

Implementasi Metode K-Means Dalam Pemetaan Kelompok Mahasiswa Melalui Data Aktivitas Kuliah

Rosmini¹, Abdul Fadlil², Sunardi³

¹Magister Teknik Informatika Universitas Ahmad Dahlan, Yogyakarta

²Sistem Informasi STMIK PPKIA Tarakanita Rahmawati, Tarakan

³Teknik Elektro Universitas Ahmad Dahlan, Yogyakarta

¹rosmini@ppkia.ac.id, ²fadlil@uad.ac.id, ³sunardi@uad.ac.id

Abstract

The graduation rate is very important in the accreditation process to improve the quality of a college. Student group mapping can be done to monitor learning outcomes by looking at student activities. This is important so students can complete their study of education and graduate on time. Data Clustering is one method of Data Mining which is unsupervised. There are two types of clustering data that is often used in the process of data clustering hierarchical (hierarchy) data clustering and non-hierarchical (non hierarchy) data clustering. K-Means is one method of clustering / grouping non hierarchy. The technique of grouping data is simple and fast. In this study built a system to classify students based on lecture activities using K-Means method which aims to predict the study period of students by mengkompokannya to facilitate the study program in supervising and evaluating the development of student studies. The data used as sample amounted to 20 student data, while the criteria used are GPA, presense, campus organization, tuition fee, job and status as a parameter or measuring tool to facilitate in grouping students. From the results of research that has been done obtained 2 clusters that is cluster A with a prediction on time graduated consisting of 10 students and cluster B, graduating not on time which amounted to 10 students.

Keywords : Data Mining, K-Means, Mapping, Akitivitas Lecture

Abstrak

Tingkat kelulusan sangat penting dalam proses akreditasi untuk meningkatkan kualitas suatu perguruan tinggi. Pemetaan kelompok mahasiswa dapat dilakukan untuk memantau hasil belajar dengan melihat aktivitas kuliah mahasiswa. Hal ini penting dilakukan agar mahasiswa dapat menyelesaikan studi pendidikan dan lulus tepat waktu. Data Clustering merupakan salah satu metode *Data Mining* yang bersifat tanpa arahan (*unsupervised*). Ada dua jenis data clustering yang sering dipergunakan dalam proses pengelompokan data yaitu *hierarchical* (hirarki) data clustering dan *non-hierarchical* (non hirarki) data clustering. K-Means merupakan salah satu metode clustering/pengelompokan non hirarki. Teknik pengelompokkan datanya sederhana dan cepat. Pada penelitian ini dibangun sebuah sistem untuk mengelompokan mahasiswa berdasarkan aktivitas kuliah menggunakan metode K-Means yang bertujuan untuk memprediksi masa studi mahasiswa dengan mengelompokannya agar memudahkan program studi dalam mengawasi dan mengevaluasi perkembangan studi mahasiswa. Data yang dijadikan sampel berjumlah 20 data mahasiswa, adapun kriteria yang digunakan yaitu IPK, presensi, organisasi kampus, tanggungan biaya kuliah, pekerjaan dan status sebagai parameter atau alat ukur untuk memudahkan dalam mengelompokan mahasiswa. Dari hasil penelitian yang telah lakukan diperoleh 2 cluster yaitu cluster A dengan prediksi lulus tepat waktu yang terdiri dari 10 mahasiswa dan cluster B, lulus tidak tepat waktu yang berjumlah 10 mahasiswa.

Kata kunci: Data Mining, K-Means, Pemetaan, Akitivitas Kuliah

1. PENDAHULUAN

Suatu perguruan tinggi yang berkualitas apabila telah memenuhi standar nasional pendidikan tinggi (akreditasi) yaitu dapat dilihat dari rata-rata lama studi mahasiswanya dan tingkat kelulusan. Berdasarkan peraturan di dalam buku II standard dan prosedur tentang akreditasi institusi perguruan tinggi oleh BAN-PT (Badan Akreditasi Nasional Perguruan Tinggi), menyatakan bahwa salah satu aspek atau standar penilaian akreditasi adalah mahasiswa dan lulusan. Standar penilaian yaitu penerimaan mahasiswa baru dan kelulusan (masa studi, IPK). Maka, jumlah mahasiswa dan kelulusan sangat penting dalam proses akreditasi untuk meningkatkan kualitas suatu perguruan tinggi. [1]

Data Clustering merupakan salah satu metode *Data Mining* yang bersifat tanpa arahan (*unsupervised*). Ada dua jenis data clustering yang sering dipergunakan dalam proses pengelompokan data yaitu *hierarchical* (hirarki) data clustering dan *non-hierarchical* (non hirarki) data clustering. [5]

K-Means merupakan salah satu metode clustering/pengelompokan non hirarki. Teknik pengelompokan datanya sederhana dan cepat. Ada banyak pendekatan untuk membuat cluster, diantaranya adalah membuat aturan yang mendikte keanggotaan dalam grup yang sama berdasarkan tingkat persamaan diantara anggota - anggotanya. Pendekatan lainnya adalah dengan membuat sekumpulan fungsi yang mengukur beberapa properti dari pengelompokan tersebut sebagai fungsi dari beberapa parameter dari sebuah clustering.

Pada penelitian ini dibangun sebuah sistem untuk mengelompokan mahasiswa berdasarkan aktivitas kuliah menggunakan metode K-Means. Manfaat penelitian ini adalah memprediksi waktu masa studi mahasiswa dengan mengelompokannya agar dapat melakukan tindakan pencegahan terhadap mahasiswa yang memiliki kecenderungan menyelesaikan studi tidak tepat waktu dan diharapkan memudahkan program studi dalam mengawasi dan mengevaluasi perkembangan studi mahasiswa.

Beberapa penelitian oleh penelitian sebelumnya yang berkaitan dengan prediksi waktu masa studi mahasiswa yang dijadikan acuan antara lain, menurut peneliti pertama yaitu Danny Himawan (2014) seorang mahasiswa dinyatakan lulus tepat waktu jika dapat menyelesaikan pendidikan selama 8 semester atau 4 tahun[1]. Peneliti kedua yaitu Dwi A dkk (2017) memprediksi masa studi dengan data kasus yaitu jumlah SKS yang sudah ditempuh saat semester 7, IPK mahasiswa, matakuliah konsentrasi yang diambil mahasiswa dan asal sekolah mahasiswa[2]. Peneliti ketiga yaitu Siska Haryati dkk (2015) memprediksi masa studi mahasiswa dengan memantau hasil belajar berupa nilai IPK dan jumlah SKS dengan menggunakan metode algoritma C4.5[3]. Peneliti keempat yaitu Silvia Lorena dkk (2014) memprediksi masa studi mahasiswa berdasar nilai akademis dengan menggunakan metode algoritma C4.5[4].

2. METODE PENELITIAN

2.1 Pengumpulan Data

Data yang dibutuhkan dalam penelitian ini adalah data IPK, Presensi, Tanggungan biaya Kuliah, Organisasi Kampus, Pekerjaan dan Status. Untuk kebutuhan pengolahan data, maka jenis file yang digunakan adalah file Microsoft Excel (.xls atau.xlsx), dapat dilihat pada tabel 1 sebagai berikut :

Tabel 1. Kriteria Akademis Dan Non Akademis

Kriteria	Keterangan
C ₀₁	IPK
C ₀₂	Presensi
C ₀₃	Tanggungjawab Biaya Kuliah
C ₀₄	Organisasi Kampus
C ₀₅	Pekerjaan
C ₀₆	Status

2.2. Konsep Teori

2.2.1. Konsep Data Mining

Data Mining merupakan teknologi baru yang sangat berguna untuk membantu perusahaan-perusahaan menemukan informasi yang sangat penting dari gudang data mereka. Beberapa aplikasi data mining fokus pada prediksi, mereka meramalkan apa yang akan terjadi dalam situasi baru dari data yang menggambarkan apa yang terjadi di masa lalu.[7]

2.2.2. Implementasi Metode K-Means

Berikut adalah langkah-langkah perhitungan dengan metode K-Means :

1. Tentukan jumlah cluster
2. Alokasikan data ke dalam cluster secara random
3. Hitung *centroid*/rata-rata dari data yang ada di masing-masing cluster
4. Alokasikan masing-masing data ke *centroid*/rata-rata terdekat
5. Kembali ke Step 3, apabila masih ada data yang berpindah cluster atau apabila perubahan nilai *centroid*, ada yang di atas nilai *threshold* yang ditentukan atau apabila perubahan nilai pada *objective function* yang digunakan di atas nilai *threshold* yang ditentukan

Salah satu alternatif penerapan *K-Means* dengan beberapa pengembangan teori-teori penghitungan terkait adalah *Euclidian Distance (L2-Norm)*. Jarak antara dua titik dirumus sebagai berikut :

$$d(x, y) = \|x - y\|^2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Keterangan :

d = determinan (*Euclidian Distance*)

x = titik pusat cluster

y = data

n = jumlah data

i = data ke-

Jarak yang terpendek antara *centroid* dengan dokumen menentukan posisi *cluster* suatu dokumen. Misalnya dokumen A mempunyai jarak yang paling pendek ke *centroid 1* dibanding ke yang lain, maka dokumen A masuk ke grup 1. Hitung kembali posisi *centroid* baru untuk tiap-tiap *centroid (C_{i..j})* dengan mengambil rata-rata dokumen yang

masuk pada *cluster* awal ($G_{i..j}$). Iterasi dilakukan terus hingga posisi *group* tidak berubah. Berikut rumus dari penentuan *centroid*.

$$C(i) = \frac{x_1 + x_2 + x_{..} + x_{...}}{\sum x} \quad (2)$$

Keterangan :

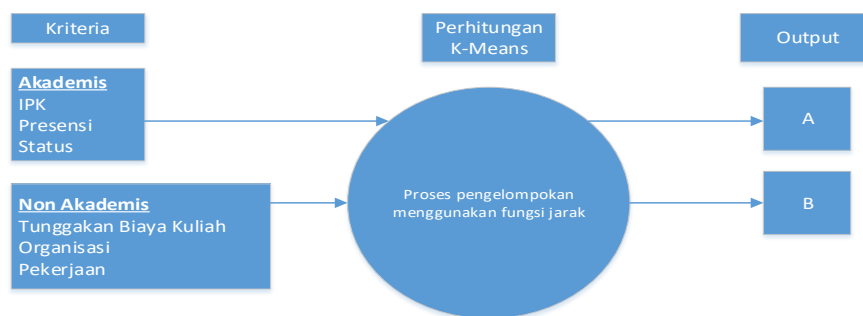
x_1 = nilai data *record* ke-1

x_2 = nilai data *record* ke-2

$\sum x$ = jumlah data *record*

2.3. Perancangan Sistem

Pada gambar rancangan sistem dapat dilihat proses dari sistem yang akan dibuat pertama menentukan kriteria yang akan menjadi inputan. Inputan tersebut digunakan dalam proses selanjutnya yaitu perhitungan k-means. Setelah proses perhitungan clustering selesai maka, ada 2 output atau hasil yaitu A dan B.



Gambar 1. Rancangan Sistem

3. HASIL DAN PEMBAHASAN

Tahap pertama yang akan dilakukan pada penelitian ini yaitu menentukan domain bagi masing-masing kriteria, yang bertujuan untuk mengelompokkan data nilai dari masing-masing mahasiswa.

Tabel 2. IPK

Parameter Ukuran	Bobot Nilai
< 1.00	1
1.01 – 2.00	2
2.01 – 3.00	3
3.01 – 3.50	4
3.51 – 4.00	5

Tabel 3. Presensi

Parameter Ukuran	Bobot Nilai
10	1
11	2
12	3
13	4
14	5

Tabel 4. Tanggungan Biaya Kuliah

Parameter Ukuran	Bobot Nilai
< 1.000.000	5
1.100.000 – 2.000.000	4
2.100.000 – 2.500.000	3
2.550.000 – 3.500.000	2
> 3.500.000	1

Tabel 5. Organisasi Kampus

Parameter Ukuran	Bobot Nilai
Sangat Aktif	1
Aktif	2
Cukup	3
Kurang	4
Tidak Aktif	5

Tabel 6. Pekerjaan

Parameter Ukuran	Bobot Nilai
Bekerja	1
Tidak Bekerja	2

Tabel 7. Status

Parameter Ukuran	Bobot Nilai
Pernah Cuti	1
Tidak Pernah Cuti	2

Tabel 8. Pengelompokan Nilai Histori

No	Nama	Kriteria					
		C01	C02	C03	C04	C05	C06
1	Qolbiah Fitri	5	5	5	1	2	2
2	Marhani	4	5	5	4	2	2
3	Siti Marlina	4	4	5	4	2	2
4	Nuryati	4	4	5	4	2	2
5	Yuliana Fatmawati	4	4	4	3	2	2
6	Novita Lestari	4	3	5	1	1	2
7	Eka Andika	4	2	5	3	1	2
8	Muhammad Habibi	4	5	3	4	2	2
9	Dwi Mei Nengtiyas	4	4	4	4	1	2
10	Abdul Malik	3	3	5	2	1	1

11	Cucun Kurnia	4	5	4	4	2	2
12	Muhammad Aldy Yansyah	4	4	4	1	1	2
13	Rezky Indah Lestari	4	3	4	4	2	2
14	Fahrozi Rahman	4	3	4	3	1	2
15	Rahmat Agung Wicaksono	4	4	1	1	2	2
16	Hamdan	3	3	2	4	1	1
17	Bima Kusuma Aji	5	5	3	4	2	2
18	Riswandi	3	3	2	4	2	2
19	Darusman	4	4	2	4	2	2
20	Vikram	4	4	2	4	2	2

Penentuan awal cluster secara acak :

1. Data **ke- 1** sebagai pusat **Cluster Ke-A**
 Qolbiah Fitri [5,5,5,1,2,2] (Tepat Waktu)
2. Data **ke- 10** sebagai pusat **Cluster Ke-B**
 Abdul Malik [3,3,5,2,1,1] (Tidak Tepat Waktu)

Setelah dilakukan pengelompokan berdasarkan masing-masing cluster kemudian hitung jarak masing-masing. [6]

Tabel 9. Perhitungan Jarak Cluster

No	NIM	C01	C02	C03	C04	C05	C06	A	B	Jarak Terpendek
1	Qolbiah Fitri	5	5	5	1	2	2	0,00	3,32	0,00
2	Marhani	4	5	5	4	2	2	3,16	3,32	3,16
3	Siti Marlina	4	4	5	4	2	2	3,32	2,83	2,83
4	Nuryati	4	4	5	4	2	2	3,32	2,83	2,83
5	Yuliana Fatmawati	4	4	4	3	2	2	2,65	2,45	2,45
6	Novita Lestari	4	3	5	1	1	2	2,45	1,73	1,73
7	Eka Andika	4	2	5	3	1	2	3,87	2,00	2,00
8	Muhammad Habibi	4	5	4	4	2	2	3,32	3,46	3,32
9	Dwi Mei Nengtiyas	4	4	4	4	1	2	3,61	2,83	2,83
10	Abdul Malik	3	3	5	2	1	1	3,32	0,00	0,00
11	Cucun Kurnia	4	5	4	4	2	2	3,32	3,46	3,32
12	Muhammad Aldy Yansyah	4	4	3	1	1	2	2,65	2,83	2,65
13	Rezky Indah Lestari	4	3	4	4	2	2	3,87	2,83	2,83
14	Fahrozi Rahman	4	3	4	3	1	2	3,32	2,00	2,00
15	Rahmat Agung Wicaksono	4	4	1	1	2	2	4,24	4,58	4,24
16	Hamdan	3	3	2	4	1	1	5,29	3,61	3,61
17	Bima Kusuma Aji	5	5	3	4	2	2	3,61	4,24	3,61
18	Riswandi	3	3	2	4	2	2	5,10	3,87	3,87
19	Darusman	4	4	2	4	2	2	4,47	4,12	4,12
20	Vikram	4	4	2	4	2	2	4,47	4,12	4,12

Misalnya menghitung jarak (distance) A, B untuk data pertama menggunakan Eulidean Distance :

$$d(x, y) = \|x - y\|^2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

$$A = \sqrt{(5-5)^2 + (5-5)^2 + (5-5)^2 + (1-1)^2 + (2-2)^2 + (2-2)^2} = 0,00 \quad (\text{minimum})$$

$$B = \sqrt{(5-5)^2 + (5-5)^2 + (5-5)^2 + (1-1)^2 + (2-2)^2 + (2-2)^2} = 3,32$$

Selanjutnya pada tiap cluster yang menjadi centroid awal pada iterasi 1 menggunakan persamaan 01, dapat dilihat pada Tabel 10.

Tabel 10. Pengelompokan Data/Persamaan 01

No	A	B
1	1	0
2	1	0
3	0	1
4	0	1
5	0	1
6	0	1
7	0	1
8	1	0
9	0	1
10	0	1
11	1	0
12	1	0
13	0	1
14	0	1
15	1	0
16	0	1
17	1	0
18	0	1
19	1	0
20	1	0

Tabel 11. Penentuan Pusat Cluster Baru

No	NIM	C01	C02	C03	C04	C05	C06
1	Qolbiah Fitri	5	5	5	1	2	2
2	Marhani	4	5	5	4	2	2
3	Siti Marlina	4	4	5	4	2	2
4	Nuryati	4	4	5	4	2	2
5	Yuliana Fatmawati	4	4	4	3	2	2
6	Novita Lestari	4	3	5	1	1	2

7	Eka Andika	4	2	5	3	1	2
8	Muhammad Habibi	4	5	4	4	2	2
9	Dwi Mei Nengtiyas	4	4	4	4	1	2
10	Abdul Malik	3	3	5	2	1	1
11	Cucun Kurnia	4	5	4	4	2	2
12	Muhammad Aldy Yansyah	4	4	3	1	1	2
13	Rezky Indah Lestari	4	3	4	4	2	2
14	Fahrozi Rahman	4	3	4	3	1	2
15	Rahmat Agung Wicaksono	4	4	1	1	2	2
16	Hamdan	3	3	2	4	1	1
17	Bima Kusuma Aji	5	5	3	4	2	2
18	Riswandi	3	3	2	4	2	2
19	Darusman	4	4	2	4	2	2
20	Vikram	4	4	2	4	2	2

Setelah proses perhitungan dan pembagian kelompok cluster dilakukan, maka pusat cluster baru dapat dibentuk dengan cara menghitung rata-rata kriteria dari masing-masing kelompok cluster yang sama. Misalnya menghitung pusat cluster baru (centroid) menggunakan rumus sebagai berikut :

$$C(i) = \frac{x_1 + x_2 + x_{..} + x_{...}}{\sum x} \quad (4)$$

$$A = \frac{5+4+4+4+4+4+5+4+4}{9} = 4.22 \text{ adalah nilai dari } C01$$

$$B = \frac{4+4+4+4+4+4+3+4+4+3+3}{11} = 3.7 \text{ adalah nilai dari } C01$$

Rata-rata dari kelompok data cluster 1 masuk sebagai nilai A, begitu pula dengan data cluster 2 masuk sebagai nilai B. Maka diperoleh pusat cluster baru/centroid baru seperti pada Tabel 12.

Tabel 12. Pusat Cluster Baru/Centroid Baru

Centroid Baru	C1	C2	C3	C4	C5	C6
A	4,22	4,56	3,22	3,00	1,89	2,00
B	3,73	3,27	4,09	3,27	1,45	1,82

Setelah pusat cluster baru terbentuk, maka perhitungan dilanjutkan untuk menghitung jarak data ke pusat cluster yang baru sekaligus penentuan kelompok cluster. Apabila masih ada perubahan kelompok cluster maka perhitungan terus dilanjutkan. Pada penulisan ini, perhitungan berhenti pada iterasi ke-3 karena sudah tidak mengalami perubahan kelompok cluster. Data akhir yang didapatkan dapat dilihat pada Tabel 13.

Tabel 13. Hasil Akhir

No	NIM	C01	C02	C03	C04	C05	C06	A	B	Cluster
1	Qolbiah Fitri	5	5	5	1	2	2	2,79	3,38	A
2	Marhani	4	5	5	4	2	2	2,05	2,24	A
3	Siti Marlina	4	4	5	4	2	2	2,05	1,56	B
4	Nuryati	4	4	5	4	2	2	2,05	1,56	B
5	Yuliana Fatmawati	4	4	4	3	2	2	0,89	1,11	A
6	Novita Lestari	4	3	5	1	1	2	3,16	2,54	B
7	Eka Andika	4	2	5	3	1	2	3,16	1,62	B
8	Muhammad Habibi	4	5	4	4	2	2	1,34	2,06	A
9	Dwi Mei Nengtiyas	4	4	4	4	1	2	1,61	1,20	B
10	Abdul Malik	3	3	5	2	1	1	3,06	1,96	B
11	Cucun Kurnia	4	5	4	4	2	2	1,34	2,06	A
12	Muhammad Aldy Yansyah	4	4	3	1	1	2	2,28	2,73	A
13	Rezky Indah Lestari	4	3	4	4	2	2	1,95	1,01	B
14	Fahrozi Rahman	4	3	4	3	1	2	1,89	0,66	B
15	Rahmat Agung Wicaksono	4	4	1	1	2	2	3,10	4,00	A
16	Hamdan	3	3	2	4	1	1	2,86	2,50	B
17	Bima Kusuma Aji	5	5	3	4	2	2	1,41	2,65	A
18	Riswandi	3	3	2	4	2	2	2,53	2,41	B
19	Darusman	4	4	2	4	2	2	1,73	2,46	A
20	Vikram	4	4	2	4	2	2	1,73	2,46	A

4. KESIMPULAN

Berdasarkan hasil akhir pengelompokan data setelah proses clustering, dapat disimpulkan sebagai berikut :

1. Cluster A adalah mahasiswa yang lulus tepat waktu sedangkan cluster B adalah mahasiswa yang lulusnya tidak tepat waktu.
2. Data pengelompokan mahasiswa ini merupakan masukan bagi dosen wali dalam membimbing dan mengawasi proses belajar mahasiswa agar bisa lulus tepat waktu.

5. SARAN

Berikut ini beberapa saran yang dapat dipergunakan untuk pengembangan penelitian selanjutnya:

1. Pada penelitian ini dapat dikembangkan menjadi sebuah aplikasi menggunakan bahasa pemrograman.
2. Untuk pengembangan penelitian selanjutnya dapat dikembangkan menggunakan perbandingan algoritma / metode lainnya.

DAFTAR PUSTAKA

- [1] Danny, H. (2014). Aplikasi Data Mining Menggunakan Algoritma ID3 Untuk Mengklasifikasi Kelulusan Mahasiswa. Semarang : Universitas Dian Nuswantoro

-
- [2] Dwi, A., Uning, L dan Edy S. (2017). Implementasu Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5. *Jurnal SCRIPT*, Vol.5 No.1
 - [3] Siska, H., Aji, S dan Eko, S. (2015). Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5. *Jurnal Media Infotama*, Vol.11 No.2.
 - [4] Selvia, L., Wendi, Z dan Ida, H. (2014). Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Memprediksi Masa Studi Mahaiswa Berdasarkan Data Nilai Akademik. *Prosiding Seminar Nasional Aplikasi & Teknologi (SNAST) 2014*.
 - [5] Yudi, A, PhD. (2007). K-Means – Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem dan Informatika* vol.3, pp.47-60.
 - [6] Rosmini, Abdul, F dan Sunardi. (2017). Perancangan Metode K-Means Clustering Dalam Pemberian Dispensasi Pembayaran Kuliah. *Seminar Nasional Teknologi Informasi dan Komunikasi (SEMANTIKOM)*.
 - [7] Mundayani., Des, S. (2016). SPK Penyeleksian Calon Presiden Mahasiswa Universitas Islam Riau Menggunakan Metode Fuzzy-AH. *IT Journal Research and Develoment (ITJRD)*. ISSN : 2528-4061.
-