

IMPLEMENTASI METODE KLASTERING K-MEANS UNTUK MENGELOMPOKAN HASIL EVALUASI MAHASISWA

FEBRIZAL ALFARASY SYAM

Dosen STMIK Dharmapala Riau

ABSTRAK

Mengelompokan hasil evaluasi akademik mahasiswa adalah salah satu basis untuk memantau perkembangan kinerja akademik mahasiswa di suatu universitas. Pengelompokan mahasiswa ke dalam kategori yang berbeda sesuai dengan kinerja mereka telah menjadi tugas yang rumit. Dalam mengelola data akademik mahasiswa penulis menggunakan metode *Data Mining* dengan Algoritma K-means *Clustering*. Hasil penelitian ini berupa pengelompokan mahasiswa berdasarkan *cluster* mahasiswa berprestasi, *cluster* mahasiswa berpotensi berprestasi, *cluster* mahasiswa berpotensi bermasalah dan *cluster* mahasiswa bermasalah. Selanjutnya pengujian dilakukan dengan menggunakan aplikasi *RapidMiner*, yang hasilnya sama dengan perhitungan analisa Algoritma K-Means yang dilakukan.

Kata Kunci : *Data Mining, Cluster, K-Means Clustering, RapidMiner.*

1. PENDAHULUAN

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. Mengelompokan 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 Riau, tepatnya di Fakultas Keguruan dan Ilmu Pendidikan. Dimana Fakultas tersebut memiliki jumlah mahasiswa hingga tahun 2015 adalah ± 5418 orang.

Data Mining merupakan penggalian makna yang tersembunyi dari kumpulan data yang sangat besar. Karena itu *Data Mining* sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik dan basis data (Aprilla et al. 2013). Penggunaan Algoritma K-Means dinilai dapat dengan

cepat dan efisien membantu memantau perkembangan kinerja mahasiswa disuatu instansi pendidikan (Arora & Badal 2013)

Dalam penelitian ini, dilakukan pengelompokan mahasiswa berprestasi dan bermasalah dengan metode Klastering K-means. 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.

2. KAJIAN LITERATUR

2.1 *Knowledge Discovery in Database (KDD)*

Data Mining sering dianggap sebagai bagian dari *knowledge discovery in database (KDD)* yaitu sebuah proses mencari pengetahuan yang bermanfaat dari data, proses KDD secara garis besar dapat dijelaskan sebagai berikut (Sunjana 2010):

1. *Data Selection*

Pemilihan (Seleksi) data dari sekumpulan data operasional perlu

dilakukan sebelum tahap penggalian informasi dalam KDD dimulai.

2. Pre-processing/Cleaning

Sebelum proses *Data Mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak.

3. Transformation

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *Data Mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat bergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. Data Mining

Data Mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu.

5. Interpretation/Evaluation

Pola informasi yang dihasilkan dari proses *Data Mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

2.2 Data Mining

Data Mining adalah proses mencari pola atau informasi menarik dalam data terpilih menggunakan teknik atau metode tertentu. Teknik, metode atau algoritma dalam *Data Mining* sangat bervariasi. *Data Mining* mampu menganalisa data yang besar menjadi informasi berupa pola yang

mempunyai arti bagi pendukung keputusan (Sunjana 2010).

2.3 Clustering

Adapun tujuan dari data *Clustering* ini adalah untuk meminimalisasikan *objective function* yang diset dalam proses *Clustering*, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu *cluster* dan memaksimalisasikan variasi antar *cluster* (Suwirmayanti et al. 2014).

Analisis *cluster* dapat dibagi menjadi teknik pengelompokan *hierarchical* (hirarki) dan teknik pengelompokan *non-hierarchical* (non-hirarki). Contoh teknik hirarki adalah *single linkage*, *complete linkage*, *average linkage*, *median* dan *Ward*. Sedangkan teknik non-hirarkis yaitu *k-means*, *adaptif k-means*, *k-medoids*, dan *fuzzy clustering*. Untuk menentukan algoritma yang baik adalah dilihat dari jenis data yang tersedia dan tujuan tertentu dari analisis (Oyelade et al. 2010).

Teknik pengelompokan saat ini dapat diklasifikasikan menjadi tiga kategori yaitu *partitional*, *hirarkis* dan *berbasis lokalitas* algoritma. Terdapat satu set objek dan kriteria *clustering* atau pengelompokan, pengelompokan *partitional* memperoleh partisi objek ke dalam *cluster* sehingga objek dalam *cluster* akan lebih mirip dengan benda-benda yang ada di dalam *cluster* dari pada objek yang terdapat pada *cluster* yang berbeda. *Partitional* mencoba untuk menguraikan dataset ke satu set *cluster* dengan menentukan jumlah *cluster* awal yang diinginkan (Varghese et al. 2011).

2.4 Algoritma K-Means

Algoritma *K-means* merupakan salah satu algoritma dengan *partitional*, karena *K-Means* didasarkan pada penentuan jumlah awal kelompok dengan mendefinisikan nilai *centroid* awalnya (Madhulatha 2012).

K-Means termasuk dalam metode *Data Mining partitioning clustering* yaitu setiap data harus masuk dalam *cluster* tertentu dan memungkinkan bagi setiap data yang termasuk dalam *cluster* tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke *cluster* lain. K-Means memisahkan data ke *K* daerah bagian terkenal karena kemudian dan kemampuannya untuk mengklasifikasi data besar dan *outlier* dengan sangat cepat (Siska 2016).

Dalam penyelesaiannya, algoritma K-Means akan menghasilkan titik *centroid* yang dijadikan tujuan dari algoritma K-Means. Setelah iterasi K-Means berhenti, setiap objek dalam dataset menjadi anggota dari suatu *cluster*. Nilai *cluster* ditentukan dengan mencari seluruh objek untuk menemukan *cluster* dengan jarak terdekat ke objek. Algoritma K-means akan mengelompokan item data dalam suatu dataset ke suatu *cluster* berdasarkan jarak terdekat (Bhoomi 2014).

Berikut ini langkah-langkah yang terdapat pada algoritma K-Means (Siska 2016):

1. Tentukan *k* sebagai jumlah *cluster* yang dibentuk
Untuk menentukan banyaknya *cluster k* dilakukan dengan beberapa pertimbangan seperti pertimbangan teoritis dan konseptual yang mungkin diusulkan untuk menentukan berapa banyak *cluster*.
2. Bangkitkan *k centroid* (titik pusat *cluster*) awal secara random
Penentuan *centroid* awal dilakukan secara random/acak dari objek-objek yang tersedia sebanyak *k cluster*, kemudian untuk menghitung *centroid cluster ke-i* berikutnya, digunakan rumus sebagai berikut :

$$v = \frac{\sum_{i=1}^n x_i}{n} \quad ; \quad i=1,2,3,\dots,n \quad (1)$$

Di mana $\Leftrightarrow v$: *centroid* pada *cluster*
 X_i : objek ke-*i*

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

3. Hitung jarak setiap objek ke masing-masing *centroid* dari masing-masing *cluster*. Untuk menghitung jarak antara objek dengan *centroid* dapat menggunakan Euclidian Distance

$$d(x,y) = \|x - y\|$$

$$= \sqrt{\sum_{i=1}^n (x_i - y_i)^2} ; i = 1,2,3,\dots,n$$

Di mana $\Leftrightarrow x_i$: objek *x* ke-*i*
 y_i : daya *y* ke-*i*
 n : banyaknya objek

4. Alokasikan masing-masing objek ke dalam *centroid* yang paling dekat.
Untuk melakukan pengalokasian objek kedalam masing-masing *cluster* pada saat iterasi secara umum dapat dilakukan dengan cara hard K-Means di mana secara tegas setiap objek dinyatakan sebagai anggota *cluster* dengan mengukur jarak kedekatan sifatnya terhadap titik pusat *cluster* tersebut.
5. Lakukan iterasi, kemudian tentukan posisi *centroid* baru dengan menggunakan persamaan (1).
6. Ulangi langkah 3 jika posisi *centroid* baru tidak sama.
Pengecekan *konvergensi* dilakukan dengan membandingkan matriks *group assignment* pada iterasi sebelumnya dengan matrik *group assignment* pada iterasi yang sedang berjalan. Jika hasilnya sama maka *algoritma kmeans cluster analysis* sudah *konvergen*, tetapi jika berbeda maka belum *konvergen* sehingga perlu dilakukan iterasi berikutnya.

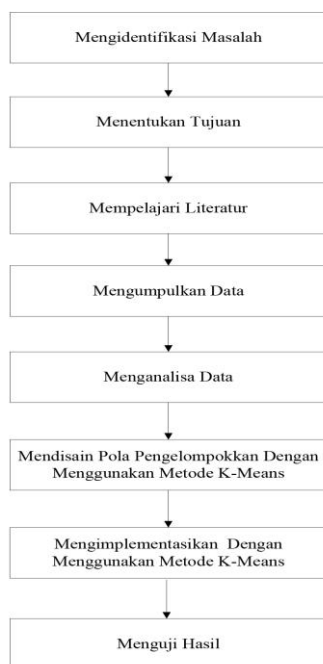
Pada penerapan metode *K-Means Cluster Analysis*, data yang bisa diolah dalam perhitungan adalah data numerik yang berbentuk angka. Sedangkan data

selain angka juga bisa diterapkan tetapi terlebih dahulu harus dilakukan pengkodean untuk mempermudah perhitungan jarak/kesamaan karakteristik yang dimiliki dari setiap objek. Setiap objek dihitung kedekatan jaraknya berdasarkan karakter yang dimiliki dengan pusat *cluster* yang sudah ditentukan sebelumnya, jarak terkecil antara objek dengan masing-masing *cluster* merupakan anggota *cluster* yang terdekat. Setelah jumlah *cluster* ditentukan, selanjutnya dipilih sebanyak 3 objek secara acak sesuai jumlah *cluster* yang dibentuk sebagai pusat *cluster* awal untuk dihitung jarak kedekatannya terhadap semua objek yang ada (Ediyanto et al. 2013).

3. METODE PENELITIAN

3.1 Kerangka Kerja

Dalam metodologi penelitian ada urutan kerangka kerja yang harus diikuti, urutan kerangka kerja ini merupakan gambaran dari langkah-langkah yang harus dilalui agar penelitian ini bisa berjalan dengan baik. Kerangka kerja yang harus diikuti bisa dilihat pada gambar 3.1.



Gambar 3.1 Kerangka Kerja

4. HASIL DAN PEMBAHASAN

4.1 Analisa Pengelompokan Data

Dalam penelitian ini, dilakukan pengelompokan atau kluster mahasiswa berprestasi dan bermasalah berdasarkan data akademik mahasiswa FKIP Universitas Riau. Dilakukan pengklasteran data akademik mahasiswa menjadi empat buah kluster, yaitu kluster mahasiswa berprestasi, berpotensi berprestasi, berpotensi bermasalah dan mahasiswa bermasalah. Jadi pada penelitian ini kluster yang akan dibentuk adalah sebanyak empat kelompok atau nilai $k = 4$. Di mana atribut yang digunakan adalah sebanyak 3 buah atribut yaitu Indeks Prestasi 1 (IP1), Indeks Prestasi 2 (IP2), Rata-rata Kehadiran (RK). Berdasarkan hasil praproses data, jumlah data yang diperoleh sebanyak 94 data, kemudian akan dipilih secara acak data sebanyak 20 buah. Data tersebut akan digunakan sebagai sampel untuk melakukan proses analisa *clustering* algoritma K-Means yang dilakukan perhitungan secara manual menggunakan aturan-aturan algoritma K-means yang telah ditetapkan.

Tabel 4.1 Sampel Data Akademik Mahasiswa

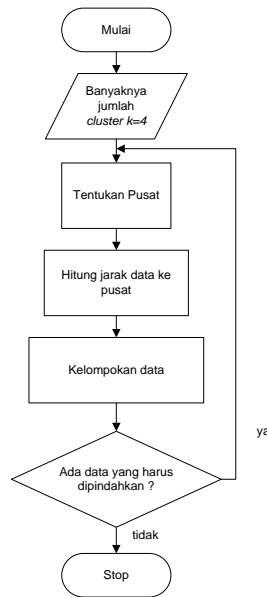
NO	NAMA	L/ P	Indeks Prestasi (IP)		Rata-rata Kehadiran (%)
			IP 1	IP 2	
D1	ADE ROZALINDA	P	3,20	3,16	98,44
D2	AFNI RANDA MUSTIANA	P	3,43	3,40	97,67
D3	AHLAKUL KARIMAH	P	3,40	3,43	97,71
D4	ALVINI LESTARI	P	3,40	3,48	100,00
D5	ANDINI LEONA SUHARDI	P	3,42	3,77	98,96
D6	AYU SITI HASANAH	P	2,84	3,44	99,34
D7	DAFID ARIANTO	L	3,03	3,86	99,69
D8	DESTI SOBRIANI	P	2,82	3,28	98,99
D9	DEVI SEPTIANITA	P	3,61	3,82	99,65
D10	DHEA DWI JAYANTIS	P	3,61	3,88	99,65
D11	FATMA NAULI BUTAR-BUTAR	P	3,40	3,50	100,00
D12	FEBRINA RAMADHANI	P	3,40	3,68	96,56
D13	FRYSCA PRIASTIWI	P	3,28	3,43	100,00
D14	HABIB ABDULLAH	L	2,76	2,78	98,33
D15	HATUN TARHAN	P	1,71	2,98	100,00
D16	MERI DWI CAHYA NINGRUM	P	1,91	2,14	92,67
D17	SYINTIA DEWI ANANTA SHINTA DEWI	P	3,57	3,76	99,06
D18	MIA SEPTIANI PUTRI	P	3,57	3,60	99,06
D19	MIFTA HURRAHMA	P	3,49	3,75	97,01
D20	MIFTAHUL BALAD	L	3,13	3,30	95,35

4.2 Perancangan dan Proses Clustering

Pada tahap ini akan dilakukan proses utama yaitu segmentasi atau pengelompokan data akademik mahasiswa FKIP Universitas Riau 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 FKIP Universitas Riau yang diperoleh, diambil 20 data untuk dijadikan sampel untuk penerapan algoritma k-means dalam pengelompokan hasil evaluasi mahasiswa. Percobaan dilakukan dengan menggunakan parameter-parameter berikut:

- a. Jumlah *cluster* : 4
- b. Jumlah data : 20
- c. Jumlah atribut : 3



Gambar 4.2 Flowchart Proses K-Means

Iterasi ke-0

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, di mana jumlah *centroid* awal dilakukan sebanyak empat *centroid* awal, nilai untuk C1 diambil dari baris data ke-3, nilai C2 diambil dari baris ke-6, nilai C3 diambil dari baris data ke-9, nilai C4 diambil dari baris data ke-16. Berikut ini nilai *centroid* awal pada penelitian:

- C1 = (3,40 ; 3,43 ; 97,71)
- C2 = (2,84 ; 3,44 ; 99,34)
- C3 = (3,61 ; 3,82 ; 99,65)
- C4 = (1,91 ; 2,14 ; 92,67)

2. Menghitung jarak dengan pusat cluster

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

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

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 0:

Perhitungan jarak data (objek) dengan *centroid* 1:

$$D1 = \sqrt{(3,40 - 3,20)^2 + (3,43 - 3,16)^2 + (97,71 - 98,44)^2} = \sqrt{0,64} = 0,80$$

$$D2 = \sqrt{(3,40 - 3,43)^2 + (3,43 - 3,40)^2 + (97,71 - 97,67)^2} = \sqrt{0,003} = 0,05$$

$$D3 = \sqrt{(3,40 - 3,40)^2 + (3,43 - 3,43)^2 + (97,71 - 97,71)^2} = \sqrt{0} = 0$$

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 0.

4. Lakukan iterasi, kemudian tentukan posisi *centroid* baru dengan menggunakan persamaan (1).

5. Ulangi langkah 3 jika posisi *centroid* baru tidak sama.

Pengecekan *konvergensi* dilakukan dengan membandingkan matriks *group assignment* pada iterasi sebelumnya dengan matrik *group assignment* pada iterasi yang sedang berjalan. Jika hasilnya sama maka *algoritma kmeans cluster analysis* sudah *konvergen*, tetapi jika berbeda maka belum *konvergen* sehingga perlu dilakukan iterasi berikutnya.

Tabel 4.15 Hasil Pengelompokan Data Iterasi 0 Sampai Iterasi 4

Iterasi ke-0				Iterasi ke-1				Iterasi ke-2				Iterasi ke-3				Iterasi ke-4						
c1	c2	c3	c4	c1	c2	c3	c4	c1	c2	c3	c4	c1	c2	c3	c4	c1	c2	c3	c4			
1	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	
1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0
1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0
0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
0	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
0	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
0	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0
0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0
1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0

Karena pada iterasi ke-4 posisi *cluster* tidak berubah / sama dengan posisi *cluster* pada iterasi ke-3, maka dapat disimpulkan bahawa proses iterasi dapat dihentikan pada iterasi ke-4 dengan hasil:

- Anggota *cluster* 1 (C1) : {D2, D3, D12, D19, D20} = 5 orang
- Anggota *cluster* 2 (C2) : {D1, D6, D8, D14, D15} = 5 orang
- Anggota *cluster* 3 (C3) : {D4, D5, D7, D9, D10, D11, D13, D17, D18} = 9 orang
- Anggota *cluster* 4 (C4) : {D16} = 1 orang

Berdasarkan hasil penelitian ini, kelompok mahasiswa berprestasi terdapat pada *cluster* 3 dan terdiri dari 9 anggota, untuk kelompok mahasiswa berpotensi berprestasi terdapat pada *cluster* 1 dan terdiri dari 5 anggota, untuk kelompok mahasiswa berpotensi bermasalah terdapat pada *cluster* 2 dan terdiri dari 5 anggota, dan untuk kelompok mahasiswa bermasalah terdapat pada *cluster* 4 dan terdiri dari 1 anggota.

4.3 Implementasi dan Pengujian

Pada bab ini merupakan tahapan tentang pembahasan metode menggunakan *software RapidMiner*. Pada tahap ini akan

digambarkan menggunakan *software RapidMiner* dan disamakan dengan pembuktian dari analisa metode terhadap permasalahan yang ada pada bab yang sebelumnya. Berdasarkan hasil implementasi dan pengujian data sampel sebanyak 20 record dengan menggunakan tanagra, menghasilkan *cluster* yang sama dengan perhitungan manual menggunakan algoritma K-Means. Selanjutnya dilakukan pengujian dengan menggunakan 94 data *record* yang menghasilkan *cluster* sebagai berikut :

Cluster	Anggota	Jumlah Anggota
Cluster 1	38, 47, 60, 63, 66, 78	6
Cluster 2	4, 5, 7, 8, 11, 13, 14, 15, 16, 18, 19, 20, 23, 26, 27, 29, 31, 34, 35, 36, 37, 39, 40, 41, 42, 45, 48, 49, 55, 61, 64, 65, 67, 70, 73, 74, 79, 80, 85, 90, 91, 92	43
Cluster 3	1, 10, 12, 17, 21, 22, 24, 25, 30, 32, 33, 44, 46, 52, 54, 57, 58, 62, 69, 75, 76, 77, 81, 82, 83, 86, 87, 88, 89, 93, 94	31
Cluster 4	2, 3, 6, 28, 43, 50, 51, 53, 56, 59, 68, 71, 72, 84	14

Berdasarkan tabel di atas dapat kita simpulkan bahwa dari pengelompokan data dapat diketahui kelompok nama mahasiswa FKIP Universitas Riau yang berprestasi terdapat pada *cluster* 3 dan mahasiswa berpotensi berprestasi terdapat pada *cluster* 1 sedangkan mahasiswa berpotensi bermasalah berada dalam *cluster* 2 dan mahasiswa bermasalah berada dalam *cluster* 4.

5. KESIMPULAN

Dari uraian yang telah ada pada bab – bab sebelumnya maka dapat ditarik kesimpulan sebagai berikut :

1. Metode *Clustering* Algoritma K-Means dapat diterapkan pada pengelompokan hasil evaluasi mahasiswa FKIP Universitas Riau, sehingga metode ini sangat membantu pihak akademik dalam menentukan mahasiswa berprestasi, berpotensi berprestasi, berpotensi bermasalah dan bermasalah.
2. Berdasarkan hasil perhitungan manual dan pengujian dengan *software RapidMiner* dengan menggunakan data akademik mahasiswa mendapatkan hasil yang sama.
3. Hasil pengelompokan data akademik mahasiswa dapat berfungsi sebagai acuan bagi perencanaan akademik untuk memantau dan mengevaluasi perkembangan kinerja akademik setiap mahasiswa.

REFERENSI

- Aprilla, D. et al., 2013. Belajar Data Mining dengan Rapid Miner. , pp.42–43.
- Arora, R.K. & Badal, D.D., 2013. Evaluating Student ' s Performance Using k-Means Clustering. *IJCST*, 4, pp.553–557.
- Bhoomi, B., 2014. Enhanced K-Means Clustering Algorithm to Reduce Time Complexity for Numeric Values. *Journal of Computer Science e and Information Technologies*, 5(1), pp.876–879.
- Ediyanto, Mara, M.N. & Satyahadewi, N., 2013. PENGKLASIFIKASIAN KARAKTERISTIK DENGAN METODE K-MEANS CLUSTER ANALYSIS. , 2(2), pp.133–136.
- Madhulatha, T.S., 2012. AN OVERVIEW ON CLUSTERING METHODS. *Journal of Engineering*, 2(4), pp.719–725.
- Oyelade, O.J., Oladipupo, O.O. & Obagbuwa, I.C., 2010. Application of k-Means Clustering algorithm for prediction of Students ' Academic Performance. (*IJCSIS*) *International Journal of Computer Science and Information Security*, 7, pp.292–295.
- 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.
- Sunjana, 2010. Aplikasi Mining Data Mahasiswa Dengan Metode Klasifikasi Decision Tree. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, pp.A24–A29.
- Suwirmayanti, P., Putra, I.K.G.D. & Kumara, I.N.S., 2014. OPTIMASI PUSAT CLUSTER K-PROTOTYPE DENGAN ALGORITMA GENETIKA. *Teknologi Elektro*, 13, pp.16–23.
- Varghese, B.M. et al., 2011. Clustering Student Data to Characterize Performance Patterns. *IJACSA*, pp.138–140.