

KLASTERISASI HASIL EVALUASI AKADEMIK MENGUNAKAN METODE K-MEANS (STUDI KASUS FAKULTAS ILMU KOMPUTER UNILAK)

Yogi Ersan Fadrial

Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Lancang Kuning
Jl. Yos Sudarso KM. 8 Rumbai, Pekanbaru, Riau, telp. 0811 753 2015
e-mail: ¹ yogiersan@unilak.ac.id

Abstrak

Perkembangan ilmu pengetahuan dan teknologi menuntut perguruan tinggi sebagai lembaga pendidikan yang formal, untuk dapat menghasilkan lulusan yang bermutu dan kompeten. Pembelajaran pada perguruan tinggi seharusnya lebih inovatif dan kreatif dalam menghasilkan lulusan serta responsif akan kebutuhan tenaga kerja. Dalam penelitian ini, dilakukan pengelompokan mahasiswa berprestasi dan bermasalah dengan metode Klastering K-means, dilakukan untuk mengklaster data akademik mahasiswa menjadi empat buah klaster, yaitu klaster mahasiswa berprestasi, berpotensi berprestasi, berpotensi bermasalah, dan klaster mahasiswa bermasalah. Hasil perhitungan manual dan pengujian dengan aplikasi menggunakan 20 sampel data akademik mahasiswa mendapatkan hasil yang sama. Di mana hasilnya adalah cluster 1 terdiri dari 5 anggota mahasiswa yang berpotensi berprestasi, cluster 2 terdiri dari 5 anggota mahasiswa yang berpotensi bermasalah, cluster 3 terdiri dari 9 anggota mahasiswa yang berprestasi sedangkan cluster 4 terdiri dari 1 anggota mahasiswa yang bermasalah.

Kata Kunci: Clustering, K-means, Evaluasi, Akademik

Abstract

The development of science and technology demands higher education as a formal educational institution to be able to produce qualified and competent graduates. Learning in tertiary institutions should be more innovative and creative in producing graduates and responsive to the needs of the workforce. In this study, students with achievement and problems were grouped using the K-means cluster method, which was conducted to cluster student academic data into four clusters, namely clusters of high achieving students, potential achievers, potential problems, and problem student clusters, the results of manual calculations and testing with the application using 20 samples of student academic data get the same results. Where the result is cluster 1 consists of 5 student members with potential achievement, cluster 2 consists of 5 student members who have the potential to have problems, cluster 3 consists of 9 student members who excel, while cluster 4 consists of 1 student member with problems.

Keyword: Clustering, K-means, Evaluation, Academic

1. PENDAHULUAN

Data mining adalah ilmu komputer dan analisisnya langkah dengan proses "Penemuan Pengetahuan dalam Database", atau KDD. Ini adalah proses mengenali pola dalam data besar set. Educational Data Mining (EDM) adalah sub-domain dari Data Mining yang berhubungan dengan data dari database akademik yang

digunakan untuk mengembangkan berbagai teknik dan untuk mengenali pola yang unik. Pengetahuan yang didapat bisa kemudian digunakan untuk menawarkan saran kepada para perencana akademik di lembaga pendidikan tinggi untuk meningkatkan pengambilan keputusan mereka. Proses, untuk meningkatkan kinerja akademik mahasiswa, untuk menurunkan tingkat kegagalan, untuk memahami perilaku mahasiswa dalam suatu cara yang lebih baik, untuk membantu *instruktur*, untuk meningkatkan pengajaran, dan untuk membangun *model regresi* dan pohon keputusan untuk diprediksi kinerja siswa dalam hal nilai atau persentase mereka. *Algoritma K-means* merupakan salah satu metode *clustering non hirarki* yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*. Metode ini mempartisi data ke dalam *cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam *cluster* yang lain.

Indeks Prestasi Kumulatif (IPK) adalah indikator yang umum digunakan untuk mengukur kinerja akademik mahasiswa di suatu universitas. Banyak universitas menetapkan IPK minimal yang harus dipertahankan untuk melanjutkan program sarjana. Dan ada juga batas IPK yang dikatakan sebagai indeks prestasi akademik yang baik. Oleh karena itu, IPK masih tetap faktor yang paling umum digunakan oleh para perencana akademik untuk *mengevaluasi* kemajuan akademik setiap mahasiswa. IPK merupakan cerminan kinerja akademik secara keseluruhan mahasiswa selama berada di universitas. Banyak faktor yang dapat menjadi hambatan bagi mahasiswa untuk mencapai dan mempertahankan IPK yang tinggi. Faktor-faktor hambatan ini dapat ditargetkan oleh perencana akademik dalam mengembangkan strategi untuk meningkatkan pembelajaran dan prestasi akademik mahasiswa dengan cara memantau perkembangan kinerja akademik mereka. Penelitian [1] Metode Data Mining dengan algoritma k-means clustering. Data Mining merupakan penggalian informasi dari sejumlah data yang besar. Informasi yang dihasilkan berupa kelompok nama pelanggan yang penggunaannya airnya tergolong boros, sedang dan hemat, jadi pihak PDAM Kab.50 Kota. Berdasarkan 3 cluster yang telah dilakukan pengujian menggunakan RapidMiner bahwa pelanggan terbanyak terdapat pada cluster ke-3 yang tergolong pada pemakaian air hemat.

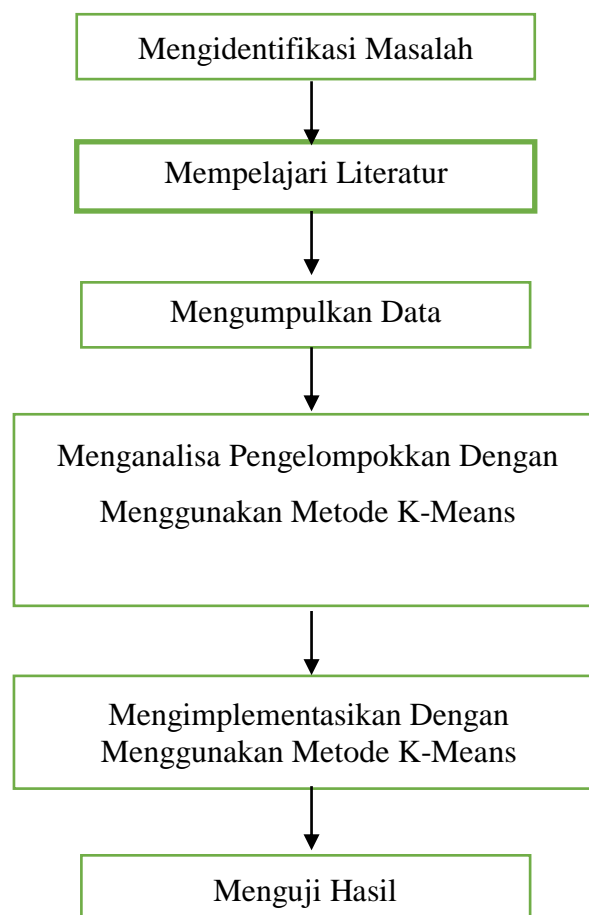
Oleh karena itu, pengelompokan hasil evaluasi akademik mahasiswa adalah salah satu basis untuk memantau perkembangan kinerja akademik mahasiswa di suatu universitas. Penelitian yang penulis lakukan adalah di salah satu Perguruan Tinggi di Pekanbaru-Riau yaitu Universitas Lancang Kuning, tepatnya di Fakultas Ilmu Komputer. Di mana Fakultas tersebut memiliki jumlah mahasiswa hingga tahun 2017 adalah ± 1014 orang. Pengelompokan mahasiswa ke dalam kategori yang berbeda sesuai dengan kinerja mereka telah menjadi tugas yang rumit. Dengan bantuan metode *Klustering K-Means* dapat memungkinkan untuk menemukan karakteristik kunci dari kinerja akademik mahasiswa.

Penelitian [2] Metode K-means clustering bertujuan untuk mempartisi dan observasi ke dalam kluster di mana setiap pengamatan milik kluster dengan jarak terdekat sehingga dapat mengurangi pengaruh outlier pada hasil. [3] Sistem Inferensi Fuzzy Mamdani dalam penentuan jurusan di SMA N 1 Bireuen. Variabel inputnya adalah NIPA, NIPS, IQ, Minat dan kapasitas kelas. Variabel outputnya adalah IPA dan IPS. Dari pengujian data output, diperoleh nilai output IPA dan IPS untuk Sistem Inferensi Fuzzy. Hasil peminatan yang dilakukan oleh algoritma Fuzzy C-Means (FCM) dapat dijelaskan bahwa pada pelaksanaan penjurusan ke kelas XI, dari sebanyak 278 data 82,01% yang tepat

dalam memilih peminatan. Sedangkan penjurusan dengan cara manual berdasarkan pilihan individu siswa hanya 63,67%, sehingga penjurusan dengan FCM lebih tinggi 18,34%. Akurasi peminatan algoritma FCM disajikan pada gambar di atas. [4] Pada penelitian ini, penggunaan algoritma *K-Means Clustering* digunakan untuk pengelompokan penderita *Thalassaemia*. Penentuan titik pusat awal (*centroid*) sangat berpengaruh terhadap jumlah iterasi yang akan dihitung. Penggunaan algoritma *K-means Clustering* pada penelitian ini dapat dijadikan acuan yang dapat dilakukan oleh pengguna. Mulai dari menjadi acuan pelaksanaan tranfusi dan menjaga kadar Hb, persediaan labu darah, dan lain sebagainya.

Penelitian ini bertujuan untuk menganalisa data akademik mahasiswa, membangun pola pengelompokan menggunakan metode Klastering K-Means sehingga dapat menghasilkan pengetahuan atau knowledge dan menguji data yang sudah diolah dengan menggunakan tools yang ada yaitu RapidMiner

2. METODE PENELITIAN



Gambar 1 Kerangka Kerja

Berdasarkan Gambar 1 di atas, berikut adalah penjabaran urutan-urutan langkah kerja sebagaimana berikut :

3.1 Mengidentifikasi Masalah

Masalah yang diidentifikasi dalam penelitian ini adalah untuk mengelompokkan mahasiswa ke dalam kategori yang berbeda sesuai dengan kinerja mereka.

3.2 Mempelajari Literatur

Mempelajari literatur-literatur yang akan dipakai sebagai bahan referensi dalam penelitian ini. Adapun literatur yang dipakai adalah dari jurnal-jurnal ilmiah, modul pembelajaran dan buku tentang *Data Mining* khususnya yang berkaitan dengan Metode *K-Means*. Literatur-literatur ini akan menjadi pedoman untuk melakukan penelitian agar membantu dan memudahkan proses penelitian.

3.3 Mengumpulkan Data

Metode pengumpulan data dilakukan dengan cara meminta data akademik mahasiswa angkatan 2017 yaitu indeks prestasi (IP) semester 1,2 dan 3. Bahan penelitian tersebut diperoleh dari pihak akademik Program Studi Teknik Informatika dan Sistem Informatika, Fakultas Ilmu Komputer, Universitas Lancang Kuning. Selain pengambilan data akademik mahasiswa, juga dilakukan tanya jawab kepada pihak akademik, hal ini dilakukan guna memperoleh data tambahan yang diperlukan di luar dari data akademik mahasiswa tersebut

3.4 Menganalisa Pengelompokkan Dengan Menggunakan Metode K-Means

Pada tahap ini data-data yang telah dikumpulkan akan dianalisa kemudian di disain menjadi suatu pola yang akan diolah dan di implementasikan dengan metode-metode yang telah ditentukan. Disain pola tersebut adalah merubah data-data mentah yang telah diperoleh menjadi data-data yang siap untuk diolah sehingga diharapkan hasil dari pengolahan tersebut dapat sesuai dengan yang diharapkan.

3.4.1 Mengimplementasikan Metode K-Means Menggunakan Program dan Aplikasi

Data-data yang diperoleh, selanjutnya dianalisis dan diimplementasikan guna menghasilkan *informasi* yang berguna. Dalam proses analisa dan implementasi datanya, digunakan aturan-aturan metode *K-Means*. Penganalisaan data dengan metode *K-Means* ini diharapkan akan menghasilkan pengetahuan atau *knowledge* sehingga tujuan dari penelitian ini akan dapat tercapai, peneliti menggunakan Microsoft Office Excel 2007 dan Program Aplikasi *RapidMiner*.

3.4.2 Pengujian Hasil

Pada tahap ini, hasil analisa data yang telah diolah sebelumnya akan diuji dengan menggunakan aplikasi yang ditentukan. Adapun mekanisme pengujian hasil adalah sebagai berikut :

- a. Pertama dari pengumpulan data, yaitu data yang dikumpulkan berupa data akademik mahasiswa angkatan 2017 Program Studi Teknik Informatika yang diperoleh dari pihak Akademik Fakultas Ilmu Komputer dan Sistem Informatika Universitas Lancang Kuning. Kemudian data tersebut akan dianalisa dan ditentukan variabel-variabel dalam penggunaan metode *K-Means*.
 - b. Data yang telah dianalisa, di implementasikan menggunakan aturan-aturan pada metode *K-Means* dan menghitungnya secara manual.
 - c. Kemudian data yang telah dianalisa tersebut juga akan diolah dan diuji dengan menggunakan sebuah *tools* aplikasi yaitu *RapidMiner*.
-

d. Kemudian hasil yang diperoleh dari perhitungan data secara manual menggunakan aturan-aturan metode *K-Means*, akan dibandingkan dengan perhitungan data menggunakan aplikasi *RapidMiner*.

3. HASIL DAN PEMBAHASAN

3.1 Perancangan dan Proses *Clustering*

Pada tahap ini akan dilakukan proses utama yaitu segmentasi atau pengelompokan data akademik mahasiswa Fakultas ilmu komputer Universitas Lancang Kuning dengan menggunakan metode *clustering* algoritma K-Means. Berikut adalah diagram flowchart dari algoritma dengan asumsi banyaknya jumlah *cluster* $k = 4$ sesuai dengan penelitian. Dari banyak data akademik mahasiswa Fakultas ilmu komputer Universitas Lancang Kuning yang diperoleh, di ambil 20 data untuk dijadikan sampel penerapan algoritma k-means dalam pengelompokan hasil evaluasi mahasiswa. Percobaan dilakukan dengan menggunakan parameter-parameter berikut:

- Jumlah *cluster* : 4
- Jumlah data : 20
- Jumlahatribut : 3

Pada tabel dibawah ini merupakan sampel data yang digunakan untuk melakukan percobaan perhitungan manual.

Tabel 1 Sampel Data Akademik Mahasiswa Fasilkom

NO	NIM	NAMA LENGKAP	IPK SMT 1	IPK SMT 2	IPK SMT 3
1	1755201040	JULIUS PRATAMA. T	3.19	3.16	3.25
2	1755201046	RIAN ANANDA PUTRA	3.41	3.43	2.56
3	1755201047	ARBI	3.18	3.41	3.3
4	1755201051	ELMIZA	3.35	3.34	3.35
5	1755201057	MARSEL FIO IPANDI	3.31	3.41	3.25
6	1755201013	RIZKI HADINATA	3.05	2.83	2.7
7	1755201016	RISKI RIDHO SAPUTRA	2.85	2.85	3.25
8	1755201028	RIDHO HIDAYAT	2.55	2.55	2.85
9	1755201032	MUHAMMAD AFRIDEL	2.77	2.77	3.25
10	1755201034	MUHAMMAD ADE RAZIK	2.9	3.05	3
NO	NIM	NAMA LENGKAP	IPK SMT 1	IPK SMT 2	IPK SMT 3
...
19	1755201053	SAPARUDIN	3.3	3.4	3.4

20	1755201054	ARPAN	3.47	3.6	3.6
----	------------	-------	------	-----	-----

Iterasi ke-1

1. Menentukan pusat *cluster* awal.

Menentukan *centroid* awal dilakukan secara acak dari data/objek yang tersedia sebanyak jumlah *cluster* k . Nilai *centroid* awal pada penelitian ini dilakukan pemilihan secara acak, dimana jumlah *centroid* awal dilakukan sebanyak empat *centroid* awal, nilai untuk C1 diambil dari baris data ke-14, nilai C2 diambil dari baris ke-16, nilai C3 diambil dari baris data ke-11, nilai C4 diambil dari baris data ke-8. Berikut ini nilai *centroid* awal pada penelitian:

$$C1 = (2,17 ; 2,17 ; 2,45)$$

$$C2 = (3,5 ; 3,57 ; 3,7)$$

$$C3 = (1,4 ; 1,4 ; 1,4)$$

$$C4 = (2,55 ; 2,55 ; 2,85)$$

2. Menghitung jarak dengan pusat *cluster*

Untuk menghitung jarak setiap data yang ada terhadap pusat *cluster* dalam penelitian ini penulis menggunakan rumus *Euclidean Distance*:

$$d(x,y) = ||x - y|| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad ; i = 1,2,3,\dots,n \dots (4.3)$$

Di mana:

x_i : objek x ke- i

y_i : daya y ke- i

n : banyaknya objek

Berikut ini adalah perhitungan jarak dengan *Euclidean Distance* untuk iterasi 1 dengan centroid 1:

$$D1 = \sqrt{(3,19 - 2,17)^2 + (3,16 - 2,17)^2 + (3,25 - 2,45)^2} = 1,6311039$$

$$D2 = \sqrt{(3,41 - 2,17)^2 + (3,43 - 2,17)^2 + (2,56 - 2,45)^2} = 1,7712425$$

$$D3 = \sqrt{(3,18 - 2,17)^2 + (3,43 - 2,17)^2 + (3,3 - 2,45)^2} = 1,8111322$$

$$D4 = \sqrt{(3,35 - 2,17)^2 + (3,34 - 2,17)^2 + (3,35 - 2,45)^2} = 1,8897883$$

$$D5 = \sqrt{(3,31 - 2,17)^2 + (3,41 - 2,17)^2 + (3,25 - 2,45)^2} = 1,8647252$$

.....

$$D20 = \sqrt{(3,47 - 2,17)^2 + (3,6 - 2,17)^2 + (3,6 - 2,45)^2} = 1,58474$$

Setelah perhitungan jarak dengan *Euclidean Distance* untuk iterasi 1 dengan centroid 1 dilakukan, maka tahapan selanjutnya melakukan perhitungan dengan centroid 2, 3 dan 4 dengan rumus yang sama dengan perhitungan diatas.

3. Pengelompokan data

Alokasikan masing-masing data ke dalam *centroid* yang paling terdekat. Dalam mengalokasikan kembali data ke dalam masing-masing *cluster* didasarkan pada perbandingan jarak antara data dengan *centroid* setiap *cluster* yang ada, data dialokasikan secara tegas ke dalam *cluster* yang mempunyai jarak ke *centroid*

terdekat dengan data tersebut. Berikut ini merupakan hasil perbandingan jarak antara data dengan *centroid* setiap *cluster* yang ada berdasarkan perhitungan jarak dengan *Euclidean Distance* untuk iterasi 1.

Tabel 2 Hasil Pengelompokan Data Iterasi 1

	C1	C2	C3	C4		C1	C2	C3	C4
D1	1.63	0.68	3.12	0.97	0	1	0	0	0
D2	1.77	1.15	3.08	1.26	0	1	0	0	0
D3	1.81	0.54	3.29	1.16	0	1	0	0	0
	C1	C2	C3	C4	C1	C2	C3	C4	
D4	1.89	0.44	3.37	1.23	0	1	0	0	
D5	1.86	0.51	3.33	1.22	0	1	0	0	
D6	1.13	1.32	2.54	0.59	0	0	0	1	
D7	1.25	1.07	2.76	0.58	0	0	0	1	
D8	0.67	1.63	2.18	0.00	0	0	0	1	
D9	1.17	1.17	2.68	0.51	0	0	0	1	
D10	1.27	1.06	2.74	0.63	0	0	0	1	
D11	1.51	3.80	0.00	2.18	0	0	1	0	
D12	2.20	4.49	0.69	2.87	0	0	1	0	
D13	1.87	4.13	0.41	2.53	0	0	1	0	
D14	0.00	2.30	1.51	0.67	1	0	0	0	
D15	0.31	2.59	1.21	0.97	1	0	0	0	
D16	2.30	0.00	3.80	1.63	0	1	0	0	
D17	1.97	0.50	3.42	1.32	0	1	0	0	
D18	2.17	0.25	3.65	1.52	0	1	0	0	
D19	1.92	0.40	3.41	1.26	0	1	0	0	
D20	2.25	0.11	3.74	1.58	0	1	0	0	

Hasil iterasi ke-1 adalah:

- Anggota *cluster* 1 (C1) = 2 orang
- Anggota *cluster* 2 (C2) = 10 orang
- Anggota *cluster* 3 (C3) = 3 orang
- Anggota *cluster* 4 (C4) = 5 orang

Iterasi ke-2

1. Penentuan pusat *cluster* baru

Tentukan posisi centroid baru dengan cara menghitung nilai rata-rata dari data-data yang ada pada centroid yang sama.

$$\text{Pusat } \textit{cluster} \text{ baru} = \frac{\sum_{i=1}^n x_i}{n}; i = 1, 2, 3, \dots, n \dots\dots (4.4)$$

Dimana:

x_i : objek ke- i (anggota *cluster*)

n : banyaknya objek/jumlah objek yang menjadi anggota *cluster*

$$C1 = \left(\frac{2,17 + 2,04}{2}; \frac{2,17 + 2,04}{2}; \frac{2,45 + 2,2}{2}; \right) = (2,105; 2,105; 2,325)$$

$$C2 = \left(\frac{3,19 + 3,41 + 3,18 + 3,35 + 3,31 + 3,5 + 3,5 + 3,5 + 3,3 + 3,47}{10}; \frac{3,16 + 3,43 + 3,41 + 3,34 + 3,41 + 3,57 + 3,39 + 3,57 + 3,4 + 3,6}{10}; \frac{3,25 + 2,56 + 3,3 + 3,35 + 3,25 + 3,7 + 3,23 + 3,45 + 3,4 + 3,6}{10}; \right) = (3,371; 3,428; 3,309)$$

$$C3 = \left(\frac{1,4 + 1 + 1,33}{3}; \frac{1,4 + 1 + 1,33}{3}; \frac{1,4 + 1 + 1}{3}; \right) = (1,24334; 1,24334; 1,13334,)$$

$$C4 = \left(\frac{3,05 + 2,85 + 2,55 + 2,77 + 2,9}{5}; \frac{2,83 + 2,85 + 2,55 + 2,77 + 3,05}{5}; \frac{2,7 + 3,25 + 2,85 + 3,25 + 3}{5}; \right) = (2,824; 2,81; 3,01)$$

2. Menghitung jarak dengan pusat *cluster* yang baru.

Setelah diketahui pusat *cluster* baru, maka dilakukan perhitungan jarak dari masing-masing data terhadap pusat *cluster* dengan menggunakan *Euclidean Distance*. Berikut ini adalah perhitungan jarak dengan *Euclidean Distance* untuk iterasi 2 dengan centroid 1:

$$D1 = \sqrt{(3,19 - 2,105)^2 + (3,16 - 2,105)^2 + (3,25 - 2,325)^2} = 1,17$$

$$D2 = \sqrt{(3,41 - 2,105)^2 + (3,43 - 2,105)^2 + (2,56 - 2,325)^2} = 1,87$$

$$D3 = \sqrt{(3,18 - 2,105)^2 + (3,41 - 2,105)^2 + (3,3 - 2,325)^2} = 1,95$$

$$D4 = \sqrt{(3,35 - 2,105)^2 + (3,34 - 2,105)^2 + (3,35 - 2,325)^2} = 2,03$$

$$D5 = \sqrt{(3,31 - 2,105)^2 + (3,41 - 2,105)^2 + (3,25 - 2,325)^2} = 2,00$$

.....

$$D20 = \sqrt{(3,47 - 2,105)^2 + (3,6 - 2,105)^2 + (3,6 - 2,325)^2} = 2,39$$

Setelah perhitungan jarak dengan *Euclidean Distance* untuk iterasi 2 dengan centroid 1 pusat *cluster* baru, maka tahapan selanjutnya melakukan perhitungan dengan centroid 2, 3 dan 4 dengan rumus yang sama dengan perhitungan diatas.

3. Pengelompokan data

Setelah diketahui jarak dari masing-masing data terhadap pusat *cluster* dengan menggunakan *Euclidean Distance*, maka tahap selanjutnya adalah melakukan pengelompokan data tersebut, dimana *cluster* dengan jarak terdekat diberi nilai satu (1) sedangkan untuk jarak yang jauh diberi nilai nol (0). Berikut hasil pengelompokan data untuk iterasi 2:

Tabel 3 Hasil Pengelompokan Data Iterasi 2

	C1	C2	C3	C4		C1	C2	C3	C4
D1	1.77	0.33	3.46	0.56	0	1	0	0	
D2	1.87	0.75	3.39	0.96	0	1	0	0	
D3	1.95	0.19	3.62	0.76	0	1	0	0	
D4	2.03	0.10	3.71	0.82	0	1	0	0	
D5	2.00	0.09	3.67	0.81	0	1	0	0	
D6	1.25	0.91	2.87	0.38	0	0	0	1	
D7	1.40	0.78	3.11	0.24	0	0	0	1	
D8	0.82	1.29	2.52	0.41	0	0	0	1	
D9	1.32	0.89	3.02	0.25	0	0	0	1	
D10	1.41	0.68	3.08	0.25	0	0	0	1	
D11	1.36	3.41	0.35	2.57	0	0	1	0	
D12	2.05	4.10	0.37	3.26	0	0	1	0	
D13	1.72	3.73	0.18	2.91	0	0	1	0	

D14	0.16	1.94	1.86	1.07	1	0	0	0
D15	0.16	2.22	1.55	1.37	1	0	0	0
D16	2.45	0.44	4.13	1.23	0	1	0	0
D17	2.10	0.16	3.75	0.92	0	1	0	0
D18	2.31	0.24	3.98	1.11	0	1	0	0
D19	2.06	0.12	3.74	0.85	0	1	0	0
D20	2.39	0.35	4.07	1.18	0	1	0	0

Hasil iterasi ke-2 adalah: Anggota *cluster* 1 (C1) = 2 orang, Anggota *cluster* 2 (C2) = 10 orang, Anggota *cluster* 3 (C3) = 3 orang, Anggota *cluster* 4 (C4) = 5 orang
Sedangkan berikut tabel perbandingan hasil pengelompokan data iterasi ke-1 dan iterasi ke-2:

Tabel 4 Hasil Pengelompokan Data Iterasi 1 dan Iterasi 2

Iterasi ke-1					Iterasi ke-2			
C1	C2	C3	C4		C1	C2	C3	C4
0	1	0	0	D1	0	1	0	0
0	1	0	0	D2	0	1	0	0
0	1	0	0	D3	0	1	0	0
0	1	0	0	D4	0	1	0	0
0	1	0	0	D5	0	1	0	0
0	0	0	1	D6	0	0	0	1
0	0	0	1	D7	0	0	0	1
0	0	0	1	D8	0	0	0	1
0	0	0	1	D9	0	0	0	1
0	0	0	1	D10	0	0	0	1
0	0	1	0	D11	0	0	1	0
0	0	1	0	D12	0	0	1	0
0	0	1	0	D13	0	0	1	0

1	0	0	0	D14	1	0	0	0
1	0	0	0	D15	1	0	0	0
0	1	0	0	D16	0	1	0	0
0	1	0	0	D17	0	1	0	0
0	1	0	0	D18	0	1	0	0
0	1	0	0	D19	0	1	0	0
0	1	0	0	D20	0	1	0	0

Karena pada iterasi ke-2 posisi *cluster* sudah tidak berubah / sama dengan posisi *cluster* pada iterasi ke-1 maka proses iterasi selanjutnya tidak perlu dilakukan, dapat disimpulkan bahwa proses iterasi dapat dihentikan pada iterasi ke-2 dengan hasil:

Anggota *cluster* 1 (C1) : {D14, D15} = 2 orang

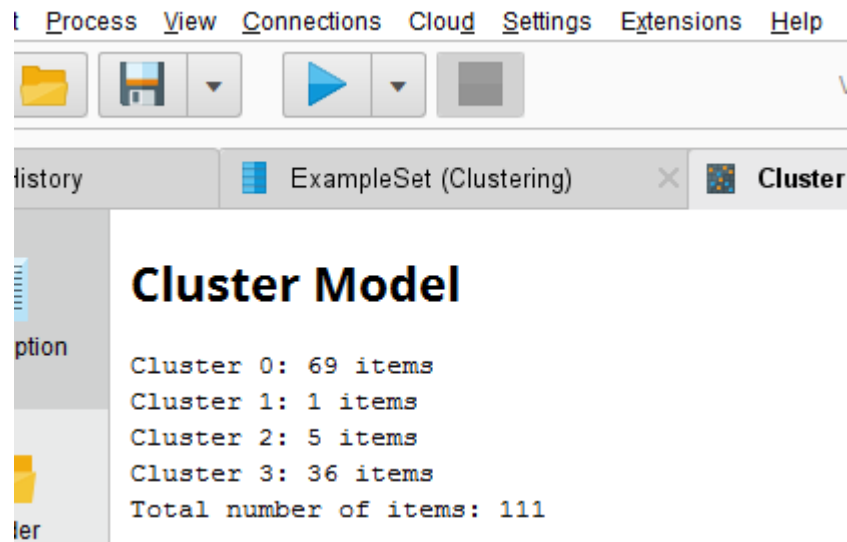
Anggota *cluster* 2 (C2) : {D1, D2, D3, D4, D5, D16, D17, D18, D19, D20} = 10 orang

Anggota *cluster* 3 (C3) : {D11, D12, D13} = 3 orang

Anggota *cluster* 4 (C4) : {D6, D7, D8, D9, D10} = 5 orang

Tabel 5 Perbandingan Hasil K-Means

Perhitungan Manual			Software <i>RapidMiner</i>		
Cluster	Anggota Kelompok	Jumlah Anggota	Cluster	Anggota Kelompok	Jumlah Anggota
1	D14, D15	2	0	D14, D15	2
2	D1, D,2 D3, D4, D5, D16, D17, D18, D19, D20	10	1	D1, D,2 D3, D4, D5, D16, D17, D18, D19, D20	10
3	D11, D12, D13	3	2	D11, D12, D13	3
4	D6, D7, D8, D9, D10	5	3	D6, D7, D8, D9, D10	5



Gambar 2 Tampilan Hasil Cluster (Text View 111 Data)

Berdasarkan gambar 2, dapat dilihat jumlah anggota dari masing-masing *cluster* yaitu *cluster 0* memiliki anggota sebanyak 69 orang, *cluster 1* memiliki anggota sebanyak 5 orang, *cluster 2* memiliki anggota sebanyak 1 orang, *cluster 3* memiliki anggota sebanyak 36 orang. Selanjutnya *Centroid Table* yang merupakan tampilan hasil perhitungan *centroid* akhir K-Means. Tampilan *Centroid Table* dapat dilihat pada gambar 2

Attribute	cluster_0	cluster_1	cluster_2	cluster_3
IPK 1	3.380	0	1.145	2.777
IPK 2	3.406	0	1.145	2.778
IPK 3	3.422	0	1.080	2.889

Gambar 3 Tampilan Hasil Cluster (Centroid Table 111 Data)

Berdasarkan gambar 3, dapat kita lihat bahwa *cluster 0* merupakan kelompok mahasiswa berprestasi, *cluster 1* merupakan kelompok mahasiswa bermasalah, *cluster 2* merupakan kelompok mahasiswa berpotensi bermasalah dan *cluster 3* merupakan kelompok mahasiswa berpotensi berprestasi.

4. KESIMPULAN

Algoritma K-Means dapat diterapkan pada pengelompokan hasil evaluasi mahasiswa Fakultas Ilmu Komputer Universitas Lancang kuning, Berdasarkan hasil perhitungan manual dan pengujian dengan *software RapidMiner* dengan menggunakan 20 sampel data akademik mahasiswa mendapatkan hasil yang sama. Di mana hasilnya adalah *cluster 1* terdiri dari 2 anggota mahasiswa yang berpotensi bermasalah, *cluster 2* terdiri dari 10 anggota mahasiswa yang berprestasi, *cluster 3* terdiri dari 3 anggota mahasiswa yang bermasalah, sedangkan *cluster 4* terdiri dari 5 anggota mahasiswa yang berpotensi berprestasi. Sedangkan pengujian dengan 111 data akademik mahasiswa teknik informatika menggunakan *software RapidMiner* hasilnya adalah *cluster 0* terdiri dari 69 anggota mahasiswa yang berprestasi, *cluster 1* terdiri dari 5 anggota mahasiswa

yang bermasalah, *cluster 2* terdiri dari 1 anggota mahasiswa yang berpotensi bermasalah sedangkan *cluster 3* terdiri dari 36 anggota mahasiswa yang berpotensi berprestasi. Hasil pengelompokan data akademik mahasiswa dapat berfungsi sebagai acuan bagi perencana akademik untuk memantau dan mengevaluasi perkembangan kinerja akademik setiap mahasiswa.

DAFTAR PUSTAKA

- [1] Asroni, dan Adrian R, 2015 "Penerapan *Metode K-Means* Untuk *Clustering* Mahasiswa Berdasarkan Nilai Akademik Dengan *Weka Interface*" Studi Kasus Pada Jurusan *Teknik Informatika* UMM Magelang " jurnal ilmiah semesta teknika vol. 18, no. 1, 76-82.
- [2] Fan Z, dan Yan S, 2016 "*Clustering of College Students Based on Improved K-means Algorithm*", © IEEE and DOI 10.1109/ICS.2016.138, *International Computer Symposium*. Suwirmayanti.
- [3] Heni S, dan Acep Irham G, 2017. "Penerapan Data Mining Dalam Pengelompokan Penderita *Thalassaemia*" jurnal *teknologi dan sistem informasi* - vol. 03 no. 02 299-305.
- [4] Jacob j, Jha K, Kotak P, dan Puthran S, 2015. "*Educational Data Mining Techniques and their Applications*" ©IEEE.
- [5] Metisen B M, dan Sari H L, 2015 "*Analisis Clustering Menggunakan Metode K-Means* Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila", *Jurnal Media Infotama* Vol. 11 No. 2.
- [6] Nainjha R, 2015. "*Data Mining And Knowledge Discovery In Database*" *International Journal of Engineering Trends and Technology (IJETT)* - Volume23 No2.
- [7] Jaroji, Danuri, Fajri P P, 2016 "K-Means Untuk Menentukan Calon Penerima Beasiswa Bidik Misi Di Polbeng" jurnal inovtek polbeng - seri informatika, Vol. 1, No. 1
- [8] Siska, S T, 2016. "Analisa Dan Penerapan Data Mining Untuk Menentukan Kubikasi Air Terjual Berdasarkan Pengelompokan Pelanggan Menggunakan *Algoritma K-Means Clustering*" *Jurnal Teknologi Informasi & Pendidikan*, 9(1), pp.86-93.
- [9] Soeleman A, Syukur A, dan Reza A, 2017 "Penentuan Jurusan Siswa Sekolah Menengah Atas Disesuaikan Dengan Minat Siswa Menggunakan *Algoritma Fuzzy C-Means*", *Jurnal Teknologi Informasi, Volume 13 Nomor 1*.
- [10] Singh I, Sabitha A, dan Bansal A, 2016 "*Student performance analysis using clustering algorithm*" 978-1-4673-8203-8/16/ IEE.
- [11] Widya S A, dan Dedy A, 2016 " Pengelompokan Minat Baca Mahasiswa Menggunakan Metode *K-Means*" *Jurnal Ilmiah ILKOM* Volume 8 Nomor 2.

