilai ini merupakan nilai dari koordinat keempat titik pusat *cluster* dan memberikan garis besar tiap *cluster* yaitu:

- a. Untuk pusat *cluster* 1 terdiri dari lulusan dengan kisaran lama studi 5,09 tahun dan kisaran IPK 3,15.
 - b. Untuk pusat *cluster* 2 terdiri dari lulusan dengan kisaran lama studi 7,32 tahun dan kisaran IPK 2,88.
 - c. Untuk pusat *cluster* 3 terdiri dari lulusan dengan kisaran lama studi 4,37 tahun dan kisaran IPK 3,48.
 - d. Untuk pusat *cluster* 4 terdiri dari lulusan dengan kisaran lama studi 5,91 tahun dan kisaran IPK 2,89.
3. Derajat keanggotaan lulusan untuk sertiap *cluster* pada iterasi terakhir (iterasi ke-27) dapat dilihat pada Tabel 2 di bawah ini:

Tabel 2. Derajat Keanggotaan Lulusan Untuk Setiap Cluster Pada Iterasi Terakhir

Lulusan ke	Derajat Keanggotaan (μ) Lulusan Untuk Setiap Cluster Pada Iterasi Terakhir			
	μ_{i1}	μ_{i2}	μ_{i3}	μ_{i4}
1	0,1203	0,0161	0,8209	0,0427
2	0,4042	0,0145	0,5209	0,0604
3	0,2821	0,0205	0,6276	0,0698
4	0,3930	0,0175	0,5211	0,0684
5	0,1137	0,0148	0,8317	0,0398
6	0,0709	0,0073	0,9005	0,0213
7	0,8566	0,0096	0,0742	0,0596
8	0,2463	0,0134	0,6905	0,0498
9	0,6898	0,0198	0,0622	0,2282
10	0,4090	0,0383	0,0808	0,4719
⋮	⋮	⋮	⋮	⋮
93	0,0810	0,7202	0,0538	0,1450

Dari derajat keanggotaan lulusan pada iterasi terakhir dapat diperoleh informasi mengenai kecenderungan lulusan untuk masuk ke *cluster* mana. Derajat keanggotaan terbesar menunjukkan bahwa kecenderungan tertinggi lulusan untuk masuk menjadi anggota *cluster* tersebut. Misalnya untuk lulusan ke-3, dapat menjadi:

- (i). Anggota *cluster* pertama dengan derajat keanggotaan 0,2821
 - (ii). Anggota *cluster* kedua dengan derajat keanggotaan 0,0205
-

- (iii). Anggota *cluster* ketiga dengan derajat keanggotaan 0,6276
- (iv). Anggota *cluster* keempat dengan derajat keanggotaan 0,0698

Derajat keanggotaan terbesarnya terletak di *cluster* ketiga, maka lulusan ke-3 akan dimasukkan kedalam *cluster* ketiga. Hasil selengkapnya pengelompokkan ke-93 lulusan kedalam 4 *cluster* dapat dilihat pada Tabel 3 di bawah ini:

Tabel 3. Anggota Pada Keempat Cluster

<i>Cluster</i>	Beranggotakan Lulusan Nomor
1	7, 9, 11, 12, 18, 19, 21, 23, 24, 25, 28, 30, 37, 45, 48, 57, 58, 59, 60, 62, 66, 68, 69, 72, 73, 84.
2	47, 49, 54, 64, 75, 79, 85, 92, 93.
3	1, 2, 3, 4, 5, 6, 8, 29, 32, 44, 46, 55, 56, 61, 63, 71, 74, 78, 82, 83, 86, 87, 89, 90, 91.
4	10, 13, 14, 15, 16, 17, 20, 22, 26, 27, 31, 33, 34, 35, 36, 38, 39, 40, 41, 42, 43, 50, 51, 52, 53, 65, 67, 70, 76, 77, 80, 81, 88.

Dari Tabel 3 di atas dapat diperoleh:

- a. *Cluster* 1 beranggotakan lulusan dengan IPK 2,67 sampai 3,65 dan lama studi 4,75 sampai 5,50 tahun sebanyak 26 lulusan.
- b. *Cluster* 2 beranggotakan lulusan dengan IPK 2,60 sampai 3,28 dan lama studi 6,75 sampai 8,42 tahun sebanyak 9 lulusan.
- c. *Cluster* 3 beranggotakan lulusan dengan IPK 3,07 sampai 3,81 dan lama studi 4,00 sampai 4,83 tahun sebanyak 25 lulusan.
- d. *Cluster* 4 beranggotakan lulusan dengan IPK 2,38 sampai 3,51 dan lama studi 5,42 sampai 6,42 tahun sebanyak 33 lulusan.

PENUTUP

Berdasarkan *clustering* yang dilakukan diperoleh 4 *cluster*. Dari hasil keempat *cluster* tersebut diketahui bahwa pada *cluster* 4 memiliki anggota lulusan yang paling banyak dengan 33 lulusan. *Cluster* 4 terdiri dari lulusan dengan kisaran lama studi 5,91 tahun. Hal ini menunjukkan bahwa masih banyak mahasiswa jurusan Matematika di Fakultas MIPA Untan Pontianak yang menempuh lama studi lebih dari 10 semester atau 5 tahun. Hasil ini dapat dijadikan sebagai bahan pertimbangan jurusan dalam meningkatkan IPK mahasiswa untuk menyelesaikan masa studinya dengan cepat.

DAFTAR PUSTAKA

- [1]. Asfi M. Pelabelan Otomatis Citra Menggunakan *Fuzzy C-Means* Untuk Sistem Temu Kembali Citra [M.Sc thesis]. Bogor: Institut Pertanian Bogor; 2008.
- [2]. Bahar. Penentuan Jurusan Sekolah Menengah Atas Dengan Algoritma *Fuzzy C-Means* [MT thesis]. Semarang: Universitas Dian Nuswantoro; 2011.
- [3]. Kusumadewi S, Purnomo H. *Aplikasi Logika Fuzzy Untuk Pendukung Keputusan*. Edisi ke-2. Yogyakarta: Graha Ilmu; 2010.
- [4]. Hadi HY. Pembagian Kelas Kuliah Mahasiswa Menggunakan Algoritma Pengklasteran *Fuzzy*. Semarang: Universitas Diponegoro; 2005.

- [5]. Khoiruddin AA. 2007, Menentukan Nilai Akhir Kuliah Dengan *Fuzzy C-Means*. *Seminar Nasional Sistem dan Informatika (SNSI)*. 2007 Nov 16.
- [6]. Luthfi ET. *Fuzzy C-Means Untuk Clustering Data (Studi Kasus: Data Performance Mengajar Dosen)*. *Seminar Nasional Teknologi (SNT)*. 2007 Nov 24.
- [7]. Aziz A. *Aplikasi Algoritma Fuzzy C-Means Clustering Untuk Pengelompokan Lulusan*. Malang: Universitas Islam Negeri; 2010.
- [8]. Kusumadewi S. *Klasifikasi Kandungan Nutrisi Bahan Pangan Menggunakan Fuzzy C-Means*. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*. 2007 Jun 16.
- [9]. Kusumadewi S, Hartati S. *Integrasi Sistem Fuzzy dan Jaringan Syaraf*. Edisi ke-2. Yogyakarta: Graha Ilmu; 2010.

CARY LINEKER SIMBOLON : Fakultas MIPA Universitas Tanjungpura, Pontianak,
lineker_cs@yahoo.com
NILAMSARI KUSUMASTUTI : Fakultas MIPA Universitas Tanjungpura, Pontianak,
uminilam@yahoo.com
BENI IRAWAN : Fakultas MIPA Universitas Tanjungpura, Pontianak,
benicsc@yahoo.com
