

## KOMPARASI METODE CLUSTERING K-MEANS DAN FUZZY C-MEANS UNTUK MEMPREDEKSI KETEPATAN WAKTU LULUS

Galet Guntoro Setiaji<sup>1</sup>, Khoirudin<sup>2</sup>, Vensy Vydia<sup>3</sup>

<sup>1,2</sup>Program Studi Teknik Informatika Universitas Semarang

<sup>3</sup>Program Studi Sistem Informasi Universitas Semarang

<sup>1</sup>[galet@usm.ac.id](mailto:galet@usm.ac.id), <sup>2</sup>[khoirudin@usm.ac.id](mailto:khoirudin@usm.ac.id), <sup>3</sup>[vensy@usm.ac.id](mailto:vensy@usm.ac.id)

### ABSTRAK

Metode data mining merupakan metode yang banyak digunakan oleh peneliti untuk mencari sebuah informasi dari sebuah kumpulan data-data, penelitian ini menggunakan object data lulusan mahasiswa Universitas Semarang. Data yang kita ambil yaitu dari jurusan Teknik Elektro dan Teknik Sipil.

Dengan menggunakan metode data mining terutama algoritma K-Means dan Fuzzy C-Means dapat mengidentifikasi ketepatan lulusan di jurusan Teknik Elektro dan Teknik Sipil. Dimana ada tiga cluster yaitu tepat waktu, tidak tepat waktu dan over studi.

Dari hasil penelitian dihasilkan validitas *K-Means* lebih kecil dan lebih baik, dengan validitas DBI sebesar 2.022, sedangkan FCM dengan validitas XBI didapatkan validitas sebesar 11.948. Dengan kesimpulan lulusan lama studi 4 tahun lebih 2 bulan dengan IPK rata-rata 3.13, sedangkan lama studi 4 tahun dengan ipk 3.42, dan untuk mahasiswa dengan lama studi 5 tahun lebih rata-rata ipk 2.88.

Keywords : Clustering, K-Means, Fuzzy C-Means, Prediksi

### 1. Pendahuluan

Sebuah data yang bisa diolah dengan sebuah metode algoritma tertentu dan bisa menghasilkan sebuah informasi yang dapat digunakan untuk menarik sebuah keputusan. Begitu juga dengan adanya perkiraan informasi mahasiswa yang tepat kelulusannya. Kita mengambil data dari mahasiswa Universitas Semarang, yaitu Fakultas Teknik terutama prodi Teknik Sipil dan Teknik Elektro.

Dari data tersebut, yaitu data IPK, lama studi dan Program Studi yang nantinya digunakan untuk mengetahui pengelompokan berdasarkan kelulusan mahasiswa dari dua program studi. Kluster merupakan metode dengan mengelompokkan data-data kedalam kluster sehingga didalam tiap-tiap kluster berisi data yang sama atau memiliki kemiripan data. Pengelompokan kluster ini dengan cara menghitung jarak

tiap data dengan menggunakan perhitungan jarak euclidian distance (jarak terdekat).

Algoritma K-means merupakan salah satu metode yang pengelompokan data yang nantinya bisa menghasilkan kluster yang lebih stabil, dan algoritma K-Means juga merupakan algoritma yang sering digunakan untuk melakukan pengelompokan. Algoritma K-Means juga mudah melakukan modifikasi disetiap tahap seperti fungsi perhitungan jarak.

### 2. Metode

#### a. Data Mining

Data mining adalah suatu proses ekstraksi atau penggalian data yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari database yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting (Connolly dan Begg, 2010). Data mining biasa juga disebut dengan “Data atau knowledge discovery” atau menemukan pola

tersembunyi pada data. Data mining adalah proses dari menganalisa data dari prespektif

yang berbeda dan menyimpulkannya ke dalam informasi yang berguna.

### 3. Pembahasan

Data yang dikumpulkan berupa data lulusan wisudawan dengan tahun angkatan masuk mahasiswa tahun 2010 dan nilai IPK mahasiswa program studi Teknik Sipil dan Teknik Elektro, didapatkan tabel seperti dibawah ini :

NIM	NAMA	IPK	WISUDA	SMTLULUS
C.111.11.0002	ARBI SETYOKO	2.89	52	GENAP TH 2015/2016
C.111.11.0008	ANDY ROSADI	3.10	50	GENAP TH 2014/2015
C.111.11.0015	RIMBA AKASIA AJIDARMA	3.13	50	GENAP TH 2014/2015
C.111.11.0016	FAUZI ARIE HARTONO	2.92	52	GENAP TH 2015/2016
C.111.11.0017	NUR WERO WIRAWAN	3.15	50	GENAP TH 2014/2015
C.111.11.0018	RAKHMAD SUBARCAH	3.06	50	GENAP TH 2014/2015
C.111.11.0019	BUDI IRAWAN	3.08	50	GENAP TH 2014/2015
C.111.11.0020	ANDI AULIA DAVID	3.31	50	GENAP TH 2014/2015
C.111.11.0023	MUHAMAD SAMSUL HADI	2.69	52	GENAP TH 2015/2016
C.111.11.0026	SHOLIKUN NUR	3.60	50	GENAP TH 2014/2015
C.111.11.0031	FAIZAL MAHMUD	3.26	50	GENAP TH 2014/2015
C.111.11.0032	AHMAD CHOLIL	3.17	50	GENAP TH 2014/2015
C.111.11.0033	MUCHAMMAD MAULANA RIZKA T	3.03	52	GENAP TH 2015/2016
C.111.11.0034	CAHYANI FAJAR SUKOCO	3.54	52	GENAP TH 2015/2016

Tabel 1 Data mahasiswa

#### 3.1 Persiapan Data

Data yang digunakan untuk melakukan penelitian komparasi ini berupa data nilai dari mahasiswa Universitas Semarang yang diambil sampelnya dari program studi Teknik Sipil dan

Teknik Elektro. Didapatkan data sejumlah 840 mahasiswa yang sudah lulus dengan atribut berupa, nim, nilai A - E, Indeks Prestasi Kumulatif (IPK) dan Lama studi.

NIM	IPK	hr	bln	thn	A	B	C	D	E	NIM	IPK	hr	bln	thn	A	B	C	D	E
C.431.09.0004	3.1	0	0	7	18	37	10	2	10	C.431.12.0100	3.56	1	6	4	38	23	4	2	1
C.431.12.0049	3.17	30	11	3	22	37	14	5	0	C.431.12.0080	3.23	1	6	4	22	35	9	1	0
C.431.10.0055	2.77	0	0	6	12	31	29	18	9	C.431.12.0055	3.37	30	11	3	28	32	9	0	3
C.431.12.0010	3.28	30	11	3	28	28	10	1	2	C.411.12.0021	3.22	30	11	3	28	27	10	1	2
C.441.11.0041	3.17	2	6	5	26	31	11	1	8	C.411.12.0019	3.15	30	11	3	21	34	10	0	1
C.431.08.0043	2.97	1	6	8	15	38	14	3	7	C.411.12.0032	3.14	30	11	3	25	28	12	2	0
C.431.11.0080	3.01	2	6	5	19	31	16	2	3	C.431.12.0063	3.46	1	6	4	35	26	4	1	3
C.431.11.0039	2.8	0	0	5	13	33	24	5	11	C.411.12.0002	3.32	30	11	3	31	24	9	2	0
C.431.11.0068	3.01	2	6	5	14	38	18	3	3	C.431.11.0064	3.21	2	6	5	23	33	11	0	3
C.431.12.0071	3.24	1	6	4	28	27	10	3	0	C.411.12.0003	3.1	30	11	3	22	29	13	2	0

Table 2 Data Sample Nilai mahasiswa

Normalisasi data dilakukan dengan tujuan untuk mempermudah perhitungan data-data yang beda. Salah satunya dengan cara dengan melakukan normalisasi jumlah huruf A = 18, kita jadikan presentase dengan mengalikan 10% sehingga

menjadi 1,80. Sedangkan untuk tahun kita normalisasi dengan membulatkan bulan, apabila bulan diatas 6 maka ditambahkan 1 tahun. Maka data akan menjadi seperti table berikut ini.

NIM	IPK	Hr	Bln	Thn	N=thn	A	B	C	D	E	N=a	N=b	N=c	N=d	N=e
C.431.09.0004	3.1	0	0	7	7	18	37	10	2	10	1.8	3.7	1	0.2	1
C.431.12.0049	3.17	30	11	3	4	22	37	14	5	0	2.2	3.7	1.4	0.5	0
C.431.10.0055	2.77	0	0	6	6	12	31	29	18	9	1.2	3.1	2.9	1.8	0.9
C.431.12.0010	3.28	30	11	3	4	28	28	10	1	2	2.8	2.8	1	0.1	0.2
C.441.11.0041	3.17	2	6	5	5	26	31	11	1	8	2.6	3.1	1.1	0.1	0.8
C.431.08.0043	2.97	1	6	8	8	15	38	14	3	7	1.5	3.8	1.4	0.3	0.7
C.431.11.0080	3.01	2	6	5	5	19	31	16	2	3	1.9	3.1	1.6	0.2	0.3
C.431.11.0039	2.8	0	0	5	5	13	33	24	5	11	1.3	3.3	2.4	0.5	1.1
C.431.11.0068	3.01	2	6	5	5	14	38	18	3	3	1.4	3.8	1.8	0.3	0.3
C.431.12.0071	3.24	1	6	4	4	28	27	10	3	0	2.8	2.7	1	0.3	0
C.431.12.0100	3.56	1	6	4	4	38	23	4	2	1	3.8	2.3	0.4	0.2	0.1
C.431.12.0080	3.23	1	6	4	4	22	35	9	1	0	2.2	3.5	0.9	0.1	0
C.431.12.0055	3.37	30	11	3	4	28	32	9	0	3	2.8	3.2	0.9	0	0.3
C.411.12.0021	3.22	30	11	3	4	28	27	10	1	2	2.8	2.7	1	0.1	0.2
C.411.12.0019	3.15	30	11	3	4	21	34	10	0	1	2.1	3.4	1	0	0.1
C.411.12.0032	3.14	30	11	3	4	25	28	12	2	0	2.5	2.8	1.2	0.2	0
C.431.12.0063	3.46	1	6	4	4	35	26	4	1	3	3.5	2.6	0.4	0.1	0.3
C.411.12.0002	3.32	30	11	3	4	31	24	9	2	0	3.1	2.4	0.9	0.2	0
C.431.11.0064	3.21	2	6	5	5	23	33	11	0	3	2.3	3.3	1.1	0	0.3
C.411.12.0003	3.1	30	11	3	4	22	29	13	2	0	2.2	2.9	1.3	0.2	0

Tabel 3 Normalisasi Data Nilai Mahasiswa

Dan hasil akhir normalisasi adalah sebagai berikut :

NIM	IPK	THN	A	B	C	D	E
C.431.09.0004	3.10	7.00	1.80	3.70	1.00	0.20	1.00
C.431.12.0049	3.17	4.00	2.20	3.70	1.40	0.50	0.00
C.431.10.0055	2.77	6.00	1.20	3.10	2.90	1.80	0.90
C.431.12.0010	3.28	4.00	2.80	2.80	1.00	0.10	0.20
C.441.11.0041	3.17	5.00	2.60	3.10	1.10	0.10	0.80
C.431.08.0043	2.97	8.00	1.50	3.80	1.40	0.30	0.70
C.431.11.0080	3.01	5.00	1.90	3.10	1.60	0.20	0.30
C.431.11.0039	2.80	5.00	1.30	3.30	2.40	0.50	1.10
C.431.11.0068	3.01	5.00	1.40	3.80	1.80	0.30	0.30
C.431.12.0071	3.24	4.00	2.80	2.70	1.00	0.30	0.00
C.431.12.0100	3.56	4.00	3.80	2.30	0.40	0.20	0.10
C.431.12.0080	3.23	4.00	2.20	3.50	0.90	0.10	0.00
C.431.12.0055	3.37	4.00	2.80	3.20	0.90	0.00	0.30
C.411.12.0021	3.22	4.00	2.80	2.70	1.00	0.10	0.20
C.411.12.0019	3.15	4.00	2.10	3.40	1.00	0.00	0.10
C.411.12.0032	3.14	4.00	2.50	2.80	1.20	0.20	0.00
C.431.12.0063	3.46	4.00	3.50	2.60	0.40	0.10	0.30
C.411.12.0002	3.32	4.00	3.10	2.40	0.90	0.20	0.00
C.431.11.0064	3.21	5.00	2.30	3.30	1.10	0.00	0.30
C.411.12.0003	3.10	4.00	2.20	2.90	1.30	0.20	0.00

Tabel 4 Normalisasi Ke 2 Data Nilai Mahasiswa

### 3.2 Algoritma *K-Means* dan *Fuzzy C-Means*

Tahapan K-means

1. Inisialisasi : tentukan nilai K sebagai jumlah cluster yang di inginkan dan matrik ketidakmiripan (jarak) yang di inginkan. Jika perlu, tetapkan ambang batas perubahan fungsi objektif dan ambang batas perubahan posisi centroid.
2. Pilik K data dari set data X sebagai centroid
3. Alokasikan semua data ke centroid terdekat dengan metrik jarak yang sudah

ditetapkan (memperbarui cluster ID setiap data)

4. Hitung kembali centroid C berdasarkan data yang mengikuti cluster masing-masing
5. Ulangi langkah 3 dan 4 hingga kondisi konvergen tercapai yaitu (a) perubahan fungsi objektif sudah dibawah ambang batas yang di inginkan, atau (b) tidak ada data yang berpindah cluster, atau (c) perubahan posisi centroid dibawah ambang batas yang ditetapkan.

C1	2.45	2.99	1.50	0.34	0.30	3.13	0.42
C2	3.72	2.39	0.78	0.15	0.15	3.42	0.39
C3	1.90	3.14	2.40	0.89	1.40	2.88	0.54

Tabel 5 Tabel Centroid Akhir

Pada perhitungan *K-Means* didapatkan hasil fungsi objektif “0” pada interaksi ke **13**, dengan validitas *DBI* sebesar **2.021897468**.

### 3.3 *Fuzzy C-Means (FCM)*

1. Inisialisasi : tentukan jumlah cluster ( $k \geq 2$ ), tentukan bobot pangkat ( $w > 1$ , tentukan jumlah maksimal iterasi, tentukan ambang batas perubahan nilai fungsi objektif ( jika perlu juga perubahan nilai centroid)
2. Berikan nilai awal pada matriks *fuzzy pseudo-partition*, dengan syarat seperti pada persamaan

$$\sum_{j=1}^k u_{ij} = 1$$

3. Lakukan langkah 4 sampai 5 selama syarat masih terpenuhi : (1) apabila perubahan pada nilai fungsi objektif masih di atas nilai ambang batas yang ditentukan, (2) perubahan pada nilai centroid masih di atas nilai ambang

batas yang ditentukan, (3) iterasi maksimal belum tercapai.

4. Hitung nilai centroid dari masing-masing cluster menggunakan persamaan
- $$c_{ij} = \frac{\sum_{i=1}^N (u_{il})^w x_{ij}}{\sum_{i=1}^N (u_{il})^w}$$
5. Hitung kembali matriks *fuzzy pseudo-partition* (derajat keanggotaan setiap data pada setiap cluster) menggunakan persama.

Awal	463.50
Nilai J	463.50
Perubahan J	0

Tabel 6 Tabel Nilai Fungsi Objektif Iterasi ke 27  
 Dari hasil perhitungan *FCM*, fungsi objektif “0” pada interkasi ke 27 dengan hasil validitas *XBI* sebesar **11.94809958**.

### 3.4 Validitas

Evaluasi sistem merupakan bagian penting dalam model pekerjaan klasifikasi dengan metode evaluasi diharapkan dapat menghitung keajurasian sebuah metode dengan data tertentu. Disini kita menggunakan dua metode yaitu metode K-means dan FCM, untuk K-Means digunakan validitas Davies-Bouldin Index (DBI) sedangkan FCM kita menggunakan Xie Beni (XBI)

Tahapan perhitungan Validitas **DBI**

1. Menghitung *Sum of square within Cluster (SSW)*

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

2. Menghitung *Sum of square between cluster (SSB)*

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

3. Mendefinisikan ukuran rasio seberapa baik nilai antara cluster ( $R_{i,j}$ )

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}}$$

4. Menghitung *Davis-Bouldin Index (DBI)*

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

Tahapan perhitungan **XBI**

$$XBI = \frac{\sum_{j=1}^K \sum_{i=1}^N u_{ij}^m \times d(x_i, c_j)^2}{N \times \min_{i,j} (d(c)^2)} = \frac{J_m(u, c)/N}{Sep(c)}$$

**4. Kesimpulan**

- a. K-Means Lebih cepat dari pada FCM, karena interaksi K-Means hingga interaksi ke **13** sedangkan FCM hingga interaksi ke **27**

- b. Dari validitas K-Means lebih kecil lebih baik, dengan validitas DBI sebesar **2.021897468**, sedangkan FCM dengan validitas XBI didapatkan validitas sebesar **11.94809958**
- c. Dengan K-Means data mahasiswa tepat waktu di 3 cluster didapatkan sebagai berikut :

	A	B	C	D	E	IPK	THN
C1	2.45	2.99	1.5	0.34	0.3	3.13	0.42
C2	3.72	2.39	0.78	0.15	0.15	3.42	0.39
C3	1.9	3.14	2.4	0.89	1.4	2.88	0.54

Dengan kesimpulan lulusan lama studi 4 tahun lebih 2 bulan dengan ipk rata-rata 3.13, sedangkan lama studi 4 tahun dengan ipk 3.42. dan untuk mahasiswa dengan lama studi 5 tahun lebih rata-rata ipk 2.88

**Daftar Pustaka**

Prasetyo, Eko (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Andi Publisher

Prasetyo, Eko (2013). *Data Mining (Konsep dan Aplikasi Menggunakan Matlab)*. Andi Publisher

Kusrini dan EmhaTaufiq Lutfi (2009). *Algoritma Data Mining*. Andi Publisher

Astuti, Fajar Hermawati (2013). *Data Mining*. Andi Publisher.

Connolly, Thomas, and Carolyn Begg. "Database Systems, 5th." (2010).

Amalia, N., Shaufiah, S., & Sa'adah, S. (2015). Penerapan Teknik Data Mining untuk Klasifikasi Ketepatan Waktu Lulus Mahasiswa Teknik Informatika Universitas Telkom Menggunakan Algoritma Naive Bayes Classifier. *eProceedings of Engineering*, 2(3).