

**KOMPARASI KINERJA ALGORITMA FUZZY C-MEANS DAN K-MEANS DALAM
PENGELOMPOKAN DATA SISWA BERDASARKAN
PRESTASI NILAI AKADEMIK SISWA
(Studi Kasus : SMP Negeri 2 Pematangsiantar)**

Nelson Butarbutar¹, Agus Perdana Windarto², Dedi Hartama³, Solikhun⁴

Mahasiswa STIKOM Tunas Bangsa Pematangsiantar¹
Dosen STIKOM Tunas Bangsa Pematangsiantar^{2,3,4}
E-mail : nbutarbutar@gmail.com¹, zhantura.gusti@gmail.com²,
solikhun@amiktunasbangsa.ac.id³, dedyhartama@amiktunasbangsa.ac.id⁴

Abstrak

Berbagai upaya telah dilakukan oleh pihak sekolah untuk meningkatkan prestasi akademik siswa sebagai upaya untuk mencapai standar pendidikan nasional. Salah satunya adalah dengan melakukan bimbingan belajar pada masing-masing siswa, namun hasilnya belum begitu memuaskan. Hal ini disebabkan karena pihak sekolah khususnya bagian pendidikan tidak memahami sepenuhnya kemampuan masing-masing siswa dalam menguasai suatu mata pelajaran khususnya mata pelajaran inti UN. Untuk mengatasi hal ini dengan memanfaatkan teknik clustering akan dilakukan pengelompokan data siswa berdasarkan prestasi nilai akademik yang sumber datanya diperoleh langsung dari bagian pendidikan. Dengan menggunakan teknik clustering, bagian pendidikan akan lebih mudah mendata siswanya berdasarkan kemampuannya masing-masing sesuai dengan prestasi nilai akademiknya. Teknik clustering yang digunakan dalam penelitian ini adalah algoritma Fuzzy C-Means dan K-Means. Berdasarkan hasil penelitian yang dilakukan dapat disimpulkan bahwa teknik clustering yang paling tepat digunakan dalam melakukan pengelompokan data siswa adalah algoritma K-Means dengan jumlah iterasi sebanyak 11 untuk mendapatkan cluster data siswa sedangkan algoritma Fuzzy C-Means membutuhkan proses iterasi yang panjang sebanyak 35 iterasi dan proses perhitungan yang rumit serta hasil cluster data siswa kurang akurat dibandingkan menggunakan algoritma K-Means.

Kata kunci : *Standar pendidikan nasional; UN, clustering; fuzzy C-Means; K- Means*

Pendahuluan

Kurikulum adalah seperangkat rencana dan pengaturan mengenai tujuan, isi dan bahan pelajaran serta cara yang digunakan sebagai pedoman penyelenggaraan kegiatan pembelajaran untuk mencapai tujuan pendidikan tertentu. Tujuan pendidikan itu adalah mencapai standar pendidikan nasional yang meliputi standar isi, proses, kompetensi lulusan, tenaga kependidikan, sarana dan prasarana, pengelolaan, pembiayaan dan penilaian pendidikan. Dalam proses pendidikan di sekolah pihak akademik harus memahami kemampuan masing-masing siswa dalam menguasai suatu ilmu yang diajarkan

disekolah karena hal ini sangat menentukan prestasi masing-masing siswa. Dalam sebuah institusi pendidikan kualitas akademik masing-masing sekolah diukur berdasarkan prestasi nilai akademik masing-masing siswa baik itu nilai akademik setiap semester maupun nilai Ujian Nasional.

Berbagai penelitian telah dilakukan dengan menggunakan teknik *clustering* dalam mengelompokkan data siswa berdasarkan data nilai. Miftahus S., *at al.*, (2012) menggunakan teknik klasifikasi dalam mengelompokkan data nilai siswa untuk memprediksi tingkat kelulusan siswa dalam menghadapi UN. Suprihatin, (2011)

menggunakan teknik *clustering* algoritma *K-Means* dalam menentukan nilai ujian. Anita B. H., et al., (2013) dalam penelitiannya menggunakan algoritma *Fuzzy C-Means* dan TOPSIS dalam mengelompokkan data siswa berdasarkan nilai prestasi akademik mata pelajaran peminatan, minat terhadap jurusan, nilai IQ serta kuota kelas yang tersedia.

Pada saat ini masing-masing sekolah khususnya pada tingkat Sekolah Menengah Pertama berusaha mempersiapkan masing-masing siswanya dalam menghadapi Ujian Nasional yang akan dilaksanakan pada saat akhir studi. Berbagai upaya telah dilakukan seperti mengadakan bimbingan belajar, namun hasil yang diharapkan belum begitu memuaskan. Hal ini tentu disebabkan karena pihak sekolah khususnya bagian pendidikan tidak memahami kemampuan masing-masing siswa dalam menguasai suatu mata pelajaran khususnya mata pelajaran inti UN. Untuk itu melalui penelitian ini diusulkan sebuah teknik *clustering* untuk mengelompokkan data siswa berdasarkan kemampuannya dalam menguasai materi mata pelajaran berdasarkan nilai akademiknya. Dengan menggunakan teknik *clustering* bagian pendidikan akan lebih mudah mendata siswanya sesuai dengan kemampuannya masing-masing, sehingga pada saat kenaikan kelas bagian pendidikan dapat menempatkan siswa dalam satu kelas sesuai dengan kemampuan masing-masing siswa agar pihak guru ataupun wali kelas dapat menerapkan metode pengajaran yang tepat sehingga setiap siswa dapat meningkatkan prestasi akademiknya. Dalam penelitian ini teknik *clustering* yang digunakan adalah algoritma *Fuzzy C-Means* dan algoritma *K-Means*. Setelah dilakukan pengelompokan menggunakan kedua algoritma tersebut, selanjutnya akan dilakukan analisis komparasi kinerja kedua algoritma tersebut dalam melakukan pengelompokan data siswa.

Penelitian ini dilakukan untuk memudahkan pihak sekolah khususnya bagian pendidikan dalam mendata siswa

sesuai dengan kemampuannya masing-masing melalui prestasi nilai akademik siswa dengan mengusulkan sebuah teknik *clustering* dalam mengelompokkan data siswa.

1. Mengimplementasikan teknik *clustering* metode *Fuzzy C-Means* dan *K-Means* dalam melakukan *clustering* data siswa.

Kajian Teori

Data Mining

Manusia dalam suatu organisasi, sadar atau tidak sadar telah memproduksi berbagai data yang jumlahnya sangat besar. Pada dasarnya data adalah entitas yang tidak memiliki arti, meskipun kemungkinan memiliki nilai di dalamnya. *Data mining* merupakan disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data, sehingga *data mining* sering juga disebut *Knowledge Discovery in Database* (KDD).

Konsep utama dari sebuah transformasi dalam *data mining* adalah suatu kumpulan data yang bersumber dari *database* yang berukuran besar yang diekstrak dan dirangkum untuk menemukan suatu pola berupa informasi yang bermanfaat dan mengandung sebuah pengetahuan yang dapat digunakan untuk membantu pengambil keputusan dalam mengambil kebijakan yang tepat untuk kepentingan sebuah organisasi.

Data mining dapat juga didefinisikan sebagai gabungan dari suatu pengalaman, nilai, informasi kontekstual dan juga pandangan pakar yang memberikan suatu *framework* untuk mengevaluasi dan menciptakan pengalaman baru dan informasi (Thomas H. Davenport, Laurence Prusak).

Clustering

Clustering adalah pengelompokan data, hasil observasi dan kasus ke dalam *class* yang mirip. Suatu *cluster* adalah koleksi data yang mirip antara satu dengan yang lain, dan memiliki perbedaan bila

dibandingkan dengan data dari *cluster* lain. Perbedaan utama algoritma *clustering* dengan klasifikasi adalah *clustering* tidak memiliki target/class/label, jadi *unsupervised learning*. *Clustering* sering digunakan sebagai tahap awal dalam proses *data mining*, dengan hasil *cluster* yang terbentuk akan menjadi input dari algoritma berikutnya yang digunakan. Ada beberapa pendekatan yang digunakan dalam mengembangkan metode *clustering*. Dua pendekatan utama tersebut adalah *clustering* dengan pendekatan partisi dan *clustering* dengan pendekatan hirarki. Disamping kedua pendekatan tersebut, ada juga *clustering* dengan pendekatan *automatic mapping* (*Self-Organising Map/SOM*).

Algoritma K-Means

Algoritma *K-Means* merupakan salah satu algoritma *clustering* yang digunakan untuk mempartisi data kedalam beberapa *cluster*, dimana data yang memiliki tingkat kemiripan yang tinggi dikelompokkan dalam satu *cluster* sedangkan data yang memiliki karakteristik yang berbeda dikelompokkan ke dalam *cluster* yang berbeda.

Metode Penelitian

Dalam metode *clustering* konsep utama yang ditekankan adalah pencarian pusat *cluster* secara iteratif, dimana pusat *cluster* ditentukan berdasarkan jarak minimum setiap data pada pusat *cluster*.

Langkah- langkah penyelesaian algoritma *K-Means* dapat diuraikan sebagai berikut :

1. Identifikasi data yang akan di *cluster* dan juga tentukan jumlah *cluster* data, X_{ij} ($i = 1, 2, \dots, n; j = 1, 2, \dots, m$) dimana n adalah jumlah data yang akan di *cluster* dan m adalah jumlah variabel data.
2. Pada awal iterasi pusat setiap *cluster* ditentukan secara acak (sembarang), C_{kj} ($k = 1, \dots, k; j = 1, \dots, m$).
3. Tentukan jarak setiap data terhadap pusat *cluster* dengan menggunakan formula *Euclidean* berikut :

$$d_{ik} = \sqrt{\sum_{j=1}^m (X_{ij} - C_{kj})^2} \text{ ----- (2.1)}$$

4. Kelompokkan data berdasarkan jarak minimum data terhadap pusat *cluster*.
5. Cek kondisi apakah ada data yang masih pindah pada kelompok lain, jika ia lakukan iterasi berikutnya dengan menghitung pusat *cluster* berdasarkan nilai rata-rata dari data yang menjadi anggota pada *cluster* yang terbentuk hasil iterasi sebelumnya dengan menggunakan rumus :

$$C_{kj} = \frac{\sum_{i=1}^p X_{ij}}{p}; \text{ ----- (2.2)}$$

Dimana $X_{ij} \in \text{cluster ke-}k$ dan adalah jumlah anggota *cluster*

6. Jika tidak ada lagi data yang pindah pada *cluster* lain, maka proses iterasi berhenti.

Dalam algoritma *K-Means* suatu data di kelompokkan pada suatu *cluster* jika data tersebut memiliki jarak minimum terhadap pusat *cluster* yang dapat dihitung dengan menggunakan rumus (2.1). Untuk memudahkan pemahaman kita tentang implemetasi penggunaan algoritma *K-Means*, contoh hasil perhitungan algoritma *K-Means* dapat kita lihat pada tabel 1 berikut :

Tabel 1. Contoh Hasil Perhitungan Menggunakan Algoritma *K-Means*

Derajat Keanggotaan ³			Jarak Minimum	Cluster ³		
d ₁	d ₂	d ₃		C ¹	C ²	C ³
2.3764	3.6245	1.2547	1.2547	0	0	1
5.2635	1.2356	2.3746	1.2356	0	1	0
3.5476	2.3669	3.4648	2.3669	0	1	0
4.0887	2.0187	3.2863	2.0187	0	1	0
1.2384	3.4658	4.2974	1.2384	1	0	0
4.2929	2.3648	1.2394	1.2394	0	0	1
2.9383	4.2722	3.0883	2.9383	1	0	0
1.0927	2.3937	1.2533	1.0927	1	0	0

Algoritma Fuzzy C-Means (FCM)

Fuzzy C-Means adalah salah satu metode *fuzzy clustering* yang pertama sekali dikembangkan oleh Dunn (1973) dan kemudian diperbaiki oleh Bezdek (1981) sebagai metode yang sering digunakan dalam pengenalan pola (*pattern recognition*). Dalam FCM pengelompokan data dilakukan berdasarkan derajat keanggotaan yang bernilai antara 0 dan 1. Dalam algoritma FCM langkah pertama yang dilakukan adalah menentukan pusat *cluster* yang akan menandai lokasi rata-rata untuk tiap *cluster*. Pada kondisi awal, pusat *cluster* ini masih belum akurat. 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 objektif yang menggambarkan jarak dari titik data yang diberikan ke pusat *cluster* yang terbobot oleh derajat keanggotaan titik data tersebut. Output dari FCM bukan merupakan *fuzzy inference system*, namun merupakan deretan pusat *cluster* dan beberapa derajat keanggotaan untuk tiap-tiap titik data.

Langkah-langkah penyelesaian algoritma *Fuzzy C-Means* dapat diuraikan sebagai berikut :

1. Identifikasi data yang akan di *cluster* berupa matriks berukuran n x m (X_{ij} = data sampel ke-i ($i = 1,2,...,n$) dan atribut ke-j ($j = 1,2,...,m$)).
2. Menentukan jumlah *cluster* (c), pangkat (w), maksimum iterasi (maxiter), *error* terkecil yang diharapkan (ϵ), fungsi objektif awal ($P_0 = 0$) dan iterasi awal ($t = 1$).
3. Membangkitkan bilangan *random* μ_{ik} ($i = 1,2,...,n; k = 1,2,...,c$) sebagai elemen-elemen matriks partisi awal U. Menghitung jumlah setiap kolom dengan rumus :

$$Q_j = \sum_{k=1}^c \sum_{i=1}^n \mu_{ik} \quad \text{-----} \quad (2.3)$$

Menghitung nilai elemen matriks partisi anggota himpunan U dengan rumus :

$$\mu_{ik} = \frac{Q_{ik}}{Q_i} \quad \text{-----} \quad (2.4)$$

4. Menghitung nilai pusat *cluster* ke-k (V_{kj} dengan $k = 1,2,...,c; j = 1,2,...,m$)

5. Men ada interasi
ke-t (2.5)

$$P_t = \sum_{i=1}^n \sum_{k=1}^c ([\sum_{j=1}^m (X_{ij} - V_{kj})^2] (\mu_{ik})^w) \quad \text{-----} \quad (2.6)$$

6. Menghitung perubahan matriks partisi U:

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}}{\sum_{k=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}} \quad \text{-----} \quad (2.7)$$

7. Cek kondisi berhenti:
 - a. Jika $(P_t - P_{t-1}) < \epsilon$ atau ($t < \text{MaxIter}$) maka iterasi berhenti.
 - b. Jika tidak : $t = t + 1$, ulangi langkah 4 sampai 7.

Dalam algoritma *Fuzzy C-Means* suatu data dikelompokkan dalam suatu *cluster*, jika data memiliki nilai jarak maksimum partisi matriks U (μ_{ik}) terhadap pusat *cluster*. Untuk memudahkan pemahaman implementasi algoritma *Fuzzy C-Means*, berikut contoh tabel 2 hasil perhitungan algoritma *Fuzzy C-Means* :

Tabel 2. Contoh hasil perhitungan menggunakan algoritma *Fuzzy C-Means*

Derajat Keanggotaan			Jarak	Klaster		
μ_{i1}	μ_{i2}	μ_{i3}	Maksimum	C	C	C
0.456	0.284	0.260	0.456	1	0	0
0.282	0.123	0.595	0.595	0	0	1
0.564	0.234	0.202	0.564	1	0	0
0.234	0.554	0.212	0.554	0	1	0
0.345	0.432	0.223	0.432	0	1	0
0.371	0.123	0.506	0.506	0	0	1
0.653	0.231	0.116	0.653	1	0	0
0.354	0.213	0.433	0.433	0	0	1

Standar Penilaian Pendidikan

Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional. Pasal 1 angka 1 menyatakan bahwa “pendidikan adalah usaha sadar dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dirinya, masyarakat, bangsa dan negara”. Standar Nasional Pendidikan berfungsi sebagai dasar perencanaan, pelaksanaan, dan pengawasan pendidikan dalam rangka mewujudkan pendidikan nasional yang bermutu. Salah satu bagian dari Standar Pendidikan Nasional adalah Standar Penilaian yang bertujuan untuk menjamin :

1. Perencanaan penilaian peserta didik sesuai dengan kompetensi yang akan dicapai dan berdasarkan prinsip-prinsip penilaian.
2. Pelaksanaan penilaian peserta didik secara profesional, terbuka, edukatif, efektif, efisien dan sesuai dengan konteks sosial budaya.
3. Pelaporan hasil penilaian peserta didik secara objektif, akuntabel dan informatif.

Standar Penilaian Pendidikan adalah kriteria mengenai mekanisme, prosedur, dan instrumen penilaian hasil belajar peserta didik.

Hasil dan Pembahasan

Penelitian ini dilakukan untuk mengetahui tingkat kemampuan masing-masing siswa dalam memahami dan menguasai ilmu yang di pelajari, dengan melakukan *clustering* terhadap data siswa yang sudah mengikuti ujian akhir semester genap berdasarkan nilai mata pelajaran inti UN (Bahasa Inggris, Bahasa Indonesia, Matematika, dan IPA) dan nilai rata-rata rapor. Hasil *clustering* data siswa nantinya akan dijadikan sebagai dasar untuk pembagian kelas oleh bagian pendidikan pada saat kenaikan kelas. Melalui hasil penelitian ini, diharapkan pihak pendidikan

dapat menentukan kebijakan yang tepat dalam menerapkan metode pengajaran yang akan dilakukan oleh masing-masing guru sesuai dengan kemampuan masing-masing siswa. Selain itu juga, melalui hasil penelitian ini bagian pendidikan dapat memantau perkembangan kemampuan masing-masing siswa dalam menguasai mata pelajaran khususnya mata pelajaran inti UN. Sehingga nantinya setiap siswa lebih siap dalam menghadapi Ujian Nasional, yang akan diadakan pada akhir masa studi serta tingkat kelulusan siswa dapat lebih di maksimalkan. Dalam penelitian ini *clustering* data siswa dilakukan dengan menggunakan algoritma *Fuzzy C-Means* dan algoritma *K-Means*. Selain untuk menentukan *clustering* data siswa, penelitian ini juga dilakukan untuk melakukan analisis komparasi kinerja metode *Fuzzy C-Means* dan metode *K-Means* dalam melakukan pengelompokan data siswa.

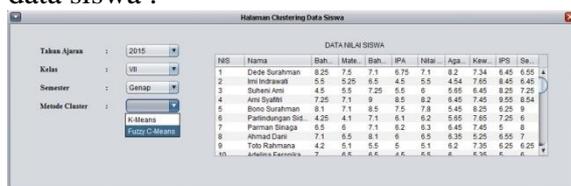
Data penelitian diperoleh langsung dari bagian pendidikan dengan jumlah data sebanyak 150 *record* data siswa. Data yang diperoleh, disimpan pada *database Mysql* agar dapat diproses menggunakan aplikasi *prototypeclustering*. Berikut adalah data nilai siswa yang berhasil diperoleh :

Tabel 3. Data nilai siswa

Nomor	Nama Siswa	Data Nilai Akademik				
		Bahasa Inggris (%)	Matematika (%)	Bahasa Indonesia (%)	IPA (%)	Rata-Rata Nilai Rapor (%)
1	Dede Surahman	8.25	7.50	7.10	6.75	7.10
2	Irni Indrawati	5.50	5.25	6.50	4.50	5.50
3	Suheni Arni	4.50	5.50	7.25	5.50	6.00
4	Arni Syahfitri	7.25	7.10	9.00	8.50	8.20
5	Bono Surahman	8.10	7.10	8.50	7.50	7.80
6	Parlindungan Sidabutar	4.25	4.10	7.10	6.10	6.20
7	Paman Sinaga	6.50	6.00	7.10	6.20	6.30
8	Ahmad Dani	7.10	6.50	8.10	6.00	6.50
9	Toto Rahmana	4.20	5.10	5.50	5.00	5.10
10	Abelina Feronika Situonus	7.00	6.50	6.50	4.50	5.50
11	Putu Donang	4.50	5.20	6.00	5.00	5.50
12	Ali Syahbana	9.00	8.10	8.00	7.50	7.80
13	Romi Dana	8.20	7.50	7.80	6.40	7.60
14	Sulistiwowati	4.30	5.30	6.50	6.50	6.30
15	Butek Manungung	6.20	5.70	7.30	6.30	6.50
16	Rano Karno Sinaga	5.50	6.20	7.30	6.60	7.00
17	Susilowardono	7.20	6.70	7.30	6.20	6.80
18	Roro Mendut	4.20	5.50	7.20	6.50	7.20
19	Endang Surat	5.30	6.40	7.60	6.90	7.00
20	Ihonsen Sidabutar	4.25	5.80	6.20	6.20	6.50
21	Parlin Santun	7.80	6.40	8.30	7.50	7.70
22	Toto Andara Seragih	8.20	9.10	7.40	6.50	7.90
23	Pangerang Hutabarat	5.20	4.60	5.20	6.30	5.30
24	Surti Nandani	4.20	4.70	5.40	6.30	6.00
25	Ortong Andi	6.30	6.30	7.30	6.20	6.60
149	Andi Sinaga	6.20	7.10	6.70	6.50	6.50
150	Permata Strait	7.20	6.50	7.10	6.85	6.90

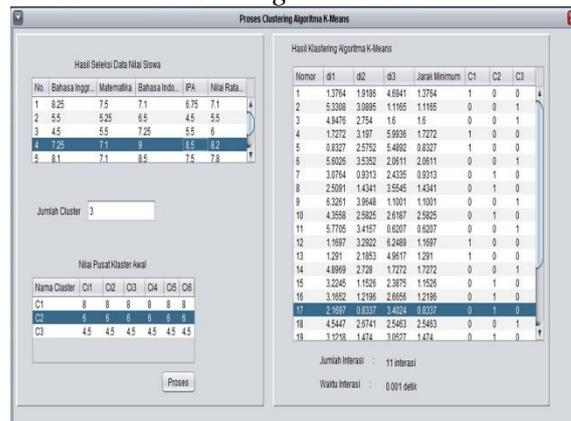
Tahap Pengolahan Data

Untuk mendapatkan *output* dari hasil penelitian yang dilakukan, pengolahan data yang dilakukan adalah menggunakan dua metode teknik clustering. Metode tersebut adalah algoritma *Fuzzy C-Means* dan *K-Means*. Untuk mendapatkan hasil perhitungan *clustering* digunakan aplikasi *prototype clustering* yang dibangun dengan menggunakan bahasa pemrograman java. Berikut adalah tampilan aplikasi clustering data siswa :



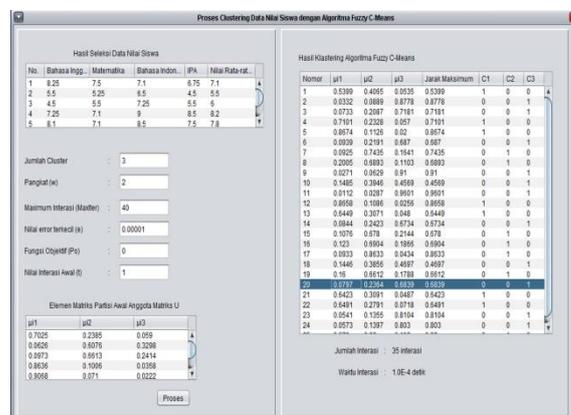
Gambar 1. Halaman cluster data siswa
Halaman Cluster Data Siswa merupakan halaman yang digunakan untuk

memilih metode yang digunakan untuk melakukan *clustering* data siswa.



Gambar 2. Halaman proses algoritma *K-Means*

Halaman Proses Algoritma *K-Means* merupakan halaman yang digunakan untuk menampilkan hasil proses perhitungan algoritma *K-Means*. Dari hasil perhitungan dapat kita lihat bahwa proses iterasi berhenti pada iterasi ke-11 dengan waktu iterasi 0.001 detik.



Gambar 3. Halaman proses algoritma *Fuzzy C-Means*

Halaman Proses Algoritma *Fuzzy C-Means* merupakan halaman yang digunakan untuk menampilkan hasil perhitungan algoritma *Fuzzy C-Means*. Dari hasil perhitungan dapat kita lihat bahwa proses iterasi berhenti pada iterasi ke-35 dengan waktu iterasi 1.0E-4 detik.

Algoritma K-Means

Proses pengelompokan data menggunakan algoritma *K-Means* dilakukan pertama sekali dengan

melakukan proses identifikasi terhadap data yang akan dicluster. Dalam hal ini data yang akan dicluster adalah data siswa yang ada pada tabel 3 di atas. Setelah data yang akan dicluster diketahui maka langkah selanjutnya adalah menentukan pusat cluster awal yang ditentukan secara acak. Langkah selanjutnya adalah melakukan perhitungan algoritma *K-Means* dengan menggunakan aplikasi *prototypeclustering*. Hasil *clustering* ditentukan berdasarkan jarak minimum setiap data pada pusat *cluster* atau proses iterasi *clustering* berhenti jika tidak ada lagi data yang berpindah pada *cluster* lain. Berdasarkan hasil percobaan yang dilakukan, proses iterasi algoritma *K-Means* berhenti pada iterasi ke-11. Hasil *clustering* menggunakan algoritma *K-Means* dapat kita lihat pada tabel 4 berikut :

Tabel 4. Hasil *clustering* data siswa menggunakan algoritma *K-Means*

Nomor	Derajat Keanggotaan			Jarak Minimum	Cluster		
	d ₁	d ₂	d ₃		C ₁	C ₂	C ₃
1	1.3764	1.9186	4.6941	1.3764	1	0	0
2	5.3308	3.0895	1.1165	1.1165	0	0	1
3	4.9476	2.7540	1.6000	1.6000	0	0	1
4	1.7272	3.1970	5.9936	1.7272	1	0	0
5	0.8327	2.5752	5.4892	0.8327	1	0	0
6	5.6026	3.5352	2.0611	2.0611	0	0	1
7	3.0764	0.9313	2.4335	0.9313	0	1	0
8	2.5091	1.4341	3.5545	1.4341	0	1	0
9	6.3261	3.9648	1.1001	1.1001	0	0	1
10	4.3558	2.5825	2.6187	2.5825	0	1	0
11	5.7705	3.4157	0.6207	0.6207	0	0	1
12	1.1697	3.2922	6.2489	1.1697	1	0	0
13	1.2910	2.1853	4.9617	1.2910	1	0	0
14	4.8969	2.7280	1.7272	1.7272	0	0	1
15	3.2245	1.1526	2.3875	1.1526	0	1	0
16	3.1652	1.2196	2.6656	1.2196	0	1	0
17	2.1697	0.8337	3.4024	0.8337	0	1	0
18	4.5447	2.6741	2.5463	2.5463	0	0	1
19	3.1218	1.4740	3.0527	1.4740	0	1	0
20	4.8443	2.6397	1.6567	1.6567	0	0	1
.....
.....
149	2.7579	0.7272	2.8906	0.7272	0	1	0
150	1.9263	0.6998	3.5197	0.6998	0	1	0

Algoritma Fuzzy C-Means

Proses pengelompokan data menggunakan algoritma *Fuzzy C-Means* dilakukan dengan terlebih dahulu melakukan proses identifikasi terhadap data yang akan dicluster. Dalam hal ini data yang akan dicluster adalah data siswa seperti yang terdapat pada tabel 3. Setelah data diidentifikasi, langkah selanjutnya adalah melakukan inisialisasi jumlah *cluster*, tingkat keaburan data (W), maximum iterasi, akurasi (e), nilai fungsi objektif awal, dan iterasi awal.

Dalam proses *clustering* menggunakan algoritma *Fuzzy C-Means* langkah pertama yang dilakukan adalah menentukan nilai berikut :

1. Jumlah Cluster (c) = 3
2. Pangkat (W) = 2
3. Maksimum Iterasi = 40
4. Error Terkecil (e) = 10⁻⁵

5. Fungsi Objektif Awal(P_0) = 0

6. Interas Awal = 1

Dalam algoritma *Fuzzy C-Means* proses interasi berhenti jika nilai hasil perhitungan fungsi objektif bernilai positif dan $(P_t - P_{t-1}) < e$ (*error* terkecil). Setelah dilakukan proses interasi menggunakan *Fuzzy C-Means*, proses interasi berhenti pada interasi ke-35 dengan hasil *clustering* data siswa ditentukan berdasarkan jarak maksimum data pada derajat keanggota *cluster* data. Hasil *clustering* data siswa menggunakan *Fuzzy C-Means* dapat kita lihat pada tabel 5 berikut :

Tabel 5. Hasil *clustering* data siswa menggunakan algoritma *Fuzzy C-Means*

Nomor	Derajat Keanggotaan			Jarak Maksimum	Cluster		
	μ_{i1}	μ_{i2}	μ_{i3}		C_1	C_2	C_3
1	0.5399	0.4065	0.0535	0.5399	1	0	0
2	0.0332	0.0889	0.8778	0.8778	0	0	1
3	0.0733	0.2087	0.7181	0.7181	0	0	1
4	0.7101	0.2328	0.0570	0.7101	1	0	0
5	0.8674	0.1126	0.0200	0.8674	1	0	0
6	0.0939	0.2191	0.6870	0.6870	0	0	1
7	0.0925	0.7435	0.1641	0.7435	0	1	0
8	0.2005	0.6893	0.1103	0.6893	0	1	0
9	0.0271	0.0629	0.9100	0.9100	0	0	1
10	0.1485	0.3946	0.4569	0.4569	0	0	1
11	0.0112	0.0287	0.9601	0.9601	0	0	1
12	0.8658	0.1086	0.0256	0.8658	1	0	0
13	0.6449	0.3071	0.0480	0.6449	1	0	0
14	0.0844	0.2423	0.6734	0.6734	0	0	1
15	0.1076	0.6780	0.2144	0.6780	0	1	0
16	0.1230	0.6904	0.1866	0.6904	0	1	0
17	0.0933	0.8633	0.0434	0.8633	0	1	0
18	0.1446	0.3856	0.4697	0.4697	0	0	1
19	0.1600	0.6612	0.1788	0.6612	0	1	0
20	0.0797	0.2364	0.6839	0.6839	0	0	1
.....
.....
149	0.0833	0.8324	0.0843	0.8324	0	1	0
150	0.0550	0.9257	0.0193	0.9257	0	1	0

Hasil dan Pembahasan

Berdasarkan perhitungan yang dilakukan dengan menggunakan *prototype* aplikasi *clustering* menggunakan algoritma *K-Meanscluster* data siswa yang terbentuk sebanyak 3 *cluster* dengan anggotanya terdiri dari :

- a) *Cluster* pertama (C_1) dengan anggota terdiri dari 1,4,5,12,13,21,22,30,31,36, 37,38,43,44,48,54,55,56,61,62,63,64,65, 71,72,78,80,81,85,86,87,88,90,97,102, 108,109,116,121,122,123,126,130,131, 135,136,137,142,143,147.
- b) *Cluster* kedua (C_2) dengan anggota terdiri dari 7,8,10,15,16,17,19,25,29, 32,33,34,41,45,46,47,52,53,57,58,66,70, 74,75,76,79,89,93,94,95,96,100,101,103, 105,106,107,113,114,115,117,120,125, 129,132,133,134,140,141,144,146,149, 150.
- c) *Cluster* ketiga (C_3) dengan anggota terdiri dari 2,3,6,9,11,14,18,20,23,24, 26,27,28,35,39,40,42,49,50,51,59,60,67, 68,69,73,77,82,83,84,91,92,98,99,104, 110,111,112,118,119,124,127,128,138, 139,145,148.

Berdasarkan perhitungan dengan menggunakan teknik *Fuzzy C-Means* maka data siswa dikelompokkan kedalam 3 *cluster* dengan anggotanya terdiri dari :

- a) *Cluster* pertama (C_1) yang anggotanya terdiri dari 1,4,5,12,13,21,22,30,31,36, 37,38,43,44,48,55,56,61,63,64,65,71,72, 78,81,85,86,87,88,90,97,102,108,109, 116,122,126,130,131,135,136,142,143, 147.
- b) *Cluster* kedua (C_2) dengan anggotanya terdiri dari 7,8,15,16,17,19,25,29,32, 41,45,46,47,52,53,54,57,58,62,66,70,74, 76,79,80,93,94,95,96,100,101,103,105, 106,107,113,114,115,117,120,121,123, 125,129,132,133,134,137,140,141,146, 149,150.
- c) *Cluster* ketiga (C_3) dengan anggotanya terdiri dari 2,3,6,9,10,11,14,18,20,23, 24,26,27,28,33,35,39,40,42,49,50,51,59, 60,67,68,69,73,75,77,82,83,84,89,91,92, 98,99,104,110,111,112,118,119,124,127, 128,138,139,144,145,148.

Analisis Komparasi Kinerja Algoritma Fuzzy C-Means dan K-Means

Untuk menganalisis komparasi kinerja kedua algoritma tersebut dalam melakukan pengelompokan data siswa berdasarkan prestasi akademik, data nilai siswa akan dikelompokkan kedalam 3 cluster dengan ketentuan cluster pertama (C1) rata-rata nilai siswa adalah lebih besar atau sama dengan 7.24, dan cluster kedua dengan rata-rata nilai siswa lebih besar atau sama dengan 6 dan lebih kecil dari 7.24, sedangkan untuk cluster ketiga rata-rata nilai siswa lebih kecil dari 6. Dari hasil interasi kedua algoritma tersebut dapat kita lihat bahwa hasil interasi algoritma *K-Means* lebih akurat dalam melakukan pengelompokan data siswa, seperti terlihat pada data nomor 10 dan nomor 33 dimana nilai rata-rata siswa adalah 6 dan 6.14 pada algoritma *K-Means* data siswa dikelompokkan pada cluster kedua sesuai dengan ketentuan yang diinginkan sedangkan hasil cluster algoritma *Fuzzy C-Means* data siswa dikelompokkan pada cluster ketiga yang tentunya tidak sesuai dengan syarat yang diinginkan sebelumnya. Jadi dapat disimpulkan bahwa algoritma *K-Means* lebih tepat digunakan untuk mengelompokkan data siswa. Perbedaan hasil clustering algoritma *K-Means* dan *Fuzzy C-Means* dapat kita lihat pada tabel 6 berikut :

Tabel 6. Hasil clustering kedua algoritma

Nomor	Nama	Rata-rata Nilai	Cluster Hasil Algoritma K-Means	Cluster Hasil Algoritma Fuzzy C-Means
1	Deke Suralan	7.24	C1	C1
2	Iini Indrawati	5.65	C3	C3
...
9	Toto Rahmana	4.98	C3	C3
10	Adhina Feranda Situmorang	6	C2	C3
...
33	Iwan Subih	6.14	C2	C2
33	Iwan Pradana	6.14	C2	C3
34	Toto Anji Suralan	6.78	C2	C2
...
53	Dora Siring	6.22	C2	C2
54	Permana Sihombing	7.28	C1	C2
55	Rogel Situmorang	7.56	C1	C1
...
61	Widia Pardamean Manik	7.36	C1	C1
62	Beta Siagian	7.34	C1	C2
63	Kuku Siagian	8.52	C1	C1
...
74	Abdi Subih	7.2	C2	C2
75	Abdi Praman	6.1	C2	C3
76	Panza Pradana	7.16	C2	C2
...
79	Toto Sihombing	7.88	C2	C2
80	Yenny Siagian	7.28	C1	C2
81	Beski Manik	8.3	C1	C1
...
88	Ferdinand Sihombing	8.4	C1	C1
89	Marsanda Manik	6.86	C2	C3
90	Endang Manik	7.4	C1	C1
...
120	Dhik Borna	6.66	C2	C2
121	Andika Andar	7.28	C1	C2
122	Pakem Manurung	8.04	C1	C1
123	Pakem Siagian	7.24	C1	C2
...
136	Tito Sihombing	7.66	C1	C1
137	Naschi Sihombing	7.28	C1	C2
138	Kanti Siagian	5.6	C3	C3
...
143	Cakrawati Sals	8.54	C1	C1
144	Susi Subih	6.88	C2	C3
145	Rana Indri	5.16	C3	C3
...
149	Anji Siagian	6.6	C2	C2
150	Permana Siret	6.91	C2	C2

Kesimpulan

Teknik *clustering* dapat digunakan untuk meningkatkan prestasi akademik siswa dengan cara melakukan *clustering* data siswa berdasarkan kemampuannya melalui prestasi nilai akademik, sehingga pihak akademik dapat memantau perkembangan prestasi siswa dan dapat menentukan kebijakan yang tepat dalam menentukan metode pengajaran yang tepat sesuai dengan kemampuan siswa. Berdasarkan hasil penelitian yang dilakukan teknik *clustering* yang paling tepat digunakan adalah metode *K-Means*.

Daftar Pustaka

- Maria E. S., et al, 2015, *Sistem Penentuan Jurusan Sekolah Menengah Atas Negeri 1 Karangmojo*, Seminar Nasional Teknologi Informasi dan Multimedia 2015, STMIK AMIKOM Yogyakarta, ISSN : 2302 – 3805.
- Aniq Noviciatie Ulfah dan Shofwatul Uyun, 2015, *Analisis Kinerja Algoritma Fuzzy C-Means dan K-Means pada Data Kemiskinan*, Jatisi Vol. 1 No. 2, ISSN : 2407 - 4322
- Suprihatin, 2011, *Clustering K-Means Untuk Penentuan Nilai Ujian*, JUSI Vol. I No. 1, ISSN : 2087 – 8737.
- Mardiani, 2014, *Perbandingan Algoritma K-Means dan EM Untuk Clusterisasi Nilai Mahasiswa Berdasarkan Asal Sekolah*, Citec Journal Vol. I No. 4, ISSN : 2354 - 5771
- Erfan Agil Putranto, et al, 2012, *Sistem Pendukung Keputusan Penjurusan Siswa Kelas X SMA Negeri 2 dengan Metode Fuzzy C-Means dengan Penggunaan Daya Dukung Minat*, JURNAL ITSMART Vol. 1 No. 2, ISSN : 2301 – 7201
- Cakra Ramadhana, et al, 2013, *Data Mining dengan Algoritma Fuzzy C-Means Clustering dalam Kasus Penjualan di PT. Sepatu Bata*, Seminar Nasional Teknologi Informasi dan Komunikasi Terapan, ISBN : 979-26-0266-6.
- Tb. Ai Munandar, et al, 2013, *Clustering Data Nilai Mahasiswa Untuk Pengelompokan Konsentrasi Jurusan Menggunakan Fuzzy Cluster Means*, Seminar Nasional Aplikasi Teknologi Informasi (SNATI) Yogyakarta, ISSN : 1907 – 5022.
- Idni Irsalina, et al, 2014, *Clustering Gender Berdasarkan Nilai Maksimum Minimum Amplitudo Suara Berbasis Fuzzy C-Means (FCM)*, Prosiding SNATIF Ke-1, ISBN : 978-602-1180-04-4.
- Idni Irsalina, et al, 2014, *Clustering Gender Berdasarkan Nilai Maksimum Minimum Amplitudo Suara Berbasis Fuzzy C-Means (FCM)*, Prosiding SNATIF Ke-1, ISBN : 978-602-1180-04-4
- Anita Budi Hastuti, et al, 2013, *Implementasi Metode Fuzzy C-Means dan TOPSIS Dalam Membangun Sistem Pendukung Keputusan Penentuan Jurusan SMA (Studi Kasus : Penentuan Jurusan di SMA Negeri 1 Wonosari)*, JURNAL DASI, Vol. 14 No. 2, ISSN : 1411 – 3201.
- Miftahus Sholihin dan Aizatus Sholikhayah, 2012, *Prediksi Tingkat Kelulusan Siswa Dalam UAN di SMP Negeri 2 Deket Dengan Menggunakan Metode Teorema Bayes*, Jurnal Teknik, Vol. 4 No. 2, ISSN : 2085 – 0859.