



Pemodelan Prediksi Predikat Kelulusan Mahasiswa menggunakan Fuzzy C-Means berbasis Particle Swarm Optimizatin

Hardi Jamhur

Sistem Informasi/STIKOM Binaniga

Email: hardi@stikombinaniaga.ac.id

ABSTRACT

Analysis and excavation of information on educational data is an inseparable part of organizing student academic systems in universities. The excavation of information is known as Educational Data mining (EDM), which is a discipline that focuses on the application of techniques and data mining devices specifically in the field of education. One of the EDM processes is related to the need for prediction especially to gain knowledge regarding the progress of student studies. Practically the activity of getting information / knowledge about the progress of student studies is difficult to do conventionally considering the size of the data volume is quite large. The approach to data mining clustering on student study progress data for prediction of predicate graduation tends to be not optimal. Fuzzy C-means (FCM) modeling optimized with Particle Swarm Optimization (PSO) can produce a more optimal prediction performance. In the case of prediction of student graduation prediction, the application of PSO-based FCM algorithm model produces more optimal results, with predictive accuracy of 86% while the FCM algorithm modeling is 79%.

Keywords: Educational Data mining; clustering; Fuzzy c-means; Particle Swarm Optimization; prediction.

A. PENDAHULUAN

1. Latar Belakang

Analisis dan penggalian informasi data akademik memiliki kedudukan strategis di dalam pengelolaan data akademik mahasiswa di perguruan tinggi. Analisis ini penting dilakukan, salah satunya untuk mendapatkan informasi pencapaian hasil studi mahasiswa terutama pada pencapaian tingkat kelulusan/yudisium setiap mahasiswa. Bentuk analisis dari penggalian informasi ini berupa prediksi predikat yudisium yang lazim berlaku di perguruan tinggi yaitu “Memuaskan”, “Sangat Memuaskan”, dan “Dengan Pujian”. Predikat yudisium dijadikan acuan oleh lulusan perguruan tinggi untuk rekomendasi mendapatkan kesempatan kerja dan lainnya. Analisis dan penggalian informasi secara konvensional membutuhkan waktu yang lama dan rumit mengingat jumlah variabel dan volume data yang banyak. Pendekatan komputasi *data mining* dapat dilakukan untuk kebutuhan tersebut.

Analisis data dengan *data mining* dapat dilakukan dengan berbagai teknik pemodelan. Untuk pemodelan pengelompokan data kelulusan dapat dilakukan dengan pendekatan *clustering* dengan mengelompokkan data atas kesamaan tertentu. Pendekatan *clustering* dapat menggunakan beberapa metode seperti Fuzzy c-means, The Gustafson-Kessel, Possibilitic Fuzzy c-means, The Kernel-based Fuzzy c-means, K-means Clustering, dll.

Penggalian informasi dari data pendidikan menggunakan metode *clustering* telah dilakukan oleh banyak peneliti. Sumanto dan Wahono melakukan penelitian tentang peminatan tugas akhir mahasiswa. Dari hasil penelitian yang dilakukan disimpulkan bahwa Fuzzy C-means dapat membantu keakurasian sebesar 80% untuk pemilihan peminatan tugas akhir

mahasiswa. Penelitian Rusdiana tentang penerapan metode *fuzzy c-means* untuk mengelompokkan data predikat kelulusan. Hasil penerapan metode *fuzzy c-means* menunjukkan akurasi 81,6% dalam menentukan predikat kelulusan. Penelitian Akbar tentang penjurusan kelas di SMA berdasarkan nilai evaluasi mata pelajaran menggunakan *fuzzy c-means clustering*. Hasil dari penelitian menunjukkan bahwa ketepatan pemilihan jurusan atau pemilihan konsen-trasi dengan pendekatan *fuzzy c-means* mencapai tingkat akurasi 72%. Kusnadi melakukan analisis prediksi tingkat kelulusan siswa pada pembelajaran berbasis elearning. Hasil penelitian menunjukkan bahwa metode *fuzzy c-means* mampu memperoleh akurasi penentuan kelulusan siswa sebesar 78%.

2. Identifikasi Masalah

Permasalahan utama yang dapat diidentifikasi pada prediksi predikat kelulusan adalah :

- Menemukan informasi tentang pencapaian predikat kelulusan di dalam kumpulan data akademik sulit dan rumit dilakukan secara konvensional.
- Pendekatan komputasi *data mining clustering* untuk mendapatkan prediksi predikat kelulusan belum optimal / tidak akurat.

3. Rumusan Masalah

Berdasarkan pada identifikasi masalah di atas dan kerangka pemecahan dengan pendekatan metode *clustering*, maka dapat dirumuskan

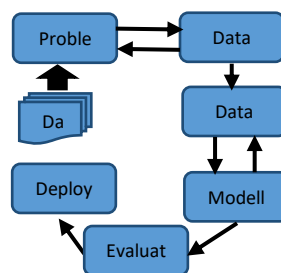
- Bagaimana menyusun kerangka pemodelan prediksi predikat kelulusan mahasiswa dengan pendekatan *Fuzzy c-means (FCM)* berbasis *Particle Swarm Optimization (PSO)*?
- Sejauhmana tingkat akurasi prediksi predikat kelulusan menggunakan al-goritma FCM dengan integrasi ke dalam algoritma PSO ?.

4. Literature Review

a. Konteks Data Mining pada Educational Data Mining (EDM)

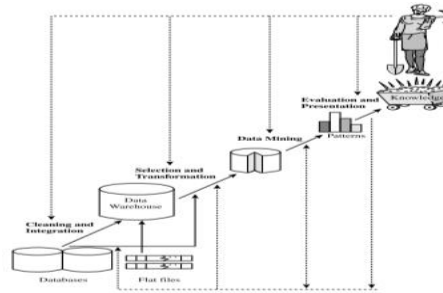
Pengertian *Data mining* merujuk pada pengertian *Knowledge Data Discovery (KDD)*. KDD menggambarkan seluruh proses ekstraksi pengetahuan dari data. Dalam konteks ini, pengetahuan berarti hubungan dan pola antara elemen data. Proses analisis berbagai jenis data dan mengekstraksi informasi yang berguna di dalamnya, itulah *data mining*. Jadi pada dasarnya "*data mining*" digunakan secara eksklusif untuk tahap penemuan proses KDD atau dikatakan bahwa *Data mining* adalah langkah analisis di dalam KDD. Tugas *data mining* bersifat semi otomatis berdasarkan ukuran (kuantitas) data yang tersedia. Pengumpulan data, persiapan, interpretasi dan pelaporan bukan bagian dari *data mining* melainkan adalah bagian dari KDD. Fase-fase dalam *data mining* seperti Gambar 1.

Meningkatnya jumlah data dan kompleksitas analisis dari data perguruan tinggi mendukung untuk proses *data mining*. *Data mining* yang dilakukan terhadap data pendidikan dikenal dengan *educational data mining (EDM)* yaitu suatu disiplin yang berfokus pada penerapan alat dan teknik *data mining* yang terkait dengan pendidikan. *EDM* menitikberatkan pada analitik akademik dengan penggunaan teknik statistik dan penggalian data yang akan membantu dosen dan penasihat akademik menjadi lebih proaktif dalam mengidentifikasi mahasiswa yang berisiko dan meresponsnya dengan semestinya. Sementara Baker dan Yacef mendefinisikan *EDM* sebagai disiplin yang berkaitan dengan pengembangan metode untuk mengeksplorasi jenis data unik yang berasal dari data pendidikan, dan menggunakan metode tersebut untuk lebih memahami mahasiswa dan pengaturan lainnya.



Gambar 1. Fase-fase Data Mining

Proses Data Mining dapat diilustrasikan sebagai Gambar 2.



Gambar 2. Proses Data Mining

b. Segmentasi Clustering dengan Fuzzy C-Means (FCM)

Fuzzy C-means Clustering (FCM) merupakan salah satu metode clustering tiap-tiap titik data dalam suatu cluster ditentukan oleh derajat keanggotaannya. 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 dan 1. Teknik ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981.

Menurut Kusumadewi dan Purnomo, konsep dasar FCM pertama kali adalah menentukan pusat cluster, yang akan menandai lokasi rata-rata untuk tiap cluster. Pada kondisi awal, pusat cluster ini masih belum akurat. Tiap-tiap titik data memiliki derajat keanggotaan untuk tiap-tiap cluster. Dengan cara memperbaiki cluster dan derajat keanggotaan tiap-tiap titik data secara berulang, maka akan dapat dilihat bahwa pusat cluster akan bergerak menuju lokasi yang tepat. Perulangan ini didasarkan pada minimisasi fungsi objektif yang menggambarkan jarak dari titik data yang diberikan ke pusat cluster yang terbobot oleh derajat keanggotaan titik data tersebut.

Output dari FCM adalah sebuah daftar pusat cluster dan beberapa fungsi keanggotaan untuk tiap data. Informasi ini dapat digunakan untuk membangun suatu fuzzy inference system.

Tahap dari algoritma FCM adalah sebagai berikut:

- 1) Input data yang akan di cluster X, berupa matriks berukuran n x m (n=jumlah sampel data, m=atribut setiap data). X_{ij} =data sampel ke-I (i=1,2,...,n), atribut ke-j (j=1,2,...,m).

Tentukan

- Jumlah cluster = c; Pangkat = $w > 1$; Maksimum iterasi = MaxIter;
- Error terkecil yang diharapkan = ϵ (galat terkecil);
- Fungsi objektif awal = $P_0 = 0$; Iterasi awal = $t = 1$;

- 2) Bangkitkan bilangan random μ_{ik} , $i=1,2,\dots,n$; $k=1,2,\dots,c$; sebagai elemen-elemen matriks partisi awal U.

- 3) Hitung jumlah setiap kolom:

$$Q_i = \sum_{k=1}^c \mu_{ik} \dots\dots\dots(1.1)$$

dengan $j=1, 2, \dots, n$, dengan $\mu_{ik} = \mu_{ik}/Q_i \dots\dots\dots(1.2)$

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

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w \cdot X_{ij})}{\sum_{i=1}^n ((\mu_{ik})^w)} \dots\dots\dots(1.3)$$

- 5) Hitung fungsi objektif pada iterasi ke-t, P_t ;

$$P_t = \sum_{i=1}^n \sum_{k=1}^c ((\sum_{j=1}^m (X_{ij} - V_{kj})^2) (\mu_{ik})^w) \dots\dots\dots(1.4)$$

- 6) Hitung perubahan matriks partisi ;

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-1/w-1}}{[\sum_{k=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-1/w-1}]} \dots\dots(1.5)$$

dengan $i=1,2,\dots,n$; dan $k=1,2,\dots,c$.

- 7) Cek kondisi berhenti

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

c. Particle Swarm Optimization (PSO)

PSO adalah algoritma pencarian berbasis populasi yang diinisialisasi dengan populasi solusi acak dan digunakan untuk memecahkan masalah optimasi. PSO adalah metode optimasi *heuristic* global yang diperkenalkan oleh dokter Kennedy dan Everhart pada tahun 1995 berdasarkan penelitian terhadap perilaku kawanan burung dan ikan.

Setiap partikel dalam PSO dikaitkan dengan kecepatan partikel terbang melalui ruang pencarian dengan kecepatan yang dinamis disesuaikan untuk perilaku historis mereka. Oleh karena itu, partikel memiliki kecenderungan untuk terbang menuju daerah pencarian yang lebih baik selama proses pencarian.

Untuk menghitung perpindahan posisi dan kecepatan partikel menggunakan formula:

$$V_i(t) = V_i(t-1) + c_1 r_1 [X_{pbest\ i} - X_i(t)] + c_2 r_2 [X_{Gbest} - X_i(t)] \quad \dots\dots\dots(2.1)$$

$$X_i(t) = X_i(t-1) + V_i(t) \quad \dots\dots\dots(2.2)$$

dimana:

$V_i(t)$ = kecepatan partikel I saat iterasi t ; $X_i(t)$ = posisi partikel I saat iterasi t ;

c_1 dan c_2 = *learning rates* untuk kemampuan individu (*cognitive*) dan pengaruh sosial (*group*); r_1 dan r_2 = bilangan random yang berdistribusi uniformal dalam interval 0 dan 1
 $X_{pbest\ i}$ = posisi terbaik partikel I ; X_{Gbest} = posisi terbaik global.

d. Validasi Hasil Clustering

Terdapat beberapa metode atau indeks evaluasi yang dapat digunakan untuk mengukur kualitas sebuah algoritma *clustering*. Hasil yang didapat dari beragam metode tersebut dapat berbeda karena pendekatan yang dipakai pun berbeda. Ada tiga pendekatan kriteria yang dapat digunakan dalam melakukan evaluasi, yaitu pendekatan dengan kriteria eksternal, pendekatan dengan kriteria internal, dan pendekatan dengan kriteria relatif.

Kriteria eksternal mengukur performansi dengan membandingkan struktur *clustering* dengan informasi yang sudah ada sebelumnya (*priori information*). Validasi ini dilakukan dengan mengukur derajat koresponden antara jumlah *cluster* yang diperoleh dari sebuah algoritma *clustering* dengan kategori label pada informasi priori.

Kriteria internal memperkirakan kecocokan antara struktur dan data dengan menggunakan data itu sendiri. Validasi dengan kriteria internal akan mengukur derajat sebuah partisi yang diperoleh dari algoritma *clustering* berdasarkan *proximity matrix* yang diberikan.

Kriteria relatif menentukan struktur yang lebih baik bila dibandingkan dengan struktur yang lain, seperti lebih stabil atau lebih cocok untuk data tertentu. Kriteria relatif akan mengukur secara kuantitatif kesesuaian data dengan penerapan *single-link* atau *complete-link* pada hierarkis.

Untuk validasi hasil *clustering* menggunakan kriteria eksternal memiliki keunggulan karena melakukan penilaian kualitas *cluster* yang terpisah dan tidak mengandung perkiraan-perkiraan atau prasangka. Namun, kriteria eksternal juga memiliki kerugian karena untuk pengukuran data ekspresi gen, “*gold standard*” eksternal jarang sekali ada.

Menurut Xu jika P adalah partisi yang sudah disiapkan sebelumnya dari data set X dengan N data dan tidak tergantung dari struktur *clustering* C hasil dari sebuah algoritma, maka evaluasi (validasi) C diperoleh dengan membandingkan C terhadap P . Dari sepasang data x_i dan x_j dari set data X , ada 4 kondisi berbeda terbentuk berdasarkan bagaimana x_i dan x_j ditempatkan di C dan P . Kondisi yang mungkin terbentuk adalah:

- Kondisi 1: x_i dan x_j termasuk dalam *cluster* C yang sama dan kategori yang sama dari P , dinotasi sebagai a ;
- Kondisi 2: x_i dan x_j termasuk dalam *cluster* C yang sama tetapi kategori P yang berbeda, dinotasi sebagai b ;
- Kondisi 3: x_i dan x_j termasuk dalam *cluster* C yang berbeda tetapi kategori P yang sama, dinotasi sebagai c
- Kondisi 4: x_i dan x_j termasuk dalam *cluster* C yang berbeda dan kategori P yang berbeda, dinotasi sebagai d .

Berdasarkan pada kondisi di atas jumlah pasangan x_i dan x_j di masing-masing kondisi dapat dipetakan ke dalam *contingency table*.

Tabel 1. Contingency Table

		Kriteria eksternal	
		Katagori sama	Katagori berbeda
clustering	Cluster sama	A	b
	Cluster berbeda	C	d

Jumlah total dari seluruh pasangan dapat dihitung dengan formula $N(N-1)/2$ dan dinotasikan sebagai $M = a+b+c+d$.

Menurut Xu ada beberapa validasi dengan kriteria eksternal yang umumnya digunakan untuk mengukur tingkat kesesuaian antara C dan P, yaitu:

- Rand Index [17]
 $R = (a + d)/M; \dots\dots\dots(3.1)$
- Jaccard coefficient
 $J = a/(a+b+c); \dots\dots\dots (3.2)$
- Fowlkes and Mallows index
 $FM = \sqrt{\frac{a}{a+b} \frac{a}{a+c}}; \dots\dots\dots (3.3)$
- Γ statistics
 $\Gamma = \frac{Ma - m_1 m_2}{\sqrt{m_1 m_2 (M - m_1)(M - m_2)}} \dots\dots\dots(3.4)$

dimana $m_1 = a + b$ dan $m_2 = a + c$.

e. Predikat Kelulusan Mahasiswa di Perguruan Tinggi

Setiap perguruan tinggi menetapkan nilai kelulusan bagi mahasiswanya dan memberi predikat kelulusan berupa label/sebutan kelulusan atau sering disebut sebagai yudisium. Persyaratan kelulusan pada setiap jenjang pendidikan berbeda-beda dan ditetapkan oleh masing-masing perguruan tinggi.

Persyaratan kelulusan untuk yudisium S1, misalnya, memenuhi syarat:

- Lulus matakuliah inti dengan nilai minimal B;
- Memenuhi Indeks Prestasi Kumulatif sesuai ketentuan program studi (misal IPK $\geq 2,0$);
- Membuat penelitian dan atau karya ilmiah berupa proyek atau skripsi;
- Memenuhi beban sks minimal (misal 144 sks).

Sementara itu kualifikasi lulusan juga dinyatakan dalam bentuk pencapaian Indeks Prestasi Kumulatif seperti Tabel 2:

Tabel 2. Predikat Kelulusan

IPK	Predikat kelulusan
2,0 – 2,75	Memuaskan
2,75 – 3,50	Sangat Memuaskan
3,50 – 4,00	Dengan Pujian

Predikat kelulusan ini memiliki pengaruh terhadap kepuasan mahasiswa terutama dikaitkan dengan kebutuhan persyaratan tertentu untuk memasuki studi lanjut, melamar pekerjaan, dan persyaratan lain seperti penerimaan beasiswa dan sebagainya. Kedudukan predikat kelulusan menjadi hal strategis bagi mahasiswa dalam menyelesaikan pendidikannya di perguruan tinggi.

f. Kerangka Konsep Pemecahan Masalah

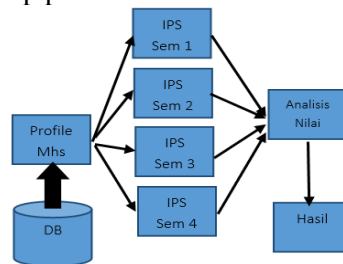
Kerangka pengelompokan (*clustering*) Fuzzy c-means dimulai dengan menangkap data dengan atribut yang mewakili, yaitu profil mahasiswa berupa histori data akademik mahasiswa yang telah terekam. Struktur analisis dapat diilustrasikan sebagaimana Gambar 3.

Sumber data yang dijadikan bahan untuk penelitian ini adalah data akademik dengan struktur analisis sebagai berikut :

- Pemilihan dan ekstraksi data

Pada langkah ini hanya atribut yang sangat relevan dengan proses *clustering* yang dipilih untuk *data mining*. Pra-pengolahan data menghilangkan nilai dan redundansi yang hilang dianggap sebagai *noise* di kumpulan data.

- Fase *FCM Clustering*
Pada tahap ini, algoritma *clustering* di-terapkan pada dataset untuk membagi *record* mahasiswa menjadi sejumlah *cluster* berdasarkan tanda kesamaan dan membandingkan data pengguna baru dengan *cluster* yang lebih mirip
- Fase *FCM clustering* yang diintegrasikan ke algoritma *PSO*
Pada tahap ini algoritma *PSO* diintegrasikan ke dalam iterasi algoritma *FCM* untuk peningkatan nilai *fitness* setiap partikel.



Gambar 3. Struktur analisis nilai

Kerangka algoritma yang dapat dikembangkan untuk pemecahan masalah dapat disusun sebagai berikut:

- Kerangka Algoritma *FCM*:
 - a) Inisialisasi parameter dengan ukuran populasi P , c_1 , c_2 , dan m
 - b) Hitung pusat *cluster* untuk setiap partikel; Hitung jarak Euclidian d_{ij} , $i=1,2,\dots,n$; $j=1,2,\dots,c$; Ubah fungsi keanggotaan m_{ij} , $i=1,2,\dots,n$; $j=1,2,\dots,c$;
 - c) Jika kondisi berakhirnya algoritma *FCM* tidak terpenuhi, ulangi langkah a).
- Kerangka integrasi *PSO* ke dalam *FCM* digambarkan melalui algoritma berikut:
 - a) Menginisialisasi parameter *FCM-PSO* dengan ukuran populasi P , c_1 , c_2 , w , dan m ;
 - b) Membuat swarm dengan partikel (X , P_{best} , G_{best} , dan V berupa matriks $n \times c$); inisialisasi posisi partikel dengan *FCM*;
 - c) Inisialisasi X , V , P_{best} untuk setiap partikel dan G_{best} untuk swarm;
 - d) Hitung pusat cluster untuk setiap partikel; Hitung nilai fitness setiap partikel;
 - e) Hitung P_{best} untuk masing-masing partikel;
 - f) Hitung G_{best} untuk swarm tersebut; Ubah pergerakan matriks setiap partikel;
 - g) Ubah posisi matriks setiap partikel;
 - h) Jika kondisi berakhirnya *FCM-PSO* tidak terpenuhi, ulangi langkah 2. a).

B. METHOD

Metode penelitian yang digunakan dalam penelitian ini adalah metode eksperimen sederhana dengan menggunakan kelompok eksperimen dan kelompok kontrol, dengan tahapan sebagai berikut:

1. Pengumpulan dan pengolahan data awal

Data sekunder pada penelitian ini adalah data mahasiswa STIKOM Binaniaga tahun 2013-2015 yaitu berupa data kemajuan studi semester 1 sampai semester 4. Formasi data kemajuan studi yang diambil adalah 1) npm, 2) indeks prestasi semester 1 (ips1), indeks prestasi semester 2 (ips2), indeks prestasi semester 3 (ips3), dan indeks prestasi semester 4 (ips4).

Dari data sumber dilakukan pengambilan sampel secara acak sejumlah 30 sampel data dan 70 sampel data.

2. Penetapan Parameter

Mengawali proses *clustering* dilakukan penetapan parameter yang diperlukan pada komponen algoritma *FCM* maupun *FCM-PSO*. Pada algoritma *FCM* parameter penting yang perlu ditetapkan adalah jumlah *cluster*, maksimum iterasi dan kriteria penghentian proses.

Pada algoritma *PSO*, pelepasan jumlah partikel P yang besar sejalan dengan prinsip NisB (Nearer is Better/lebih dekat, lebih baik) karena dengan P yang besar tersebar di bidang

pencarian harapannya setidaknya ada salah satu partikel yang gerakannya semakin mendekati titik yang dicari.

Menurut Clerc P yang ideal adalah rentang $20 \leq P \leq 40$ namun dengan mempertimbangkan beban kerja perangkat keras dan data yang diolah berupa matriks maka pilihan nilai P disiapkan pada rentang 2-10 partikel.

Sementara bobot inersia (*inertia weight*) w dikemukakan oleh Shi dan Eberhart dalam adalah untuk meredam kecepatan selama iterasi yang memungkinkan kawanan (*swarm*) dapat lebih akurat dalam menuju titik konvergensi. Peredaman kecepatan ini dilakukan karena implementasinya ditemukan bahwa kecepatan partikel dalam *PSO* yang belum dimodifikasi diperbarui terlalu cepat sehingga nilai optimum menurun yang dicari (minimum atau maksimum) terlewat.

Konsep keseimbangan antara eksplorasi global dan lokal untuk mempercepat konvergensi kepada nilai optimum sesungguhnya diungkapkan dalam suatu rumus bobot inersia yang nilainya menurun linear dengan jumlah partikel dengan jumlah iterasi yang terjadi.

$$w(i) = w_{\max} - \left(\frac{w_{\max} - w_{\min}}{i_{\max}} \right) i ; \dots\dots\dots (4.1)$$

dimana w_{\max} =nilai bobot inersia; w_{\min} = nilai akhir bobot inersia
 i = iterasi sekarang; i_{\max} =iterasi minimal.

Nilai yang sesuai untuk w_{\max} dan w_{\min} adalah 0.9 dan 0.4 [22] dan $i_{\max} = 1000$. Parameter lain yang disiapkan adalah $c1, c2 = 2$, $m=2$. Kriteria penghentian iterasi menggunakan formasi : FCM , $|Jm(t+1)-Jm(t)| < 10^{-5}$, dan $FCM-PSO$, iterasi 1000 kali atau tidak ada perubahan pada $gbest$ selama 100 kali iterasi berturut-turut.

3. Inisialisasi Kelengkapan Swarm

Kawanan (*swarm*) dalam operasinya membutuhkan 4 unsur yaitu x (nilai keanggotaan), v (kecepatan/velocity), $pbest$ (posisi partikel terbaik) dan $gbest$ (posisi kawan terbaik) dalam bentuk matriks. Inisialisasi masing-masing matriks adalah:

a. Matriks x

Nilai awal dari x harus memenuhi kaidah:

$$x_{hik} \in 0, 1, \forall h = 1, 2, \dots, p, \forall i = 1, 2, \dots, n, \forall k = 1, 2, \dots, c;$$

$$\sum_{k=1}^c x_{hik} = 1, \forall h = 1, 2, \dots, p, \forall i = 1, 2, \dots, n$$

$$0 < \sum_{i=1}^n x_{hik} < n, \forall h = 1, 2, \dots, p, \forall k = 1, 2, \dots, c ;$$

b. Matriks v

Nilai awal dari v harus memenuhi kaidah

$$V_{hik} \in -1, 1, \forall h = 1, 2, \dots, p, \forall i = 1, 2, \dots, n, \forall k = 1, 2, \dots, c ;$$

$$\sum_{k=1}^c v_{hik} = 0, \forall h = 1, 2, \dots, p, \forall i = 1, 2, \dots, n$$

c. Matriks $pbest$

Nilai awal $pbest$ didapat dari iterasi pertama dimana

$$Pbest_{hik} = x_{hik}, \text{ untuk } \forall h = 1, 2, \dots, p, \forall i = 1, 2, \dots, n ;$$

d. Matriks $gbest$

Nilai awal dari $gbest$ didapat dari iterasi pertama dimana nilainya adalah dari partikel bernilai Jm terbaik menurut rumus:

$$Gbest_{ik} = x_{tik}, \forall i = 1, 2, \dots, n, \forall k = 1, 2, \dots, c$$

t = partikel terbaik (Jm terendah, karena fungsi minimasi).

4. Eksperimen/Percobaan

Clustering dengan metode *FCM* terhadap data kemajuan studi mahasiswa dilakukan untuk mengetahui atau memetakan pengelompokan (*cluster*) yang mungkin ada dan memiliki pola tertentu dengan parameter sebagai berikut:

- Matriks input berupa nilai indeks prestasi semester, semester 1 s.d semester 4.
- Jumlah *cluster* = 3 (sesuai ketentuan kelulusan/yudisum);
- Pangkat (pembobot) = $w = 2$; Maksimum iterasi = 1000; Kriteria penghentian = $\epsilon = 10^{-5}$.

Pemodelan prediksi menggunakan MatLab dengan formasi data seperti Tabel 3.

Tabel 3. Formasi data input (matriks)

No.id	IPS1	IPS2	IPS3	IPS4
1	2,45	2,77	1,83	0,82
2	3,27	2,73	2,39	3,27
...
n	3,14	3,14	3,00	3,1

Algoritma untuk proses *FCM clustering*:

- Menginisialisasi parameter *FCM* termasuk jumlah cluster (c), pangkat/pembobot (w=2), maksimum iterasi (MaxIter=1000), penghentian ($\epsilon=10^{-5}$), bobot eksponen untuk mengendalikan kekaburan (fuzziness=m), fungsi objektif awal (P0=0), iterasi awal (t=1);
- Baca data yang akan di *cluster* X, berupa matriks n x m (n=jumlah sampel data, m=atribut setiap data). Xij=data sampek ke-I (i=1,2,...,n), atribut ke-j (j=1,2,...,m).
- Bangkitkan bilangan random μ_{ik} , i=1,2,...,n; k=1,2,...,c; sebagai elemen-elemen matriks partisi awal U; Hitung jumlah setiap kolom (persamaan (1.1) dan (1.2));
- Hitung pusat cluster ke-k: V_{kj} , dengan k=1,2,...,c; dan j=1,2,...,m (persamaan (1.3));
- Hitung fungsi objektif pada iterasi ke-t, Pt (persamaan (1.4));
- Hitung perubahan matriks partisi (persamaan (1.5));
- Cek kondisi berhenti
Jika $(|Pt-Pt-1| < \epsilon)$ atau $(t > \text{MaxIter})$ maka berhenti; Jika tidak $t=t+1$, ulangi langkah ke-4.

Clustering dengan metode FCM-PSO diinisialisasi dengan parameter berikut:

- Matriks input berupa nilai indeks prestasi semester, semester 1 s.d semester 4 sebagai partikel swarm; Jumlah cluster = 3 (sesuai ketentuan kelulusan/yudisum);
- Bobot inersia = w (0,9 dan 0,4); koefisien akselerasi c1, c2; bobot eksponen m; Maksimum iterasi = 1000; Kriteria penghentian iterasi = $\epsilon = 10^{-5}$.

Algoritma untuk proses *clustering FCM-PSO*, sebagai berikut:

- Menginisialisasi parameter *PSO* dan *FCM* termasuk ukuran populasi P, c1, c2, w, dan m;
- Buat swarm dengan partikel P (X, pbest, gbest, dan V yaitu matrik nxc)
- Inisialisasi X, V, pbest untuk setiap partikel dan gbest untuk swarm
- Hitung pusat cluster untuk setiap partikel (gunakan persamaan (1.3))
- Hitung nilai fitness setiap partikel

$$f(X) = \frac{K}{I_m}$$

X = posisi dari particle, yang memperlihatkan relasi fuzzy dari set data/objek $o = \{O_1, O_2, \dots, O_n\}$

K = konstanta; $J_m =$ fungsi objektif FCM; $J_m = \sum_{j=1}^c \sum_{i=1}^n \mu_{ij}^m d_{ij}$; $d_{ij} = \|O_i - z_j\|$

- Hitung pbest untuk masing-masing partikel; Hitung gbest untuk swarm tersebut
- Ubah pergerakan matriks setiap partikel (gunakan persamaan ((2.1)); Ubah posisi matriks setiap partikel (gunakan persamaan (2.2)); Jika kondisi berakhirnya *FPSO* tidak terpenuhi, ulangi langkah 4

C. HASIL PENGAMATAN

1. Hasil Pengamatan FCM Clustering

Untuk 30 record data, diperoleh hasil *clustering* dan perbandingan terhadap predikat kelulusan sebenarnya sebagaimana Tabel 4.

Tabel 4. Perbandingan kelulusan FCM clustering dengan Kelulusan sebenarnya

id	Nilai keanggotaan			Cluster FCM			Kelulusan sebenarnya		
				C 1	C 2	C 3	S M	D P	M
1	0,050584	0,02684	0,922576			*			*
2	0,882903	0,077952	0,039145	*					*
3	0,722854	0,24315	0,033996	*			*		
3	0,61423	0,352291	0,033478	*			*		
4	0,769175	0,118273	0,112552	*			*		
5	0,754845	0,216003	0,029151	*			*		
6	0,739921	0,147703	0,112376	*			*		
7	0,263516	0,096341	0,640143			*			*
8	0,901895	0,086407	0,011697	*			*		

9	0,932192	0,045082	0,022726	*				*	
10	0,888287	0,094018	0,017695	*			*	*	
11	0,774982	0,175606	0,049412	*			*	*	
12	0,417656	0,127995	0,454349			*		*	
13	0,099601	0,060728	0,839671			*		*	
14	0,271813	0,113794	0,614394			*		*	
15	0,891872	0,096104	0,012024	*			*		
16	0,842804	0,136017	0,021179	*			*		
17	0,616737	0,356532	0,02673	*			*		
18	0,665333	0,145987	0,18868	*			*	*	
19	0,852703	0,131274	0,016023	*			*		
20	0,120427	0,854994	0,024579		*		*		
21	0,657766	0,317502	0,024731	*			*		
22	0,488585	0,487619	0,023796	*			*		
23	0,693035	0,280726	0,026239	*			*		
24	0,201851	0,732481	0,065668		*		*	*	
25	0,741502	0,234121	0,024377	*			*		
26	0,200745	0,07691	0,722345		*		*	*	
27	0,793554	0,174996	0,03145		*		*		
28	0,670977	0,264909	0,064114	*			*		
29	0,119148	0,853184	0,027668		*		*		
Total				20	4	6	20	1	9
Salah				3	3	0			
Benar				17	1	6			

Penghitungan pusat *cluster* dan perbaikan nilai keanggotaan dihentikan pada iterasi ke 36 dengan fungsi objektif (Jm) 7.828311, dan waktu olah 1,8068 detik, dengan pusat *cluster*:

3.0904 2.7196 2.8175 2.9765
 3.5338 3.2442 3.5859 3.1948
 2.7114 2.5362 1.9743 0.9603

2. Hasil Pengamatan FCM-PSO

Untuk 30 record data, diperoleh hasil *clustering* dan perbandingan terhadap predikat kelulusan sebagaimana Tabel 5.

Tabel 5. Perbandingan kelulusan FCM-PSO clustering dengan Kelulusan sebenarnya

id	Nilai keanggotaan			Cluster dengan FCM-PSO			Kelulusan yang sebenarnya		
				C1	C2	C3	S	D	M
1	0,050584	0,02684	0,922576			*			*
2	0,077952	0,039145	0,882903			*			*
3	0,722854	0,24315	0,033996	*			*		
4	0,61423	0,352291	0,033478	*			*		
5	0,769175	0,118273	0,112552	*			*		
6	0,754845	0,216003	0,029151	*			*		
7	0,739921	0,147703	0,112376	*			*		
8	0,263516	0,096341	0,640143			*			*
9	0,901895	0,086407	0,011697	*			*		
10	0,932192	0,045082	0,022726	*			*		*
11	0,888287	0,094018	0,017695	*			*		*
12	0,774982	0,175606	0,049412	*			*		*
13	0,417656	0,127995	0,454349			*			*
14	0,099601	0,060728	0,839671			*			*
15	0,271813	0,113794	0,614394			*			*
16	0,891872	0,096104	0,012024	*			*		
17	0,842804	0,136017	0,021179	*			*		
18	0,616737	0,356532	0,02673	*			*		
19	0,18868	0,145987	0,665333			*			*
20	0,852703	0,131274	0,016023	*			*		
21	0,120427	0,854994	0,024579		*		*		
22	0,657766	0,317502	0,024731	*			*		
23	0,488585	0,487619	0,023796	*			*		
24	0,693035	0,280726	0,026239	*			*		
25	0,201851	0,732481	0,065668		*		*	*	
26	0,741502	0,234121	0,024377	*			*		*
27	0,200745	0,07691	0,722345		*		*		*
28	0,793554	0,174996	0,03145		*		*		*
29	0,670977	0,264909	0,064114	*			*		*
30	0,119148	0,853184	0,027668		*		*		*
Total				18	4	8	20	1	9
Salah				1	3	0			
Betul				17	1	8			

Penghitungan pusat *cluster* dan perbaikan nilai keanggotaan dihentikan pada iterasi ke 114 dengan fungsi objektif (Jm) 3.7509e+06, dan waktu olah 2,4152 detik, dengan pusat *cluster* :

3.0904 2.7196 2.7175 2.8765
3.5338 3.2442 3.5859 3.4948
2.7114 2.5362 1.9743 0.9603

Demikian pula untuk model FCM dan FCM-PSO dengan 70 data record diperoleh:

FCM: Penghitungan pusat *cluster* dan perbaikan nilai keanggotaan dihentikan pada iterasi ke 33 dengan fungsi objektif (Jm) 14,749636, dan waktu olah 1,9422 detik, dengan pusat cluster:

2.7387 2.5872 2.1970 1.7613
3.5958 3.1695 3.4348 3.3736
3.3209 2.9708 2.9348 2.7536

FCM-PSO: Penghitungan pusat *cluster* dan perbaikan nilai keanggotaan dihentikan pada iterasi ke 154 dengan fungsi objektif (Jm) 6.893e+06, dan waktu olah 2,5522 detik, dengan pusat cluster:

3.5338 3.2442 3.5859 3.1948
3.2904 2.7196 2.7175 2.8765
2.5114 2.5362 1.9743 0.9708

Hasil rata-rata pengamatan dari kedua metode di atas (FCM dan FCM-PSO) diperoleh hasil perbandingan sebagai Tabel 6.

Tabel 6. Hasil rata-rata pengamatan

Aspek	FCM	FCM-PSO
<i>Untuk 30 data sample:</i>		
Jm Akhir	7,828311	3.7509e+06
Jm mulai konvergen (iterasi ke)	36	154
Jumlah iterasi	40	158
Waktu (cpu time)	1,806876 detik	2,415262 detik
<i>Untuk 70 data sample:</i>		
Jm Akhir	7,828311	6.8936e+06
Jm mulai konvergen (iterasi ke)	33	210
Jumlah iterasi	39	216
Waktu (cpu time)	1,9422250 detik	2,552258 detik

3. Pengukuran (validasi) Prediksi

Pengukuran (validasi) prediksi dalam penelitian ini menggunakan pengukuran validasi kriteria eksternal yaitu dengan membandingkan hasil pengukuran metode *clustering* dengan hasil penetapan pengkatagorian yang sudah diperoleh sebelumnya. Pengkatagorian predikat kelulusan mahasiswa sebagai pembanding diperoleh melalui penyaringan berdasarkan indeks prestasi kumulatif yang dapat memetakan ke dalam predikat lulus “dengan pujian”, “memuaskan”, dan “sangat memuaskan”.

Proses evaluasi/validasi mengikuti pemilahan kondisi yang dapat terbentuk dari perbandingan hasil *clustering* terhadap kriteria eksternal yang sudah diperoleh, sebagaimana Tabel 7.

Tabel 3. Tabel Kondisi Pasangan Data x_i dan x_j

kondisi	kondisi	notasi
1	Pasangan x_i dan x_j termasuk dalam <i>cluster</i> C yang sama dan kategori yang sama di P	a
2	Pasangan x_i dan x_j termasuk dalam <i>cluster</i> C yang sama dan kategori yang berbeda di P	b
3	Pasangan x_i dan x_j termasuk dalam <i>cluster</i> C yang berbeda dan kategori yang sama di P	c
4	Pasangan x_i dan x_j termasuk dalam <i>cluster</i> C yang berbeda dan kategori yang berbeda di P	d
Total pasangan x_i dan x_j		a+b+c+d

Sebaran jumlah pasangan yang terbentuk dapat dilihat pada *Contingency table* pada Tabel 8.

Tabel 4. Contingency table 30 sampel untuk algoritma FCM

		Kriteria eksternal	
		Kategori sama	Kategori berbeda
clustering	Cluster sama	155	41
	Cluster berbeda	45	181

Dengan menggunakan validasi Rand Index diperoleh

$$\begin{aligned}
 R &= (a + b)/M \\
 &= (155+181)/(155+41+45+181) \\
 &= 336/422 = 0,7962
 \end{aligned}$$

Nilai Rand Index berada pada range [0 – 1], 0 berarti hasil *clustering* tidak memiliki kesamaan/kemiripan terhadap data partisi eksternal, dan 1 berarti memiliki kesamaan terhadap data partisi eksternal. Dari Rand Index di atas diperoleh nilai $R=0,79$ sehingga dapat diartikan bahwa hasil clustering 79% sesuai dengan partisi eksternal.

Tabel 5. Contingency table 30 sampel untuk algoritma FCM-PSO

		Kriteria eksternal	
		Kategori sama	Kategori berbeda
clustering	Cluster sama	178	8
	Cluster berbeda	51	195

Rand Index

$$\begin{aligned}
 R &= (a + b)/M \\
 &= (178 + 195)/(178+8+51+195) \\
 &= (373)/(432) = 0,8634
 \end{aligned}$$

$R = 0,8634$ menunjukkan bahwa tingkat kesamaan hasil *clustering* terhadap partisi eksternal sebesar 86%.

Secara lengkap hasil pengujian untuk 30 dan 70 sampel data, seperti Tabel 10.

Tabel 6. Hasil pengujian pemodelan FCM & FCM-PSO

Validasi/data	FCM		FCM-PSO	
	30	70	30	70
Rand Index / Tingkat kesamaan hasil <i>clustering</i> C terhadap partisi P (eksternal)	79%	82%	86%	84%

Dari pengukuran uji model *FCM* dan *FCM-PSO* di atas menunjukkan bahwa tingkat akurasi *FCM-PSO* dalam memprediksi kelulusan lebih baik (optimal) dari model *FCM*.

D. KESIMPULAN

Prediksi predikat kelulusan mahasiswa adalah bagian yang strategis didalam pengelolaan data akademik/kemajuan studi mahasiswa. Kepentingan mahasiswa dalam memperoleh kualifikasi kelulusan yang dapat memudahkan dalam menempuh dunia nyata selepas pendidikan menjadi sasaran pokok pengelolaan data akademik mahasiswa tersebut. Pemanfaatan pemodelan komputasi menjadi penting dilakukan mengingat karakteristik data dan jumlahnya yang semakin banyak. Pendekatan komputasi menggunakan algoritma *Fuzzy c-means* dan *Particle Swarm Optimization* menunjukkan kemampuan olah yang dapat digunakan pada pemodelan tersebut.

Kerangka dasar pemodelan diawali dengan penyusunan data olah yang bersumber dari data akademik mahasiswa. Proses seleksi dan ekstraksi data dilakukan untuk mendapatkan formasi data yang diperlukan. Dalam penelitian ini data akademis diekstraksi untuk mendapatkan data indeks prestasi. Pemodelan *fuzzy c-means* menghasilkan peta kemampuan mahasiswa dalam bentuk peringkat kelulusan “memuaskan”, “sangat memuaskan”, dan “dengan pujian”. Peta ini dapat dipandang sebagai representasi kemampuan mahasiswa untuk dijadikan pedoman dalam mempersiapkan proses pendidikan yang dijalani.

Integrasi *particle swarm optimization* ke dalam algoritma *fuzzy c-means* menunjukkan hasil yang signifikan untuk mendapatkan nilai optimasi algoritma fuzzy c-means. Hasil penelitian menunjukkan bahwa basis *particle swarm optimization* mampu mengoptimasi jarak euclidean data ke pusat *cluster*. Pemodelan algoritma *fuzzy c-means* memang menghasilkan waktu olah lebih cepat dari *fcm-particle swarm optimization*. Untuk data yang sama algoritma *FCM* memiliki

waktu olah 1,8068 & 1,9422 detik, sementara algoritma *FCM-PSO* 2,4152 & 2,5522 detik. Namun pemodelan algoritma *FCM-PSO* menghasilkan hasil akurasi prediksi lebih tinggi dari FCM (86%).

E. REFERENCES

- [1] Sumanto and R. S. Wahono, "PENERAPAN FUZZY C-MEANS DALAM PEMILIHAN," vol. 1, no. 1, 2011.
- [2] L. Rusdiana, "APLIKASI BERBASIS FUZZY C-MEANS DALAM PENENTUAN," vol. 2, no. 2, pp. 1–9, 2016.
- [3] F. Akbar, "Clustering Data Nilai Siswa SMA Untuk Penentuan Jurusan Menggunakan Algoritma Fuzzy C-Means," vol. XII, no. 1, 2015.
- [4] Y. Kusnadi, "PREDIKSI TINGKAT KELULUSAN SISWA ELEARNING BERBASIS ALGORITMA FUZZY C-MEANS," vol. 15, no. 1, pp. 1–6, 2018.
- [5] R. Saxena, "Educational Data Mining : Performance Evaluation of Decision Tree and," vol. 14, no. April, pp. 1–10, 2015.
- [6] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, vol. 54, no. Second Edition. 2006.
- [7] R. A. Huebner, "A survey of educational data-mining research," pp. 1–13, 2012.
- [8] J. P. Campbell, P. B. DeBlois, and D. G. Oblinger, "Academic Analytics: A New Tool for a New Era," *Educ. Rev.*, vol. 42, no. October, pp. 40–57, 2007.
- [9] R. S. J. D. Baker and K. Yacef, "The State of Educational Data Mining in 2009: A Review and Future Visions," *J. Educ. Data Min.*, vol. 1, no. 1, pp. 3–16, 2009.
- [10] S. Kusumadewi and H. Purnomo, *Aplikasi Logika Fuzzy untuk Pendukung Keputusan*, 2nd ed. Yogyakarta: Graha Ilmu, 2013.
- [11] A. Naba, *Belajar Cepat Fuzzy Logic Menggunakan MATLAB*. Yogyakarta: CV. ANDI OFFSET, 2009.
- [12] J. Yan, Michael, and J. Power, *Using Fuzzy Logic (Toward Intelligent Systems)*. New York: Prentice-Hall, 1994.
- [13] A. Abraham, C. Grosan, and V. Ramos, *Swarm Intelligence in Data Mining*. Verlag Berlin Heidelberg: Springer, 2006.
- [14] Q. Bai, "Analysis of Particle Swarm Optimization Algorithm," vol. 3, no. 1, pp. 180–184, 1998.
- [15] R. Xu and D. C. Wunsch, *Clustering*. 2008.
- [16] F. Gorunescu, *Data Mining: Concepts, models and techniques*. Springer-Verlag Berlin Heidelberg, 2011.
- [17] W. M. Rand, "Objective Criteria for the Evaluation of Clustering Methods," *J. Am. Stat. Assoc.*, vol. 66, no. 336, pp. 846–850, Dec. 1971.
- [18] S. Binaniaga, *Pedoman Akademik 2017*. Bogor: STIKOM Binaniaga, 2017.
- [19] E. Mehdzadeh, S. Sadi_nezhad, and R. Tavakkoli-moghaddam, "Optimization of Fuzzy Clustering Criteria By a Hybrid PSO and Fuzzy c-means Clustering Algorithm," vol. 5, no. 3, pp. 1–14, 2008.
- [20] M. Clerc, *From Theory to Practice in Particle Swarm Optimization in B.K. Handbook of Swarm Intelligence: Concepts, Principles and Applications*. Berlin: Springer, 2011.
- [21] M. Clerc, *Particle Swarm Optimization*. London: ISTE Ltd, 2006.
- [22] B. Santosa and P. Willy, *Metoda dan Metaheuristik: Konsep dan Implementasi*. Surabaya: Guna Widya, 2011.
- [23] S. R. Rao, *Engineering Optimization: Theory and Practice 4th Edition*. Hoboken - New Jersey: John Wiley & Sons Inc, 2009.
- [24] H. Izakian, A. Abraham, and V. Snasel, "Fuzzy Clustering Using Hybrid Fuzzy C-Means and Fuzzy Particle Swarm Optimization," in *Proceeding of World Congress on Nature and Biologically Inspired Computing (NaBIC 2009)*, 2009.
- [25] W. Pang, K. Wang, C. Zhou, and L. Dong, "Fuzzy Discrete Particle Swarm Optimization For Solving Traveling Salesman Problem," in *Proceeding of the fourth international conference on computer and information technology*, 2004, pp. 796–800.