

Clustering Kinerja Akademis Mahasiswa Pascasarjana UIN Khas Jember Menggunakan *Fuzzy C-Means*

Elok Rahmawati^{1*}, Deni Arifianto², Amalina Maryam Zakiyyah³
Program Studi Teknik Informatika, Universitas Muhammadiyah Jember¹
Program Studi Manajemen Informatika, Universitas Muhammadiyah Jember^{2,3}
Email: elokrahmawati894@gmail.com^{1*}, deniarifianto@unmuhjember.ac.id²,
amalinamaryamzakiyyah@unmuhjember.ac.id³

ABSTRAK

Clustering mahasiswa program pascasarjana dilakukan untuk mengelompokkan mahasiswa ke dalam *cluster* tertentu yang memiliki kemiripan data menggunakan algoritma *Fuzzy C-means*. Data yang diambil yaitu data umur dan IPS (index prestasi mahasiswa). Metode validitas *cluster* dari hasil algoritma *Fuzzy C-means* menggunakan DBI (*Davies bouldin index*) untuk mengukur seberapa bagus *cluster* yang dihasilkan oleh algoritma *Fuzzy C-means*. Dari penelitian yang telah dibuat peneliti berikut hasil DBI *cluster* 2 adalah *cluster* yang menghasilkan nilai DBI terkecil yaitu 1,9570. Berdasarkan hasil dari identifikasi profil mahasiswa pada 2 *cluster* didapatkan bahwa *cluster* 1 memiliki data mahasiswa dengan rentang umur 21 sampai 45, nilai ips (index prestasi semester) 1 rentang 2,00 sampai 4,00 dan rata-rata jenis kelamin adalah perempuan. Sedangkan, *cluster* 2 memiliki data mahasiswa dengan rentang umur 24 sampai 45, nilai IPS (index prestasi semester) 1 rentang 1,70 sampai 3,80 dan rata-rata jenis kelamin adalah laki-laki.

Kata Kunci: DBI (*Davies bouldin index*), *Fuzzy C-means*

ABSTRACT

Clustering graduate students is done to group students into certain clusters that have similar data using the Fuzzy C-means algorithm. The data taken are age data and ips (student achievement index). The cluster validity method of the Fuzzy C-means algorithm results uses DBI (Davies bouldin index) to measure how good the cluster produced by the Fuzzy C-means algorithm is. From the research that has been made by the following researchers, the DBI results of cluster 2 are the clusters that produce the smallest DBI value, namely 1.9570. And based on the results of identifying student profiles in 2 clusters, it is found that cluster 1 has student data with an age range of 21 to 45, the value of IPS (semester achievement index) 1 ranges from 2.00 to 4.00 and the average gender is female. Meanwhile, cluster 2 has student data with an age range of 24 to 45, an IPS value (semester achievement index) 1 range of 1.70 to 3.80 and the average gender is male.

Keywords: DBI (*Davies bouldin index*), *Fuzzy C-means*

1. PENDAHULUAN

Salah satu indikator kemajuan suatu negara adalah meningkatnya kualitas pendidikan yang dimilikinya. Dalam hal ini, bidang pendidikan memiliki peran penting untuk meningkatkan kualitas kualitas diri. Tergantung pada kualitas tenaga kerja akan dipengaruhi oleh beberapa hal diantaranya pendidikan dan kesehatan. Dengan mengenyam pendidikan tinggi, manusia dapat meningkatkan kualitasnya. Salah satu perguruan tinggi dapat memperoleh pelatihan tambahan selain dari pendidikan yang diwajibkan pemerintah, yaitu dengan mengikuti kuliah.

Di Indonesia banyak sekali tempat menuntut ilmu dan jurusan yang berbeda-beda, salah satunya adalah di kota Jember yang merupakan kota yang memiliki tempat belajar yang cukup banyak dan berbagai jurusan yang bisa dipelajari untuk menimba ilmu. .bangun. Untuk mencapai efek pembelajaran yang berkualitas, UIN KHAS Jember telah menerapkan berbagai cara untuk pengembangan yang telah dilakukan seperti dengan merumuskan kurikulum, membangun gedung dengan fasilitas yang sesuai dan melengkapi dan melengkapi fasilitas khusus untuk siswa penyandang cacat, mengevaluasi kinerja fakultas dan mengubah pengajaran metode.

Rasio mahasiswa pascasarjana merupakan salah satu ukuran kualitas kampus UIN KHAS Jember karena rasio mahasiswa pascasarjana dapat mempengaruhi nilai akreditasi kampus. Sebagai contoh, studi kasus pada program Pascasarjana yang masa studi normalnya adalah 2 tahun studi atau 4 semester menunjukkan bahwa masih banyak mahasiswa yang tidak dapat mengikuti perkuliahan

dalam batas studi normal. Di bawah ini adalah persentase total data untuk mahasiswa pascasarjana lembaga agama Islam Jember 2015-2018.

Berdasarkan banyaknya mahasiswa program Pascasarjana UIN KHAS Jember terdapat berbagai macam karakter serta kondisi mahasiswa yang berbeda-beda. Sering kali mengakibatkan mahasiswa tersebut mengulang mata kuliah yang terjadi karena mendapat nilai yang tidak sesuai dengan ketentuan ke lulusan. Peneliti berkeinginan untuk mengelompokkan data mahasiswa melalui metode *Clustering*.

Dalam proses pengelompokan data mahasiswa peneliti menggunakan metode *Clustering* dikarenakan metode tersebut mampu memperoleh output kelompok-kelompok data. Menurut Satriyanto (2011), *Clustering* adalah metode pengelompokan berdasarkan ukuran kedekatan. Metode *Clustering* memiliki bermacam-macam algoritma pada proses pengolahannya. *Fuzzy C-means* merupakan salah satu dari macam-macam algoritma metode *Clustering*. 1 yang disusun oleh Fitria dkk, dengan judul Perbandingan Pengclusteran Data Iris Menggunakan Metode K-Means dan *Fuzzy C-means*. Menyimpulkan bahwa hasil dari pengclusteran data user knowledge modeling dengan menggunakan metode K-Means dan *Fuzzy C-means* ternyata metode *Fuzzy C-means* yang menghasilkan nilai validitasnya yang mendekati 1 yaitu 0,2854 sedangkan menggunakan metode K-Means menghasilkan nilai 0,1866 yang artinya metode *Fuzzy C-means* lebih baik dari pada metode K-Means untuk melakukan *Clustering*.

Oleh karenanya peneliti ingin menggunakan metode *Fuzzy C-means* untuk penelitian peneliti, dikarenakan metode *Fuzzy C-means* menghasilkan output yang lebih baik untuk proses pengklasteran data. Maka peneliti mengajukan judul penelitian "*Clustering* Kinerja Akademis Mahasiswa Pascasarjana UIN KHAS Jember Menggunakan *Fuzzy C-means*".

2. KAJIAN PUSTAKA

A. Data Mining

Data mining adalah proses pengumpulan data yang mengandung sejumlah besar informasi yang akan digunakan untuk tujuan tertentu (Setiawan, 2021).

B. Clustering

Clustering adalah metode dalam pengolahan data *mining* yang menghasilkan *output* dari proses pengolahan data nya adalah sebuah kelompok atau kelas-kelas yang berkarakteristik sama atau mirip.

C. Fuzzy C-means

Output dari *Fuzzy C-means* adalah deretan pusat *cluster*, dan setiap titik data memiliki beberapa derajat keanggotaan (Rahakbauw dkk, 2017). Informasi ini dapat digunakan untuk membangun sistem *inferensi fuzzy*. Berikut Persamaan 1 - 5 perhitungan *Fuzzy C-means*.

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

$$\mu_{ik} = \frac{\mu_{ik}}{Q_i} \quad (2)$$

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * x_{ij})}{\sum_{i=1}^n (\mu_{ik})} \quad (3)$$

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left(\left[\sum_{j=1}^m (X_{ij} - V_{kj}) \right]^2 \right) (\mu_{ik})^w \quad (4)$$

$$\mu_{ik} = \frac{\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2\right]^{-1}}{\sum_{k=1}^c \left[\sum_{j=1}^m (X_{ij} - V_{kj})^2\right]^{-1}} \quad (5)$$

D. Normalisasi Min-Max

Normalisasi Min-Max adalah proses yang digunakan untuk menormalkan semua variabel data menjadi data dengan bobot yang sama (Shofiani, 2017). Dapat dilihat pada Persamaan 6.

$$X_n = \frac{x_0 - x_{min}}{x_{max} - x_{min}} \quad (6)$$

E. Davies Bouldin Index

Tujuan dari DBI adalah menentukan nilai atau *value* dari beberapa *cluster* yang diperoleh untuk menjadikan salah satu nilai *cluster* yang paling valid untuk diambil atau digunakan.

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (7)$$

Untuk menghitung nilai DBI nilai yang harus di cari terlebih dahulu adalah:

1. SSW (*Sum of Square Within-Cluster*)

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i) \quad (8)$$

2. SSB (*Sum of Square Between-Cluster*)

$$SSB_{i,j} = d(c_i, c_j) \quad (9)$$

3. Ratio

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (10)$$

3. METODE PENELITIAN

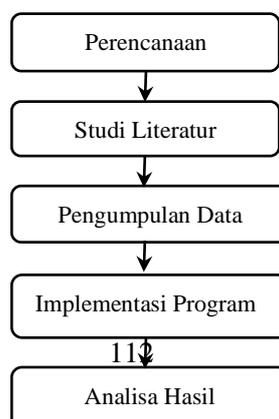
A. Pendekatan Penelitian

Metode penelitian kuantitatif mengacu pada observasi yang mengukur skala nilai sampai batas tertentu. Penelitian kuantitatif mencakup semua jenis penelitian berdasarkan persentase, rata-rata, atau kalkulasi statistik lainnya. Alat yang digunakan misalnya tabel dan diagram.

B. Jenis Penelitian

Riset terapan menekankan pada manfaat praktis dari hasil riset dalam memecahkan masalah spesifik dan menemukan produk baru yang bermanfaat bagi kehidupan.

C. Langkah-langkah Penelitian



Gambar 1. Metode Penelitian

D. Perencanaan

Pada tahap perencanaan terdapat 3 uraian yang akan dilakukan untuk tahap ini yaitu merumuskan masalah, *survey*, observasi.

E. Studi Literatur

Adapun buku dan jurnal yang diambil referensi dalam studi literatur ini adalah Buku Aplikasi Logika Fuzzy (Kusumadewi & Purnomo, 2004).

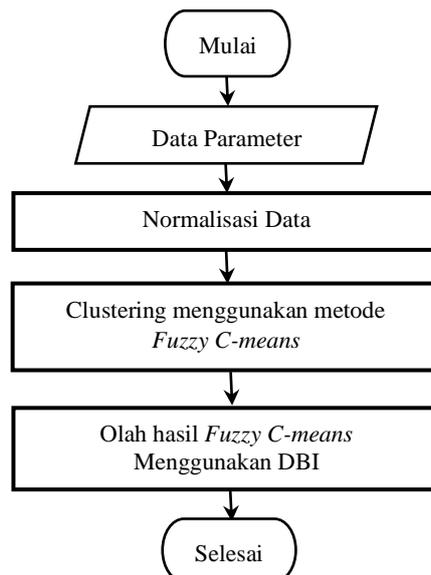
F. Analisis Data

Pada proses penumpukan data terdapat 4 hal yang diperhatikan yaitu:

1. Seleksi data Bertujuan untuk menyeleksi data parameter yang diperlukan dalam proses data *mining*.
2. *Cleaning* Bertujuan untuk menjamin kualitas data yang dipilih, pada titik ini masalah yang akan diselesaikan adalah duplikasi data.
3. Integrasi data bertujuan mengintegrasikan data ke dalam database, karena datanya lebih dari satu tabel, maka perlu dilakukan tahapan penggabungan data dari berbagai tabel yang berbeda.
4. Normalisasi parameter yang bertujuan untuk menyeimbangkan data nilai parameter dan menyederhanakan data nilai parameter dengan skala *value* 0 sampai 1 menggunakan Persamaan 6.

G. Implementasi Program

Pada Gambar 2 diagram alur dibawah ini akan dijelaskan bagaimana langkah perancangan sistem.



Gambar 2. Diagram Alur Sistem

H. *Clustering* data menggunakan metode *Fuzzy C-means*

Menentukan nilai parameter awal berdasarkan data dari tabel 1 ditentukan beberapa komponen perhitungan sebagai berikut:

1. banyaknya *cluster* yang diinginkan = c
2. pangkat (pembobot) = w
3. maksimum iterasi = T
4. error terkecil = 10^{-5}
5. fungsi objektif awal = P_0

6. iterasi awal $= 1$

Untuk menentukan pusat *cluster* menggunakan Persamaan 3. Fungsi objektif pada iterasi pertama P1 dihitung dengan menggunakan Persamaan 4. Menghitung perubahan Matriks Partisi dengan menggunakan Persamaan 5.

I. Validitas *Cluster* dengan Metode *Davies Bouldin Index* (DBI)

Pada Tabel 1 menjelaskan hasil dari nilai anggota *cluster* dan Tabel 2 menunjukkan hasil pusat *cluster*.

Tabel 1. Hasil nilai anggota *cluster*

Data cenderung masuk ke <i>cluster</i> ke-	
1	2
0,9932	0,9189
0,9932	0,8950
	0,7914

Sumber: Hasil DBI

Tabel 2. Hasil Pusat *Cluster*

<i>Cluster</i>	Pusat <i>cluster</i>		
1	0,0745	0,7622	0,9218
2	0,0662	0,6139	0,0080

Sumber: Hasil Pusat *Cluster*

Setelah memperoleh nilai anggota *cluster* dan nilai pusat *cluster* maka dapat dihitung SSW, SSB, dan Ratio. Tabel 3 merupakan hasil nilai DBI.

Tabel 3. Nilai Rmax dan DBI

R	Data <i>cluster</i> ke-i		Rmax	DBI
	1	2		
Data <i>cluster</i> ke-	1	0	2,5558	2,5558
	2	2,5558	2,5558	

Sumber: Hasil Perhitungan DBI

J. Analisa Hasil

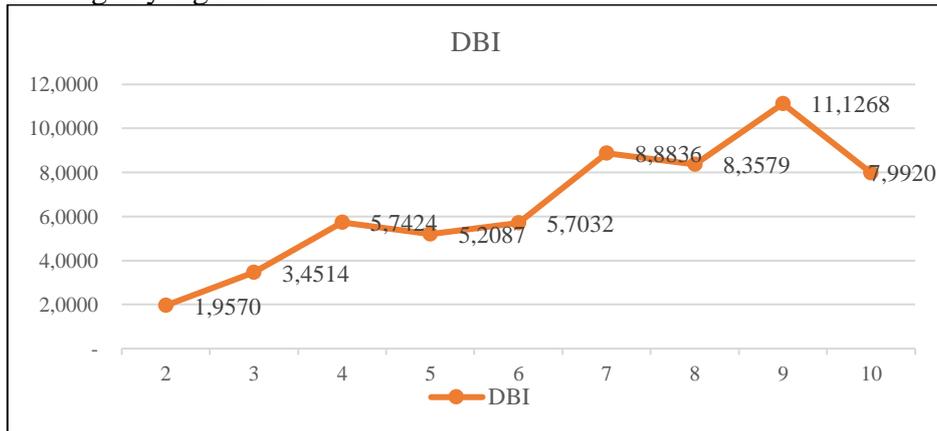
Analisa hasil adalah tahapan terakhir dari penelitian ini. Dengan di ujinya data yang didapatkan kemudian data tersebut diolah dengan algoritma *Fuzzy C-means* dan DBI (*Davies Boildin Index*). Dalam analisa hasil ini berikut lingkungan uji coba dan jadwal proyek penelitian:

1. Lingkungan uji coba Pascasarjana UIN KHAS Jember.
2. Alat dan bahan yang digunakan pada penelitian ini adalah:
 - a. Laptop core i5
 - b. Aplikasi *text editor vs code*
 - c. Xampp
 - d. Data parameter

4. HASIL DAN PEMBAHASAN

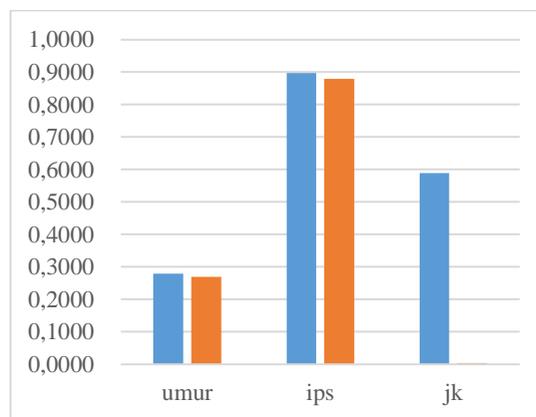
Berdasarkan pada Gambar 3 adalah Grafik nilai DBI menunjukkan bahwa pada di titik 1,9570 adalah nilai DBI terkecil dari 10 titik nilai DBI, dimana titik 1,9570 adalah 2 *cluster*. Maka 2 *cluster*

adalah *cluster* yang paling optimal dari pengelompokan *cluster* yang lain. Evaluasi hasil DBI yang bernilai paling kecil menunjukkan bahwa nilai DBI yang paling baik karena semua anggota *cluster* relatif dekat satu dengan yang lain.



Gambar 3. Grafik Nilai DBI

Dan hasil pengelompokan pada Gambar 4 adalah data profil mahasiswa berdasarkan parameter umur, IPS dan jenis kelamin dengan *cluster* 1 sebanyak 457 mahasiswa dan *cluster* 2 sebanyak 270 mahasiswa.



Gambar 4. grafik rata-rata kedekatan
 Sumber: Hasil Penelitian

5. KESIMPULAN

A. Kesimpulan

Berdasarkan penelitian yang dilakukan oleh peneliti dalam pengelompokan seleksi kinerja akademis mahasiswa program pascasarjana UIN KHAS Jember menggunakan algoritma *Fuzzy C-means*. Diperoleh hasil perhitungan DBI dari hasil masing-masing DBI maka dapat disimpulkan bahwa *cluster* 2 merupakan *cluster* paling baik karena nilai DBI *cluster* 2 menunjukkan nilai yang paling mendekati 0.

Selanjutnya mengidentifikasi profil di masing-masing *cluster*. Pada Gambar 4 adalah hasil identifikasi dari profil 2 *cluster*. Berdasarkan grafik identifikasi 2 *cluster* didapatkan bahwa *cluster* 1 memiliki data mahasiswa dengan umur 21 sampai 45, nilai IPS 2,00 sampai 4,00 dan rata-rata jenis kelamin adalah perempuan. Sedangkan pada *cluster* 2 memiliki data mahasiswa dengan umur 24 sampai 45, nilai IPS 1,70 sampai 3,80 dan rata-rata jenis kelamin adalah laki-laki.

B. Saran

Adapun saran agar aplikasi ini dapat berfungsi dengan lebih optimal yaitu adalah:

1. Metode yang digunakan untuk perhitungan pengelompokkannya dapat menggunakan metode selain *Fuzzy C-means*.
2. Menentukan nilai validitas pada *cluster* dapat menggunakan perhitungan selain perhitungan *davies bouldin index* (DBI).
3. Menyeimbangkan bobot data parameter agar data parameter memiliki bobot yang sama dapat menggunakan perhitungan selain perhitungan normalisasi min-max.

6. DAFTAR PUSTAKA

- Kusumadewi, S., & Purnomo, H. (2004). *Aplikasi Logika Fuzzy*. Yogyakarta: Graha Ilmu.
- Rahakbauw, D.L., Ilwaru, V.Y.I., dan Harury, M.H. (2017). Implementasi Fuzzy C-Means *Clustering* Dalam Penentuan Beasiswa, *Jurnal Ilmu Matematika dan Terapan*, 11(1), 1-11.
- Satriyanto, E. (2011). *Clustering*. Surabaya: Institut Teknologi Sepuluh Nopember.
- Setiawan, R. (2021). Apa itu Data Mining dan Bagaimana Metodenya? [Online] Diakses pada <https://www.dicoding.com/blog/apa-itu-data-mining/#:~:text=Berdasarkan%20uraian%20di%20atas%2C%20data,dan%20untuk%20eksplorasi%20atau%20exploratory>.
- Shofiani, N. (2017). *Segmentasi Supplier Menggunakan Metode Kmeans Clustering (Studi Kasus: PTPN X PG Meritjan)*. Surabaya: Institut Teknologi Sepuluh Nopember.