

Aplikasi K-Means Clustering Untuk Mengelompokkan Data Kinerja Akademik Mahasiswa

Rudi Rosadi¹, Akmal², Akik Hidayat³, Bagus Kharismawan⁴

^{1,2,3}Prodi Teknik Informatika, Departemen Ilmu Komputer

Fakultas MIPA Universitas Padjadjaran

Jalan Raya Bandung Sumedang km-21

e-mail: r.rosadi@unpad.ac.id, akmal@unpad.ac.id, akik@unpad.ac.id, baguskharismawan@gmail.com

Abstrak – Indeks Prestasi Kumulatif (IPK) biasanya dipergunakan sebagai indikator untuk melihat kinerja akademik mahasiswa. Evaluasi kinerja akademik adalah salah satu dasar untuk memantau perkembangan prestasi akademik mahasiswa di tingkat pendidikan yang lebih tinggi. Dalam mengevaluasi data kinerja akademik mahasiswa, ada kalanya data mahasiswa tersebut dikelompokkan terutama bila data tersebut berjumlah besar, sehingga pola hubungan data di didalam kelompok maupun antar kelompok dapat diungkap. Pengelompokkan tersebut dapat dilakukan dengan metode *clustering*, salah satunya dengan algoritma K-Means. Selanjutnya algoritma ini dapat diterapkan pada sekumpulan data nilai mahasiswa dari Fakultas FMIPA Unpad.

Kata kunci: *Clustering*, Algoritma *K-Means*, Indeks Prestasi Kumulatif (IPK), Kinerja Akademik.

1. Pendahuluan

Indeks Prestasi Kumulatif (IPK) merupakan angka yang menunjukkan prestasi atau kinerja akademik mahasiswa secara kumulatif mulai dari semester pertama hingga semester akhir yang pernah ditempuh. Evaluasi kinerja akademik adalah salah satu dasar untuk memantau perkembangan prestasi akademik mahasiswa di tingkat pendidikan yang lebih tinggi. Dalam mengevaluasi data kinerja akademik mahasiswa, ada kalanya data mahasiswa tersebut dikelompokkan terutama bila data tersebut berjumlah besar, sehingga pola hubungan data di didalam kelompok maupun antar kelompok dapat diungkap.

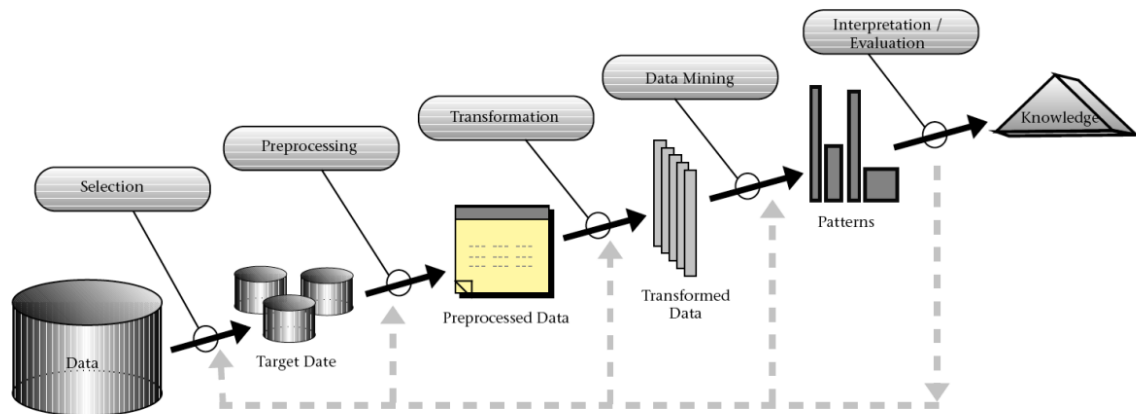
Dengan bantuan teknik *clustering*, mahasiswa dapat dikelompokkan berdasarkan prestasi akademiknya. *Clustering* adalah salah satu metode dari *data mining* yang merupakan proses pengelompokan objek data ke dalam kelas-kelas berbeda yang disebut *cluster* sehingga objek yang berada dalam suatu *cluster* mempunyai perbedaan dengan objek yang berada dalam suatu *cluster* lainnya. Salah satu metode yang digunakan dalam *clustering* adalah algoritma *K-Means*. Algoritma ini dapat digunakan untuk mengumpulkan sekumpulan objek kedalam sejumlah *cluster* yang diinginkan.

2. Metode Penelitian

2.1 Data Mining

Data Mining mengacu kepada proses atau metode untuk mengekstrak atau “menambang” pengetahuan atau pola dari sejumlah data yang besar. *Data mining* muncul diakibatkan oleh tersedianya jumlah data yang besar dan kebutuhan untuk mengubah data tersebut menjadi informasi yang berguna atau pengetahuan (Jiawei Han dan Micheline Kamber, 2006).

Knowledge Discovery in Database (KDD) adalah keseluruhan proses untuk mengkonversi data mentah menjadi suatu pengetahuan yang bermanfaat (Jiawei Han dan Micheline Kamber, 2006). Pada gambar 1 menunjukkan tahapan-tahapan untuk mendapatkan *knowledge* dari proses *data mining*.



Gambar 1 Proses KDD

2.2 Algoritma *K-Means*

Metode *K-Means cluster* adalah metode pengelompokan yang bertujuan mengelompokkan individu sedemikian hingga jarak setiap individu ke pusat kelompok dalam satu kelompok adalah minimum (Dillon, 1984). Proses *clustering* dengan menggunakan algoritma *K-Means* memiliki langkah-langkah sebagai berikut :

1. Tentukan jumlah *cluster* (sebanyak *K cluster*)
2. Menentukan *centroid* setiap *cluster* secara acak.
3. Menghitung jarak setiap objek ke *centroid*
4. Memasukkan objek ke *centroid* terdekat.
5. Perbaharui kembali *centroid* setiap *cluster*.
6. Kembali ke langkah 3, apabila masih ada data yang berpindah *cluster*.

Jarak antara data dengan centroid dihitung dengan menggunakan rumus *Euclidian* sebagai berikut :

$$C_n(x, y) = \sum_{n=1}^m \sqrt{|x_n - y_n|^2} \tag{1}$$

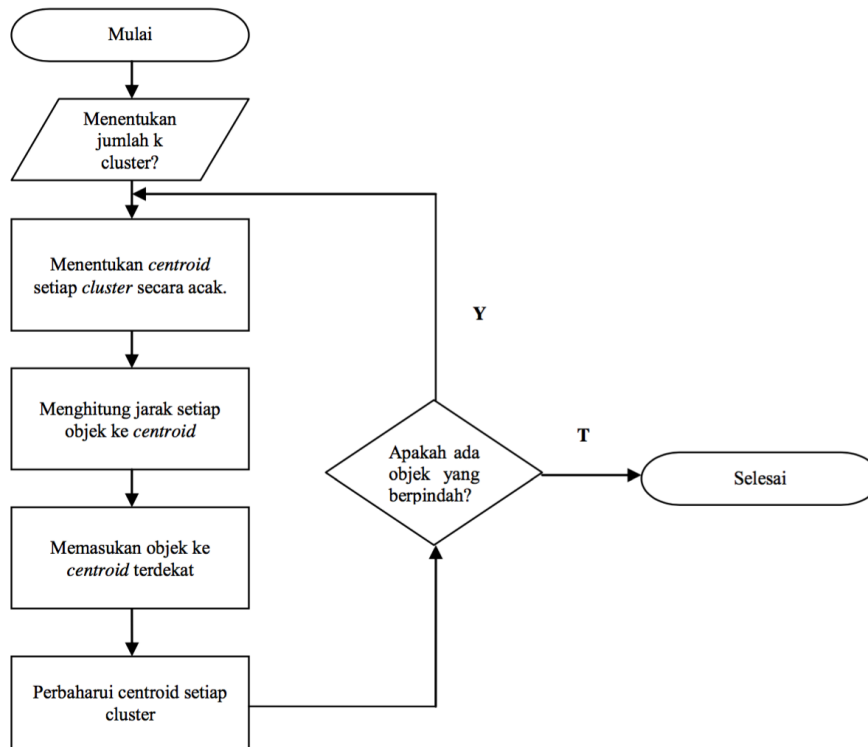
Dimana :

x = Objek Data.

y = Data *Cluster*.

m= Jumlah Atribut.

Langkah-langkah Algoritma *K-Means* dapat digambarkan dengan diagram pada gambar 2 :



Gambar 2 Diagram Langkah-langkah Algoritma K-Means

2.3 Kinerja Akademik

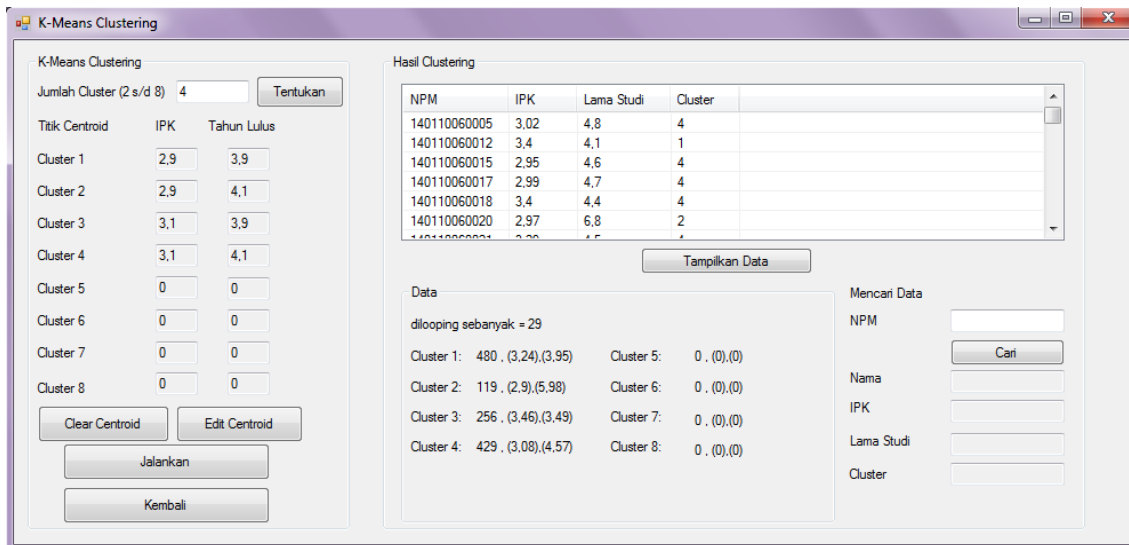
Kinerja akademik merupakan suatu hasil yang dicapai dalam bidang akademik. Evaluasi kinerja akademik mahasiswa dapat dilihat dari IPK (Indeks Prestasi Kumulatif) mahasiswa tersebut. Indeks Prestasi Kumulatif (IPK) merupakan angka yang menunjukkan prestasi atau kinerja akademik mahasiswa secara kumulatif mulai dari semester pertama sampai dengan semester akhir yang telah ditempuh (<http://unpad.ac.id>, 2013).

3. Hasil dan Pembahasan

3.1 Implementasi Program

Implementasi metode *K-Means* dapat dilihat dalam uji coba pengelompokan data menggunakan aplikasi. Pada penelitian ini jumlah data yang digunakan ada sebanyak 1284 *record* data yang akan dikelompokkan menjadi empat buah *cluster*.

Hasil clustering yang telah dilakukan oleh sistem akan ditampilkan seperti pada gambar 3. Dari gambar tersebut terlihat bahwa telah seluruh data mahasiswa telah dikelompokkan oleh sistem sebanyak 4 *cluster* dengan *centroid* yang berbeda-beda. Pengkategorian yang digunakan untuk mengkategorikan tiap *cluster* terdapat pada tabel 1.



Gambar 3 Proses K-Means Clustering

Tabel 1 Kategori Evaluasi Kinerja Akademik Mahasiswa berdasarkan Proses *Clustering*

Cluster	Centroid IPK	Centroid Lama Studi	Kategori
1	3,24	3,95	“Baik”
2	2,9	5,98	“Kurang Baik”
3	3,46	3,49	“Sangat Baik”
4	3,08	4,57	“Cukup Baik”

Hasil pengelompokan data akademik mahasiswa yang dikaitkan dengan jurusan/departemen yang dikaitkan dengan jurusan yang ditempuh mahasiswa dapat dilihat pada tabel 2 dan dalam bentuk prosentase dapat dilihat pada tabel 3.

Tabel 2 Jumlah Mahasiswa per-cluster menurut Departemen yang ditempuh

Departemen	Hasil Pengelompokan				Total Mahasiswa per-departemen
	Cluster 1 (3.24,3.95)	Cluster 2 (2.9,5.98)	Cluster 3 (3.46,3.49)	Cluster 4 (3.08,4.57)	
Matematika	79	27	73	79	258
Fisika	87	27	32	81	227
Kimia	124	24	31	90	269
Biologi	67	28	3	122	220
Statistika	123	13	117	57	310
Total	480	119	256	429	1284

Tabel 3 Jumlah Mahasiswa per-*cluster* Menurut Departemen yang Ditempuh (Dalam Persen) dengan proses *k-means clustering*

Departemen	Hasil Pengelompokan dalam %				Total Mahasiswa per-departemen
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	
	(3.24,3.95)	(2.9,5.98)	(3.46,3.49)	(3.08,4.57)	
Matematika	30,62	10,47	28,29	30,62	100
Fisika	38,33	11,89	14,1	35,68	100
Kimia	46,1	8,92	11,52	33,46	100
Biologi	30,45	12,73	1,36	55,45	100
Statistika	39,68	4,19	37,74	18,39	100
Total	37,38	9,27	19,94	33,41	100

Dari tabel 3 diatas, dapat dirangkum bahwa terdapat 19.94% mahasiswa termasuk dalam *cluster 3* (kategori “Sangat Baik”), lalu diikuti dengan 37,38% mahasiswa di *cluster 1* (kategori “Baik”), 33,41% mahasiswa di *cluster 4* (kategori “Cukup Baik”), dan 9,27% di *cluster 2* (kategori “Kurang Baik”). Selain itu dapat dilihat persentase mahasiswa lulusan departemen Statistika yang termasuk dalam *cluster 3* (“Sangat Baik”) adalah yang tertinggi, diikuti oleh mahasiswa departemen Matematika, Fisika, Kimia dan Biologi.

4. Kesimpulan

Berdasarkan hasil dan pembahasan yang telah dilakukan, kesimpulan yang dapat diambil dari penelitian ini adalah sebagai berikut :

Metode *K-Means* dapat diimplementasikan ke dalam program untuk mengelompokkan data akademik mahasiswa. Berdasarkan hasil proses *clustering* terhadap 1284 data akademik mahasiswa menggunakan aplikasi. 4 *cluster* cocok digunakan sebagai hasil pengelompokan data mahasiswa karena setiap *cluster* tidak memiliki *centroid* yang berdekatan sehingga dapat memiliki kategori yang berbeda-beda. Cluster dengan kataegori kinerja akademik terbaik yaitu cluster 3 (“Sangat Baik”) berjumlah 256 mahasiswa, diikuti oleh 480 mahasiswa di cluster 1 (“Baik”), 429 mahasiswa di cluster 4 (“Cukup Baik”), dan 119 mahasiswa di cluster 2 (“Kurang Baik”). Sedangkan pada hasil pengelompokkan dengan metode *fuzzy c-means*, cluster dengan kataegori kinerja akademik terbaik yaitu cluster 4 (“Sangat Baik”) berjumlah 570 mahasiswa, diikuti oleh 422 mahasiswa di cluster 2 (“Baik”), 232 mahasiswa di cluster 3 (“Cukup Baik”), dan 60 mahasiswa di cluster 1 (“Kurang Baik”). Dari hasil pengelempokkan berdasarkan algoritma *K-Means* yang dikaitkan dengan Jurusan/Departement yang ditempuh, presentase mahasiswa lulusan Departemen Statistika, yang mendapatkan kinerja akademik “Sangat Baik” adalah yang tertinggi.

Daftar Pustaka

- [1] Han, Jiawei.,Kamber, Micheline. 2006. *Data Mining Concepts and Techniques*. San Francisco: Morgan Kaufmann Publishers.
- [2] Jang, JSR Sun, CT dan Mizutani, E. 2004. *Neuro Fuzzy and Soft Computing*. Singapore. Pearson Education.
- [3] Kusrini., Luthfi. 2009, *Algoritma Data Mining*. Yogyakarta : Andi Offset.
- [4] Arwan, Ahmad Khoiruddin. 2007; *Menentukan Nilai Akhir Kuliah dengan Fuzzy C-Means*. Bali ; Universitas Islam Indonesia.
- [5] Bagus, Kharismawan. 2015. *Aplikasi K-Means dan Fuzzy C-Means Clustering Untuk Mengelompokkan Data Mahasiswa FMIPA Unpad*. Bandung :Universitas Padjadjaran.