

Analisis Perancangan Algoritma Fuzzy C-Means dalam Menentukan Dosen Pembimbing Tugas Akhir

Yunita Sari Siregar¹

Strata 1 Program Studi Teknik Informatika
Universitas Harapan Medan
Indonesia
¹yunie.boreg@yahoo.com

Putri Harliana²

Strata 1 Program Studi Teknik Informatika
Universitas Harapan Medan
Indonesia
²cimoputri@gmail.com

Abstrak - Algoritma *fuzzy c-means* merupakan suatu teknik pengelompokan data dimana keberadaan tiap-tiap titik data dalam suatu cluster ditentukan oleh derajat keanggotaan. *Fuzzy c-means* adalah algoritma pengelompokan yang terawasi, karena pada algoritma *fuzzy c-means* jumlah cluster yang akan dibentuk perlu diketahui terlebih dahulu. Dalam penelitian ini, model *fuzzy c-means* akan digunakan untuk menentukan dosen pembimbing tugas akhir sesuai dengan bidang keahliannya. Hasil dari penelitian yang dilakukan tidak sampai pada tahap pengujian, hanya sebatas perancangan perhitungan algoritma *fuzzy c-means*. Dari hasil analisis perancangan yang dilakukan dapat diketahui beberapa syarat dalam menentukan dosen pembimbing. Manfaat penelitian ini agar pembagian dosen pembimbing merata sesuai dengan bidang dan keahlian masing-masing dosen.

Kata Kunci : *Fuzzy C-Means*, Dosen Pembimbing, *Data Mining*

1. PENDAHULUAN

Algoritma *fuzzy c-means* merupakan suatu teknik pengelompokan data dimana keberadaan tiap-tiap titik data dalam suatu cluster ditentukan oleh derajat keanggotaan. *fuzzy c-means* adalah algoritma pengelompokan yang terawasi, karena pada algoritma *fuzzy c-means* jumlah cluster yang akan dibentuk perlu diketahui terlebih dahulu. Konsep dasar algoritma *fuzzy c-means* adalah menentukan pusat kelompok yang akan menandai lokasi rata-rata untuk tiap-tiap *cluster*. Pada kondisi awal, pusat *cluster* ini masih belum akurat. Tiap-tiap titik data memiliki derajat keanggotaan pada tiap-tiap *cluster*. Dengan cara memperbaiki pusat *cluster* dan derajat keanggotaan tiap-tiap titik data secara berulang-ulang, maka didapat lokasi pusat *cluster* optimal. [6]

Penelitian berjudul Clustering Dengan Algoritma Fuzzy C Means Untuk Rekomendasi Pemilihan Bidang Keahlian Pada Program Studi Teknik Informatika yang dilakukan oleh Muhammad Faisal Mirza dari jurusan Teknik Informatika Universitas Dian Nuswantoro Semarang pada tahun 2009. Penelitian ini membahas tentang penggunaan algoritma FCM untuk mengelompokkan mahasiswa

berdasarkan transkrip nilai mata kuliah prasyarat untuk rekomendasi penjurusan. [6]

Berdasarkan penelitian yang sudah pernah dilakukan oleh peneliti lain dan dari uraian di atas penulis tertarik untuk melakukan penelitian perancangan analisis algoritma *fuzzy c-means* dalam penentuan dosen pembimbing. Untuk studi kasus dilaksanakan di Universitas Harapan Medan karena selama ini masih menggunakan pembagian secara manual dan tidak ada pembagian yang secara merata sesuai dengan bidang dan keahlian dosen tersebut.

Dari latar belakang yang telah dijelaskan sebelumnya maka masalah dari penelitian ini adalah seberapa banyak syarat yang diperlukan untuk menentukan dosen pembimbing menggunakan algoritma fuzzy c-means. Dan tujuan yang ingin dicapai pada penelitian ini adalah untuk menentukan dosen pembimbing sesuai dengan minat dan bidang keahlian dosen dengan metode clustering algoritma fuzzy c-means. Manfaat yang diharapkan dari penelitian ini adalah agar pembagian dosen pembimbing tugas akhir merata sesuai dengan kepangkatan dan bidang keahlian masing-masing dosen.

2. TINJAUAN PUSTAKA

2.1 Data Mining

Data *mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data *mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang besar [4]. Data *mining* didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau seringnya semiotomatis. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi. Data yang dibutuhkan dalam jumlah besar [6].

Karakteristik data *mining* sebagai berikut :

- Data *mining* berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- Data *mining* biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.
- Data *mining* berguna untuk membuat keputusan yang kritis, terutama dalam strategi.

Berdasarkan beberapa pengertian yang telah disebutkan diatas, maka dapat ditarik kesimpulan bahwa data *mining* adalah suatu teknik untuk menggali informasi yang tersembunyi pada gunung data. Kata *mining* sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar [6]

Data *mining* juga dapat diartikan sebagai pengeksktrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Data *mining* berkaitan dengan bidang ilmu-ilmu lain, seperti *database system*, *data warehousing*, statistik, *machine learning*, *information retrieval*, dan komputasi tingkat tinggi. Selain itu, data *mining* didukung oleh ilmu lain seperti *neural network*, pengenalan pola, *spatial data analysis*, *image database*, *signal processing*. Istilah data *mining* kadang disebut juga *knowledge discovery* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar [5].

2.2 Fuzzy C-Means

Clustering merupakan proses pengelompokan data dalam kelas-kelas atau *cluster-cluster* sehingga data dalam suatu cluster memiliki tingkat kesamaan yang tinggi antara data satu dengan yang lainnya tetapi

sangat berbeda dengan data pada *cluster* lain. *Clustering* juga dapat dianggap sebagai bentuk kompresi data, di mana sejumlah besar sampel diubah menjadi sejumlah kecil perwakilan prototipe atau *cluster*. Tergantung pada data dan aplikasi, berbagai jenis ukuran kesamaan dapat digunakan untuk mengidentifikasi kelas, di mana ukuran kesamaan mengontrol bagaimana *cluster* terbentuk. Beberapa contoh nilai-nilai yang dapat digunakan sebagai parameter kesamaan termasuk jarak, konektivitas dan intensitas [5].

Clustering dapat diterapkan ke dalam data yang kuantitatif (numerik), kualitatif (kategorikal), atau kombinasi dari keduanya. Data dapat merupakan hasil pengamatan dari suatu proses. Setiap pengamatan dapat memiliki n variabel pengukuran dan dikelompokkan dalam n dimensi vektor $z_k = [z_{1k}, \dots, z_{nk}]^T$, $z_k \in R^n$. Sebuah himpunan dari N pengamatan dinotasikan $Z = \{ z_k \mid k = 1, 2, \dots, N \}$ dan direpresentasikan sebagai matriks $n \times N$. *Cluster* secara umum merupakan wujud himpunan bagian dari suatu himpunan data dan metode *clustering* dapat diklasifikasikan berdasarkan himpunan bagian yang dihasilkan: apakah *fuzzy* atau *crisp (hard)* [4].

Fuzzy C-Means adalah suatu teknik *clustering* (pengelompokan) data di mana keberadaan titik-titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaan. Penentuan titik *cluster* dilakukan secara berulang-ulang hingga diperoleh data yang akurat berdasarkan derajat keanggotaannya. Perulangan ini didasarkan pada minimalisasi fungsi obyektif yang menggambarkan jarak dari titik data ke pusat *cluster* yang terbobot oleh derajat keanggotaan. Akibat adanya derajat keanggotaan tersebut, maka suatu titik data bisa dimiliki lebih dari satu kelompok. Metode ini merupakan minimasi dari fungsi obyektif [5].

3. METODE PENELITIAN

Metodologi penelitian menggunakan algoritma *fuzzy c means* dalam menentukan dosen pembimbing tugas akhir, dalam hal ini penulis melakukan beberapa tahapan sebagai berikut:

1. Penelitian Awal

Pada tahapan ini dikumpulkan bahan penelitian dari berbagai sumber pustaka, seperti buku, jurnal (baik cetak maupun *online*), prosiding, majalah, artikel dan sumber lain yang relevan dalam ilmu pengetahuan.

2. Pengumpulan Data

Sumber data dari penelitian ini dari Sekolah Tinggi Teknik Harapan program studi Teknik Informatika.

3. Inisialisasi Data
Data yang terkumpul diidentifikasi dan di klasifikasikan sesuai dengan kelompoknya. Selain itu juga menentukan validitas data dan variabel yang akan dipakai. Pada tahap inisialisasi data dilakukan pembuatan *interval* variabel *fuzzy c means*.
4. Proses Penentuan Himpunan Data
Dari beberapa variabel *fuzzy c means* yang sudah ditentukan maka selanjutnya dilakukan penentuan himpunan data dimana jumlah cluster yang diharapkan $1 < c < N$, nilai pembobot $m > 1$, toleransi penghentian $\epsilon > 0$.
5. Inisialisasi Matriks
Hasil dari penentuan himpunan data dilakukanlah inisialisasi matriks partisi secara acak, $U(0) \in M_{fc}$.
6. Proses Penghitungan *Cluster Center (Means)*.
Dari inisialisasi matriks yang dihasilkan, maka akan dilakukan proses penghitungan *cluster center (means)* dimana, $v_i^{(1)} = \frac{\sum_{k=1}^N (\mu_{ik}^{(1-1)})^m z_k}{\sum_{k=1}^N (\mu_{ik}^{(1-1)})^m}$, $1 \leq i \leq c$
7. Proses Penghitungan Jarak
Hasil dari penghitungan cluster center, akan dilakukan penghitungan jarak dimana, $D_{ikA}^2 = (z_k - v_i^{(1)})^T A (z_k - v_i^{(1)})$, $1 \leq i \leq c$, $1 \leq k \leq N$
8. Perbaharui Matriks Partisi
Dari nilai yang sudah dihasilkan maka dilakukanlah pembaharuan matriks partisi untuk $1 \leq k \leq N$, Jika $D_{ikA} > 0$ untuk semua $i = 1, 2, \dots, c$

$$\mu_{ik}^{(1)} = \frac{1}{\sum_j^c 1 \left(\frac{D_{ikA}}{D_{jkA}} \right)^{\frac{2}{m-1}}}$$

Atau dengan kata lain:

$$\mu_{ik}^{(1)} = 0 \text{ jika } D_{ikA} > 0 \text{ dan } \mu_{ik}^{(1)} \in [0,1] \text{ dengan } \sum_{i=1}^c \mu_{ik}^{(1)} = 1$$

Ulangi sampai $\|U^{(l)} - U^{(l-1)}\| < \epsilon$.

9. Kesimpulan
Dari beberapa proses yang dilakukan, maka dapat diambil kesimpulan bahwa akan terus dilakukan proses iterasi sampai hasil yang diperoleh maksimal dan sesuai dengan yang penulis harapkan.

Dalam uji kasus pada sistem *clustering* diperlukan data yang akan menjadi sumber acuan dan pembuatan model tersebut, adapun prosedur dari

pemodelan data dapat dijelaskan dari beberapa langkah sebagai berikut:

1. Data masukan diambil dari data Universitas Harapan Medan rogram studi Teknik Informatika.
2. Data merupakan data asli yang kemudian dilanjutkan dengan proses himpunan data. Proses himpunan data bertugas mengubah data asli dari masukan menjadi data *fuzzy* berdasarkan himpunan *fuzzy* yang telah ditetapkan.
3. Setelah menjadi data *fuzzy* kemudian dilanjutkan ke penghitungan matriks partisi, dimana terlebih dahulu dengan ditentukan jumlah *cluster* yang diharapkan dengan nilai bobot dan nilai toleransi yang selanjutnya akan dilakukan penghitungan secara berulang sampai diperoleh hasil yang diharapkan.
4. Keluaran yang dihasilkan dari proses penghitungan *fuzzy c means* yang telah dilakukan adalah hasil pembagian pembimbing.

4. HASIL DAN PEMBAHASAN

Analisis Data

Analisis data merupakan tahapan untuk melakukan penganalisaan terhadap data-data yang digunakan untuk perancangan sistem yang akan dibuat, dalam hal ini penulis mengambil data dari tempat studi kasus yang dilakukan yaitu di Universitas Harapan Medan pada program studi Teknik Informatika.

Data yang diperlukan adalah data sekunder dari Universitas Harapan Medan program studi Teknik Informatika. Data tersebut merupakan data dosen tetap program studi Teknik Informatika dari tahun 2015-2017 dimana data tersebut terdapat nama dosen beserta dengan NIDN, pendidikan dan jabatan fungsional. Jumlah data hanya 41, karena dosen tetap program studi Teknik Informatika berjumlah 41 orang.

Analisis Perancangan Algoritma *Fuzzy C-Means*

Sebelum masuk ke dalam proses algoritma *fuzzy c-means*, data yang diperoleh harus terlebih dahulu dinormalisasi agar tidak terjadi ketimpangan jumlah nilai antara variabel satu dengan yang lain dikarenakan satuan pengukuran berbeda.

Untuk melakukan normalisasi data, dibutuhkan beberapa syarat untuk menentukan dosen pembimbing. Adapun syaratnya antara lain:

1. Kepangkatan yang dimiliki oleh dosen, kepangkatan terdiri dari dosen biasa, asisten ahli, lector, dan lector kepala.
2. Bidang Keahlian.

Bidang keahlian ditinjau dari rekam jejak penelitian yang telah dilakukan oleh dosen. Tidak hanya itu, bidang keahlian juga ditinjau dari mata kuliah yang diampu.

Setelah syarat terpenuhi maka akan didapatkan jumlah variable yaitu 3 variabel (X,Y,Z). Dikarenakan algoritma fuzzy c-means hanya dapat membaca nilai (angka), maka syarat-syarat tersebut harus terlebih dahulu diubah ke dalam bentuk nilai (angka).

Untuk kepangkatan, nilai yang dipenuhi dapat berupa nilai kum angka kredit yang sudah diperoleh oleh masing-masing dosen. Misal, untuk asisten ahli nilai kum sebesar 150, untuk lektor terbagi dua, yaitu lektor 200 dan lektor 300, begitu seterusnya.

Pada bidang keahlian, dapat diambil nilai dari jumlah penelitian yang sudah dibuat oleh masing-masing dosen. Begitu pula untuk mata kuliah yang diampu, seberapa banyak mata kuliah yang diampu oleh dosen tersebut.

Untuk menentukan pembagian dosen pembimbing tugas akhir maka data akan dibagi menjadi tiga bagian sesuai dengan bidang keahlian yang ada di program studi Teknik Informatika Universitas Harapan Medan, yaitu bidang Robotika, Multimedia dan Jaringan.

Apabila ke semua nilai sudah didapat, maka dapat disusun menjadi data nilai inputan algoritma fuzzy c-means. Nilai inputannya dapat dilihat pada tabel 3.1. Tabel 3.1 data inputan dari bidang Multimedia.

Tabel 3.1 Nilai Input Algoritma Fuzzy C-Means

X	Y	Z
150	0	0
150	0	0
150	1	1
150	3	0
150	4	0
150	1	1
150	1	0
150	1	1
150	0	0
300	1	0
300	1	1
300	3	0
200	0	0
100	0	0
300	5	0

150	1	1
200	8	0
150	3	1
150	2	1
150	2	0
150	6	1
150	10	5
150	4	1
150	10	4
150	0	0
150	0	0
150	4	0
150	0	0
150	2	0
150	2	1
200	2	0
200	5	0
200	5	3
300	4	0
300	2	0
300	10	0
300	4	0
300	2	0
100	1	0
100	0	0
150	0	1

Keterangan :

Untuk X adalah nilai kepangkatan, Y adalah jumlah penelitian/jurnal yang telah dilakukan oleh dosen dan Z adalah jumlah mata kuliah yang diampu oleh dosen tersebut. Tabel di atas adalah salah satu dari tabel peminatan yaitu multimedia.

Dari hasil nilai inputan algoritma fuzzy c-means dapat dilakukan normalisasi data. Selanjutnya akan ditentukan pusat cluster dari data tersebut menggunakan perhitungan algoritma fuzzy c-means.

Pusat cluster terdiri dari 3 (tiga) bagian yaitu C1, C2, dan C3 dari masing-masing bidang Robotika, Multimedia dan Jaringan. Setelah menentukan pusat cluster maka tahap selanjutnya adalah menghitung derajat keanggotaan (U1, U2, dan U3).

Dari semua proses yang telah dilakukan, maka akan didapatkan hasil penentuan dosen pembimbing berdasarkan bidang keahliannya.

Penelitian ini hanya membahas masalah perancangan belum sampai pada tahap pengujian.

5. Kesimpulan

Kesimpulan yang dapat diambil dari penelitian ini adalah sebagai berikut :

1. Syarat penentuan dosen pembimbing berdasarkan kepangkatan, jumlah penelitian dan mata kuliah pada bidang keahliannya.
2. Penelitian ini hanya sebatas perancangan perhitungan menggunakan algoritma fuzzy c-means belum sampai pada tahap pengujian.
3. Algoritma *fuzzy c-means* adalah salah satu algoritma dalam metode clustering yang sangat baik dalam melakukan clusterisasi terhadap data dalam kondisi high level, untuk itu harus memperbanyak jumlah variabel data.

6. Daftar Referensi

- [1] Hafisah., Rustamaji, H.J., & Inayanti, Y. 2008. Sistem Pendukung Keputusan Pemilihan Jurusan Di Smu Dengan Logika Fuzzy. *Seminar Nasional Informatika*.pp. 213-218.
- [2] Haryati, S., Sudarsono, A., & Suryana, E. 2015. Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu). *Jurnal Media Infotama***11**(2): 30-138.
- [3] Jayanti, S. & Hartati, S. 2012. Sistem Pendukung Keputusan Seleksi Anggota Paduan Suara Dewasa Menggunakan Metode Fuzzy Mamdani. *Seminar Nasional Matematika dan Pendidikan Matematika*.pp. 55-66.
- [4] Kusriani & Luthfi, E.T. 2009. Algoritma Data Mining, Penerbit Andi, Yogyakarta.
- [5] Mantho, C. J. 2015. Analisis Segmentasi Citra USG Hati Menggunakan Metode Fuzzy C Means. *Citec Journal***2**(3): 256-264.
- [6] Mirza, M. F. 2009. Metode Clustering Dengan Algoritma Fuzzy C Means Untuk Rekomendasi Pemilihan Bidang Keahlian Pada Program Studi Teknik Informatika. *Jurnal Nasional***11**: 1-12.
- [7] Susanto, H., & Sudiyatno. 2014. Data Mining Untuk Memprediksi Prestasi Siswa Berdasarkan Sosial Ekonomi, Motivasi, Kedisiplinan, dan Prestasi Masa Lalu. *Jurnal Pendidikan Vokasi***4**(2): 222-231