

KLASIFIKASI DATA MAHASISWA MENGGUNAKAN METODE K-MEANS UNTUK MENUNJANG PEMILIHAN STRATEGI PEMASARAN

Totok Suprawoto

Sistem Informasi, STMIK AKAKOM Yogyakarta
e-mail : totok@akakom.ac.id, suprawoto.totok@gmail.com

ABSTRAK

Analisis cluster merupakan teknik data mining yang bertujuan untuk mengidentifikasi suatu kelompok dari objek yang memiliki karakteristik yang sama. Jumlah kelompok yang dapat diidentifikasi tergantung pada sejumlah data dan jenis dari objeknya. K-Means adalah salah satu metode clustering data yang dibagi kedalam bentuk satu atau lebih cluster/kelompok yang memiliki karakteristik sama. Clustering data mahasiswa menggunakan metode K-Means, terdiri dari nilai rerata ujian nasional (UN) dan indeks prestasi kumulatif (IPK) mahasiswa. Penelitian ini menggunakan data mahasiswa angkatan 2014/2015. Kemudian diperoleh kesimpulan bahwa kelompok mahasiswa dengan nilai rerata UN yang rendah memiliki pengaruh terhadap prestasi akademik mahasiswa yang rendah pada jenjang diploma-3(D-3) dan strata-1 (S-1). Jika mahasiswa memiliki nilai UN yang tinggi maka prestasi akademik mahasiswa juga tinggi pada semua jenjang. Dari hasil pengelompokan berdasarkan daerah asal sekolah IPK rata-rata yang tertinggi berasal dari propinsi Daerah Istimewa Yogyakarta (DIY) dan Jawa Tengah (Jateng).

Kata kunci: K-Means, Cluster, IPK, Nilai Rerata UN

ABSTRACT

Cluster analysis is data mining techniques that aim to identify a group of objects that have the same characteristics. The number of groups that can be identified depending on a number of data and the type of object. K-Means clustering is a method of data which is divided into the form of one or more clusters / groups that have the same characteristics. The use of K-Means method to perform clustering of students based on the average value of the UN and students GPA. Profile data sample of students of 2014/2015 showed that the group of students with a lower average value of the UN that has an influence on students' academic achievement is low (6.30, 1.87) and the levels of D-3 (6.45, 1.78) for S1, being the value of the UN the higher the academic achievement of students is also high (5.66, 3.12) for levels of D-3 and (8.19, 3.31) to S1.

Results from this study can be concluded that the higher the average value of the UN, the results of academic achievement as measured from it GPA's also higher. From the results based on region of school GPA highest come from Yogyakarta - Central Java.

Keywords: K-Means, Cluster, GPA, Average Value of UN

I. PENDAHULUAN

Perguruan tinggi sebagai institusi pendidikan telah memiliki data akademik dan administrasi dalam jumlah yang sangat besar, namun hanya sebagian kecil data tersebut dimanfaatkan (khususnya dalam penyusunan evaluasi diri). Data akademik mahasiswa merupakan data yang dihimpun dari hasil kegiatan proses belajar mengajar selama mengikuti studi di suatu perguruan tinggi. Data tersebut antara lain: data pribadi mahasiswa, data rencana studi, dan data hasil studi (nilai dan indeks prestasi). Disamping itu, bagian admisi pada setiap awal tahun akademik melakukan rekrutmen calon mahasiswa baru. Data mahasiswa baru yang dihimpun pada saat pendaftaran dan/atau registrasi berupa nilai ujian nasional (UN) rata-rata dan/atau nilai hasil tes masuk perguruan tinggi. Jumlah data yang terakumulasi dari tahun ke tahun perlu dilakukan analisis untuk dapat membuka peluang dihasilkannya informasi yang berguna dalam pembuatan alternatif keputusan bagi manajemen perguruan tinggi.

Untuk menarik minat calon mahasiswa yang lebih banyak, diperlukan strategi promosi yang tepat, antara lain: asal sekolah, prestasi akademik, lokasi tempat tinggal, tren keminatan calon mahasiswa, kemampuan calon mahasiswa (akademik dan/atau finansial), proses pembelajaran, serta prospek lulusan kedepannya.

Objek penelitian bersumber dari data mahasiswa yang telah melakukan registrasi di STMIK AKAKOM. Atribut data yang akan digunakan adalah nama mahasiswa, jurusan SLTA, nilai UAN, kota asal mahasiswa, program studi yang dipilih dan IPK. Hasil yang diperoleh dari penelitian ini diharapkan dapat menampilkan profil mahasiswa, keterkaitan antara nilai UN terhadap IPK mahasiswa, serta sebaran asal sekolahnya. Dari hasil analisis tersebut diharapkan dapat membantu pihak admisi perguruan tinggi dalam menyusun strategi promosi program studi yang menjadi target.

Penggalan informasi pada sebuah data yang berukuran besar (mempunyai jumlah *record* dan jumlah atribut yang cukup banyak) tidak dapat dilakukan dengan mudah. Teknologi *data mining* merupakan salah satu alat bantu untuk penggalan data pada basis data berukuran besar dengan spesifikasi kerumitan tinggi dan telah banyak

digunakan pada lingkungan aplikasi bisnis seperti perbankan, *provider* telekomunikasi, perusahaan pertambangan, perminyakan, dan lain-lain [1]. *Clustering* merupakan salah satu teknik *data mining* yang berfungsi melakukan pengelompokan sejumlah data atau objek ke dalam *cluster* (group) sehingga setiap *cluster* akan berisi data yang semirip mungkin dan berbeda dengan objek dalam *cluster* yang lainnya. Teknik klasifikasi merupakan pendekatan fungsi klasifikasi dalam data mining yang digunakan untuk melakukan prediksi atas informasi yang belum diketahui sebelumnya. Beberapa algoritma yang dapat digunakan antara lain adalah algoritma Decision Tree C.45, Artificial Neural Networks (ANN), K-Nearest Neighbor (KNN), algoritma Naive Bayes, Neural Network serta algoritma lainnya.

Klasterisasi merupakan teknik atau metode untuk mengelompokkan sejumlah besar data menjadi suatu bagian-bagian kecil data yang mempunyai atribut kemiripan dalam sifat, letak, ciri atau filter lain yang telah ditentukan. Teknik atau metode klasterisasi dapat dikelompokkan menjadi dua kategori besar, yaitu : (1) Metode Hirarki (Hierarchical Clustering) dan (2) Metode Non-Hirarki/Partisi (Partitional Hierarchical). Referensi [1] menunjukkan bahwa Algoritma K-Modes merupakan metode pengembangan dari K-Means yang mampu mengelompokkan data kategorikal dan menghasilkan klaster yang lebih stabil dengan waktu komputasi yang lebih singkat daripada metode K-Means.

Teknik klasifikasi merupakan pendekatan untuk menjalankan fungsi klasifikasi dalam data mining yaitu untuk menggolongkan data. Teknik klasifikasi ini dapat pula digunakan untuk melakukan prediksi atas informasi yang belum diketahui sebelumnya. Beberapa algoritma yang dapat digunakan antara lain adalah algoritma Decision Tree C.45, *Artificial Neural Networks* (ANN), K-Nearest Neighbor (KNN), algoritma Naive Bayes, Neural Network serta algoritma lainnya. Beberapa penelitian yang menggunakan teknik data mining pada data set akademik dan kemahasiswaan telah banyak dilakukan, antara lain [1] melakukan penelitian untuk mengklasifikasi kinerja akademik mahasiswa dengan menggunakan algoritma Supervised Learning In Quest (SLIQ) dan [2] dalam penelitian mengenai perancangan aplikasi K-Means untuk pengelompokan mahasiswa berdasarkan frekuensi kunjungan ke perpustakaan dan IPK.

Perbedaan dengan penelitian [2] adalah atribut data, profil yang dihasilkan, dan kegunaan dari hasil analisisnya. Sedangkan perbedaan dengan [3] terletak pada metode maupun profil yang dihasilkan, meskipun dari dua peneliti tersebut objeknya sama yaitu data mahasiswa. Penelitian ini difokuskan pada analisis hasil klasterisasi menggunakan metode K-Means untuk mengelompokkan sejumlah data dengan variabel: kota asal, nilai UAN dan IPK lulusan. Sumber data berasal dari bagian akademik dan admisi STMIK AKAKOM Yogyakarta.

II. METODE

Penelitian ini menggunakan teknik klasifikasi dengan K-Means untuk mengelompokkan data kategorikal hingga menghasilkan klaster yang lebih stabil. Selanjutnya melakukan analisis lebih mendalam terhadap hasil klasterisasi tersebut untuk membantu bagian admisi dalam menentukan kebijakan mengenai promosi program studi di lingkungan STMIK AKAKOM.

Analisis *cluster* merupakan salah satu teknik data mining yang bertujuan untuk mengidentifikasi sekelompok objek yang mempunyai kemiripan karakteristik tertentu yang dapat dipisahkan dengan kelompok objek lainnya, sehingga objek yang berada dalam kelompok yang sama relatif lebih homogen daripada objek yang berada pada kelompok yang berbeda. Jumlah kelompok yang dapat diidentifikasi tergantung pada banyak dan variasi data objek. Tujuan dari pengelompokan sekumpulan data objek ke dalam beberapa kelompok yang mempunyai karakteristik tertentu dan dapat dibedakan satu sama lainnya adalah untuk analisis dan interpretasi lebih lanjut sesuai dengan tujuan penelitian yang dilakukan. Model yang diambil diasumsikan bahwa data yang dapat digunakan adalah data yang berupa data interval, frekuensi dan biner. Set data objek harus mempunyai variabel dengan tipe yang sejenis tidak campur antara tipe yang satu dengan lainnya.

Algoritma K-Means dimulai dengan pemilihan secara acak K, yang merupakan banyaknya kluster yang ingin dibentuk. Kemudian ditetapkan nilai-nilai K secara random, untuk sementara nilai tersebut menjadi pusat dari kluster atau biasa disebut dengan *centroid/mean*. Selanjutnya dihitung jarak dari setiap data yang ada terhadap masing-masing centroid menggunakan rumus yang sudah disediakan hingga diketemukan jarak yang paling dekat dari setiap data dengan centroid. Klasifikasi setiap data berdasarkan kedekatannya dengan *centroid*.

A. Pengumpulan Data

Pengumpulan data dilakukan melalui penghimpunan data mahasiswa angkatan 2014/2015 yang bersumber dari bagian akademik dan admisi. Data dari bagian akademik meliputi: nomor induk mahasiswa (NIM), nama, dan IPK, sedang data dari admisi, meliputi: NIM, nama, asal sekolah, kota asal, dan nilai UN rata-rata. Selanjutnya kedua jenis data diintegrasikan kedalam satu file. Data dipisahkan berdasarkan jenjang studi yaitu D-3 dan S-1, selanjutnya hasil integrasi dilakukan cleaning data, untuk memperoleh data yang valid.

Metode K-Means digunakan untuk mengelompokkan data ke dalam beberapa kelompok, data dalam satu kelompok mempunyai karakteristik yang sama dan mempunyai karakteristik yang berbeda dengan data dari kelompok yang lain. Metode ini berusaha meminimalkan variasi antar data yang ada di dalam suatu cluster dan memaksimalkan variasi data yang ada di *cluster* lainnya.

B. Analisis

Untuk menentukan *cluster*, sebarang objek atau elemen pertama dapat dipilih untuk dijadikan sebagai titik tengah (*centroid point*) *cluster*. Algoritma metode K-Means selanjutnya akan melakukan pengulangan langkah-langkah berikut sampai terjadi kestabilan (tidak ada objek yang dapat dipindah-kan):

- 1) Menentukan koordinat titik tengah setiap *cluster*,
- 2) Menentukan jarak setiap objek terhadap koordinat titik tengah,
- 3) Mengelompokkan objek-objek tersebut berdasarkan pada jarak minimumnya.
- 4) Tentukan pusat *cluster* baru
- 5) Apakah ada selisih antara *cluster* lama dengan *cluster* baru?, Jika masih ada kembali ke langkah-1, jika tidak lanjut kelangkah 6.
- 6) Selesai.

C. Prosedur Pelaksanaan

Prosedur pelaksanaan yang dilakukan dalam penelitian ini meliputi:

- 1) Menghimpun data yang bersumber dari bagian akademik dan admisi untuk mahasiswa jenjang D3 dan S1, selanjutnya diintegrasikan untuk dilakukan *data cleaning*.
- 2) Menentukan jumlah *cluster* yang akan diproses lebih lanjut dan menetapkan titik pusat dari masing-masing *cluster*
- 3) Selanjutnya langkah-langkah yang dilakukan adalah menyusun data-data yang akan diolah termasuk parameter dan alternatif yang akan dirangking, merancang dan memproses perhitungan menggunakan metode K-Means untuk mendapatkan nilai pengelompokan mahasiswa berdasar-kan Nila Rata-rata UN dan IPK.
- 4) Hasil proses yang telah dinyatakan valid selanjutnya dilakukan analisis yang dikaitkan dengan kelompok dari asal kota sekolahnya.

III.HASIL

A. Analisis Data

Data sampel awal yang digunakan dalam penelitian ini, untuk D-3 sebanyak 66 data, sedang S-1 sebanyak 336 data (perhitungan dalam lampiran). Data mahasiswa awal (sebelum dilakukan perhitungan menggunakan metode K-Means) dapat dilihat pada Tabel I.

TABEL I
DAFTAR DATA AWAL

Mhs	NAMA MAHASISWA	NOMHS	ASAL SEKOLAH	KOTA SEKOLAH	RNIL	IPK
1	ERVIN YOGA PRADANA	143110001	SMK MUHAMMADIYAH 3	KOTA	7.50	2.89
2	MUHAMMAD MUKHBIT	143110002	SMK MUHAMMADIYAH 3	KOTA	8.45	3.06
3	MUHAMMAD CAHYO NOR	143110003	SMAN 1 TURI	KAB. SLEMAN	6.21	1.93
4	NANDA DWI SURYO PUTRO	143110004	SMK (STM) MUHAMMADIYAH	KAB. SLEMAN	6.33	1.52
5	SEPTIAN SARIP HIDAYAT	143110005	MAS AL MUNIROH	KAB.	6.88	1.85
6	SITI NUR AZIZAH	143110006	SMK (SMKK) KARYA RINI YHI	KAB. SLEMAN	7.65	2.59
7	MUHAMMAD ALFIAN	143110007	SMAN 1 KARANGMOJO	KAB. WONOSARI	6.85	2.43
8	DEVI RATNA WULANDARI	143110008	SMK MUHAMMADIYAH	KAB. WONOSARI	8.20	3.20
9	YUSUF SAIFULLAH	143110009	SMK (SMEA) MA`ARIF 1 TEMON	KAB. KULON PROGO	8.25	3.33
10	BAYU YOGA PRATAMA	143110011	SMA IKIP VETERAN NGEMPLAK	KAB. SLEMAN	7.10	2.16
11	ANANG NUGROHO	143110012	SMK (STM) MUHAMMADIYAH 1	KAB. KLATEN	7.55	2.59
12	ANNISA FAUZIAH	143110013	KEJAR PAKET C KAB. SLEMAN	SLEMAN	8.12	3.18
13	KORNELIA DELVIN	143110014	SMK YPK SERUI (SMEA)	YAPEN WAR	7.15	2.79
14	BENI WAHYU ADI	143110015	SMAN 1 SEDAYU	KAB. BANTUL	6.95	2.79
15	DESI WULANDARI	143110016	SMA MUHAMMADIYAH 4	KOTA	6.85	2.46
16	MUHAMMAD IWAN NUR	143110018	SMK (STM) MUHAMMADIYAH 1	KAB. WONOSARI	5.58	3.41
17	SIGIT GALIH	143110019	SMAN 1 PUNDONG	KAB. BANTUL	7.34	2.90
..						
..						
54	SUNARTI	143110028	MAN LAB UIN YOGYAKARTA	BANTUL	7.85	3.09
55	M. AGUNG JULIWANTO	143110029	SMA TUNAS BANGSA	KAB. INDRAGIRI	6.20	1.88
56	MUHAMAD MUSLIH	143110030	SMK DIPONEGORO	KAB. SLEMAN	7.18	2.73
57	INGGRIT SISWATI PUTRI	143110031	SMAN 1 MLATI	KAB. SLEMAN	6.45	2.14
58	BIMA NUR ADITAMA	143110032	SMA GADJAH MADA	KOTA	8.25	2.56
59	MUHAMMAD TAUFIQ	143110033	SMAN 1 TIDORE	KOTA TERNATE	7.75	2.34
66	FARIZ RACHMAT	143110034	SMA MUHAMMADIYAH 7	KOTA	7.90	2.83

Selanjutnya diproses menggunakan algoritma K-Means, dengan urutan langkah sebagai berikut.

- 1) Menentukan koordinat titik tengah setiap *cluster*,
- 2) Menentukan jarak setiap objek terhadap koordinat titik tengah,
- 3) Mengelompokkan objek-objek tersebut berdasarkan pada jarak minimumnya.
- 4) Tentukan pusat *cluster* baru
- 5) Apakah ada selisih antara *cluster* lama dengan *cluster* baru, Jika masih ada kembali ke langkah-1, jika tidak lanjut kelangkah 6.
- 6) Selesai.

B. Iterasi-1

- 1) Penentuan nilai awal titik tengah.

Langkah pertama, menentukan pusat *cluster* secara acak pada data awal yang ada di Tabel I. *Cluster* pertama yang akan digunakan dalam perhitungan adalah C_1 (6.0, 2.0), artinya Nilai Rerata UN pada pengelompokan pertama adalah 6.0 dan IPK mahasiswa sebesar 2.0. *Cluster* kedua C_2 (7.5, 3.0) dan *cluster* ketiga C_3 (8.0, 3.75).

- 2) Menghitung jarak objek ke *centroid* dengan menggunakan rumus jarak Euclid.

Perhitungan jarak mahasiswa pertama dengan pusat *cluster* pertama adalah:

$$C_{11} = \sqrt{(7.5 - 6.0)^2 + (2.89 - 2.0)^2} = 1.74$$

Perhitungan jarak mahasiswa pertama dengan pusat *cluster* kedua adalah:

$$C_{12} = \sqrt{(7.5 - 7.5)^2 + (2.89 - 3.0)^2} = 0.11$$

Perhitungan jarak mahasiswa pertama dengan pusat *cluster* ketiga adalah:

$$C_{13} = \sqrt{(7.5 - 8.0)^2 + (2.89 - 3.75)^2} = 0.99$$

Perhitungan jarak mahasiswa kedua dengan pusat *cluster* pertama adalah:

$$C_{11} = \sqrt{(8.45 - 6.0)^2 + (3.06 - 2.0)^2} = 2.67$$

Perhitungan jarak mahasiswa kedua dengan pusat *cluster* kedua adalah:

$$C_{12} = \sqrt{(8.45 - 7.5)^2 + (3.06 - 3.0)^2} = 0.95$$

Perhitungan jarak mahasiswa kedua dengan pusat *cluster* ketiga adalah:

$$C_{13} = \sqrt{(8.45 - 8.0)^2 + (3.06 - 3.75)^2} = 0.82$$

Perhitungan jarak mahasiswa ketiga dengan pusat *cluster* pertama adalah:

$$C_{11} = \sqrt{(6.21 - 6.0)^2 + (1.93 - 2.0)^2} = 0.22$$

Perhitungan jarak mahasiswa ketiga dengan pusat *cluster* kedua adalah:

$$C_{12} = \sqrt{(6.21 - 7.5)^2 + (1.93 - 3.0)^2} = 1.68$$

Perhitungan jarak mahasiswa ketiga dengan pusat *cluster* ketiga adalah:

$$C_{13} = \sqrt{(6.21 - 8.0)^2 + (1.93 - 3.75)^2} = 2.55$$

Perhitungan yang sama dilakukan untuk data ke empat hingga ke enam puluh enam. Hasil perhitungan dapat dilihat pada Tabel II.

TABEL II.
JARAK SETIAP MAHASISWA ITERASI KE-1

Mhs Ke	Cluster			Mhs Ke	Cluster			Mhs Ke	Cluster		
	C1	C2	C3		C1	C2	C3		C1	C2	C3
1	1.74	0.11	0.99	23	1.51	0.33	1.15	45	2.38	1.16	1.55
2	2.67	0.95	0.82	24	2.15	0.36	0.68	46	2.77	0.99	0.60
3	0.22	1.68	2.55	25	0.23	1.72	2.60	47	0.24	1.96	2.79
4	0.58	1.89	2.79	26	1.39	0.42	1.31	48	0.38	2.12	2.99
5	0.89	1.31	2.21	27	0.47	1.36	2.23	49	2.42	0.70	0.78
6	1.75	0.44	1.21	28	2.32	0.87	1.22	50	1.48	0.33	1.22
7	0.95	0.86	1.75	29	1.78	0.71	1.43	51	1.46	0.35	1.25
8	2.51	0.73	0.59	30	2.07	0.43	0.93	52	0.16	1.70	2.57
9	2.61	0.82	0.49	31	2.76	1.02	0.76	53	0.65	1.87	2.77
10	1.11	0.93	1.83	32	1.95	0.25	0.88	54	0.88	1.01	1.91
11	1.66	0.41	1.24	33	1.94	0.36	0.99	55	1.82	0.27	1.04
12	2.43	0.65	0.58	34	0.41	1.83	2.72	56	0.91	0.95	1.76
13	1.40	0.41	1.28	35	1.56	0.30	1.10	57	1.61	0.55	1.38
14	1.24	0.59	1.42	36	0.89	0.92	1.79	58	1.67	0.30	1.15
15	0.97	0.85	1.73	37	1.86	0.45	1.15	59	1.23	0.77	1.66
16	1.47	1.96	2.44	38	1.66	0.49	1.30	60	1.46	0.35	1.24
17	1.61	0.19	1.08	39	1.99	0.52	1.12	61	2.04	0.60	0.72
18	0.72	1.31	2.07	40	0.58	1.36	2.25	62	0.27	1.71	2.60
19	2.30	0.65	0.90	41	2.04	0.82	0.90	63	0.97	0.83	1.70
20	0.67	1.38	2.14	42	0.28	1.53	2.39	64	2.10	0.30	0.62
21	0.87	1.78	2.45	43	1.99	0.44	1.03	65	1.83	0.16	0.96
22	1.25	1.03	1.90	44	1.68	0.43	1.24	66	0.67	1.30	2.20

Melakukan *clustering* objek dengan memasukkan setiap objek ke dalam *cluster* (grup) berdasarkan jarak minimumnya. Suatu data akan menjadi anggota dari suatu *cluster* (C_1 , C_2 maupun C_3) yang memiliki jarak terkecil

dari pusat *cluster*-nya [5]. Misalnya untuk data pertama, jarak terkecil ada pada *cluster* yang ke-2 yaitu 0.11, berarti mahasiswa pertama akan menjadi anggota pada kelompok *cluster* yang ke-2. Demikian juga untuk data yang kedua, jarak terkecil ada pada *cluster* yang ke-3 yaitu 0.82, berarti mahasiswa kedua akan menjadi anggota pada kelompok *cluster* yang ke-3. Posisi *cluster* setiap mahasiswa dapat dilihat pada Tabel II.

Daftar data posisi *cluster* pada Tabel III menunjukkan hasil pengelompokan dari proses perhitungan dari Tabel II, sehingga diperoleh hasil pengelompokan baru untuk dilakukan iterasi ke-2, dengan jumlah anggota *cluster* C1 sebanyak 21, C2 sebanyak 39 dan C3 sebanyak 6, dari total anggota *cluster* sebanyak 66.

TABEL III.
POSISI CLUSTER SETIAP MAHASISWA ITERASI KE-1

Mhs ke	Kelompok Cluster						
1	C2	18	C1	35	C2	52	C1
2	C3	19	C2	36	C1	53	C1
3	C1	20	C1	37	C2	54	C1
4	C1	21	C1	38	C2	55	C2
5	C1	22	C2	39	C2	56	C1
6	C2	23	C2	40	C1	57	C2
7	C2	24	C2	41	C2	58	C2
8	C3	25	C1	42	C1	59	C2
9	C3	26	C2	43	C2	60	C2
10	C2	27	C1	44	C2	61	C2
11	C2	28	C2	45	C2	62	C1
12	C3	29	C2	46	C3	63	C2
13	C2	30	C2	47	C1	64	C2
14	C2	31	C3	48	C1	65	C2
15	C2	32	C2	49	C2	66	C1
16	C1	33	C2	50	C2		
17	C2	34	C1	51	C2		

3) Langkah berikutnya menghitung pusat *cluster* baru

Pusat *cluster* baru ditentukan berdasarkan pengelompokan anggota dari masing-masing *cluster*. Berdasarkan Tabel III.3, *cluster* pertama (C₁₁) dengan parameter Nilai Rerata UN memiliki 21 anggota. Pusat *cluster* pertama yang baru dihitung berdasarkan rata-rata koordinat dari 21 anggota tersebut adalah:

$$C_{11} = (6.21 + 6.33 + 6.88 + 5.58 + 6.23 + 6.17 + 5.73 + 6.20 + 6.45 + 6.25 + 6.75 + 6.58 + 6.25 + 5.77 + 5.88 + 6.15 + 6.40 + 6.85 + 6.60 + 6.23 + 6.67)/21 = 5.98$$

Cluster kedua (C₁₂) dengan parameter Nilai Rerata UN memiliki 39 anggota, sehingga pusat *cluster* kedua yang baru dihitung berdasarkan rata-rata koordinat dari 39 anggota tersebut, dengan total nilai 275.93 adalah:

$$C_{12} = 275.93/39 = 7.08$$

Cluster ketiga (C₁₃) dengan parameter Nilai Rerata UN memiliki 6 anggota, sehingga pusat *cluster* ketiga yang baru dihitung berdasarkan rata-rata koordinat dari 6 anggota tersebut, dengan total nilai 33.02 adalah:

$$C_{13} = 33.02/6 = 5.50$$

Pusat *cluster* baru untuk parameter IPK mahasiswa adalah:

$$C_{21} = 45.08/21 = 2.15$$

$$C_{22} = 106.40/39 = 2.73$$

$$C_{23} = 19.28/6 = 3.21$$

4) Ulangi iterasi yang dimulai dari langkah 2, sampai konvergen sehingga *cluster* yang baru memiliki nilai yang tetap (tidak mengalami perubahan).

IV. PEMBAHASAN

Pada penelitian sebelumnya yang dilakukan oleh Syafrianto (2005), pada iterasi ke-4 dihentikan karena hasil perhitungan menunjukkan adanya angka pusat *cluster* yang sama. Hasil klusterisasi didapatkan 3 kelompok data, yaitu: (1) Mahasiswa dengan IPK rendah dan jarang ke perpustakaan, dengan pusat *cluster* (1.48, 1.4); (2) Mahasiswa dengan IPK cukup tinggi dan cukup sering pergi ke perpustakaan, dengan pusat *cluster* (3.35, 1.64); dan (3) Mahasiswa dengan IPK tinggi dan sering pergi ke perpustakaan, dengan pusat *cluster*(3.49, 3.5).

Sedang pada penelitian ini, pengulangan dihentikan pada iterasi ke-3 dan ke-4, karena hasil perhitungan menunjukkan adanya angka pusat *cluster* yang yang sama. Dari hasil perhitungan kelompok data mahasiswa D3 didapatkan 3 kelompok *cluster* yaitu :

- 1) Kelompok pertama yang memiliki pusat *cluster* (6.30, 1.87) dimaksudkan bahwa kelompok pertama adalah mahasiswa dengan Nilai Rerata UN = 6.30 dan IPK = 1.87. Dari hasil pengelompokan mahasiswa yang IPKnya antara 1.5 – 2.0 nilai masuknya Rerata UN-nya antara 6.0 – 7.0.
- 2) *Cluster* yang kedua memiliki pusat *cluster* (6.70, 2.77) dimaksudkan bahwa kelompok kedua adalah mahasiswa dengan Nilai Rerata UN 6.70 dan IPK 2.77, untuk mahasiswa yang IPKnya antara 2.5 – 3.0 nilai masuk Rerata UN-nya antara 7.0 - 8.0.

- 3) *Cluster* yang ketiga memiliki pusat *cluster* (5.66, 3.12) dimaksudkan bahwa kelompok ketiga adalah mahasiswa dengan Nilai Rerata UN 5.66 dan IPK 3.12, bagi mahasiswa yang IPK-nya antara 3.0 – 3.5 nilai masuk Rerata UN ≥ 8.0 .

Hasil perhitungan kelompok data mahasiswa S1 didapatkan 3 *cluster*, yaitu:

- 1) *Cluster* yang pertama memiliki pusat *cluster* (6.45, 1.78) dimaksudkan bahwa kelompok pertama adalah mahasiswa dengan Nilai Rerata UN = 6.45 dan IPK = 1.87. Dari hasil pengelompokan mahasiswa yang nilai UN ≤ 6.50 dan prestasi akademik mahasiswa dengan IPK ≤ 1.5 .
- 2) *Cluster* yang kedua memiliki pusat *cluster* (7.30, 2.59) dimaksudkan bahwa kelompok kedua adalah mahasiswa dengan Nilai Rerata UN = 7.30 dan IPK = 2.59. Dari hasil pengelompokan mahasiswa yang nilai $6.50 < UN \leq 7.50$ dan prestasi akademik mahasiswa dengan $1.5 < IPK \leq 2.5$.
- 3) *Cluster* yang ketiga memiliki pusat *cluster* (8.19, 3.31) dimaksudkan bahwa kelompok ketiga adalah mahasiswa dengan Nilai Rerata UN = 8.19 dan IPK = 3.31. Dari hasil pengelompokan mahasiswa yang nilai UN > 7.50 dan prestasi akademik mahasiswa dengan IPK > 2.5 .

Selanjutnya hasil proses pengelompokan berdasarkan nilai tersebut dikelompokkan lagi berdasarkan asal kota sekolahnya, untuk mengetahui lebih jauh apakah ada relasi antara prestasi mahasiswa terhadap sebaran asal kota sekolahnya.

Dari hasil pengelompokan mahasiswa jenjang Diploma-3 berdasarkan kota asal sekolah yang dibagi dalam 3 area, yaitu: DIY – Jateng, Jatim – Jabar, Luar Jawa, diperoleh rerata IPK sebagai berikut:

- 1) Wilayah DIY – Jateng, rerata IPK = 2.74
- 2) Wilayah Jatim – Jabar, rerata IPK = 2.47
- 3) Wilayah Luar Jawa, rerata IPK = 2.41

Dari hasil pengelompokan mahasiswa jenjang S1 berdasarkan kota asal sekolah yang dibagi dalam 3 area, yaitu: DIY – Jateng, Jatim – Jabar, Luar Jawa, diperoleh rerata IPK sebagai berikut:

- 1) Wilayah DIY – Jateng, rerata IPK = 2.70
- 2) Wilayah Jatim – Jabar, rerata IPK = 2.64
- 3) Wilayah Luar Jawa, rerata IPK = 2.60

Profil sampel data mahasiswa tahun angkatan 2014/2015 tersebut dapat disimpulkan bahwa kelompok mahasiswa dengan nilai rerata UN yang rendah memiliki pengaruh terhadap prestasi akademik mahasiswa yang rendah (6.30, 1.87) jenjang D-3 dan (6.45, 1.78) untuk jenjang S1, sedang nilai UN yang semakin tinggi maka prestasi akademik mahasiswa juga tinggi (5.66, 3.12) untuk jenjang D-3 dan (8.19, 3.31) untuk jenjang S1.

V. SIMPULAN DAN SARAN

A. Simpulan

Berdasarkan penelitian yang dilakukan, dapat disimpulkan bahwa algoritma K-Means bisa digunakan untuk mengelompokkan mahasiswa berdasarkan nilai UN dan IPK. Dari data yang dilatih, didapatkan 3 kelompok.

- 1) Kelompok data mahasiswa D3, meliputi:
 - a. *Cluster* yang pertama memiliki pusat *cluster* (6.30, 1.87);
 - b. *Cluster* yang kedua memiliki pusat *cluster* (6.70, 2.77); dan
 - c. *Cluster* yang ketiga memiliki pusat *cluster* (5.66, 3.12).
- 2) Kelompok data mahasiswa S1, meliputi:
 - a. *Cluster* yang pertama memiliki pusat *cluster* (6.45, 1.78);
 - b. *Cluster* yang kedua memiliki pusat *cluster* (7.30, 2.59); dan
 - c. *Cluster* yang ketiga memiliki pusat *cluster* (8.19, 3.31).

Dari hasil pengelompokan mahasiswa jenjang Diploma-3 berdasarkan kota asal sekolah yang dibagi dalam 3 area, yaitu: DIY – Jateng, Jatim – Jabar, Luar Jawa, diperoleh rerata IPK sebagai berikut:

- 1) Wilayah DIY – Jateng, rerata IPK = 2.74
- 2) Wilayah Jatim – Jabar, rerata IPK = 2.47
- 3) Wilayah Luar Jawa, rerata IPK = 2.41

Dari hasil pengelompokan mahasiswa jenjang S1 berdasarkan kota asal sekolah yang dibagi dalam 3 area, yaitu: DIY – Jateng, Jatim – Jabar, Luar Jawa, diperoleh rerata IPK sebagai berikut:

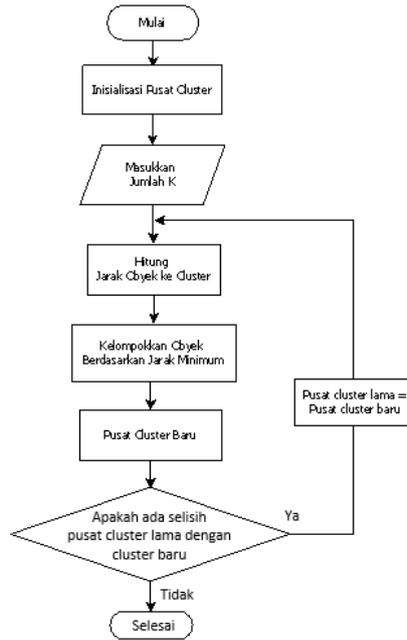
- 1) Wilayah DIY – Jateng, rerata IPK = 2.70
- 2) Wilayah Jatim – Jabar, rerata IPK = 2.64
- 3) Wilayah Luar Jawa, rerata IPK = 2.60

B. Saran

Untuk mengembangkan lebih jauh dari penelitian ini, disarankan untuk mengembangkan lebih lanjut adanya pemetaan kesesuaian antara prestasi akademik mahasiswa dikaitkan dengan asal jurusan dan prestasi di SMA/SMK.

REFERENSI

- [1] Arief Jananto,, “Memprediksi kinerja mahasiswa menggunakan teknik data mining (studi kasus data akademik mahasiswa unisbank””, Tesis Tidak Terpublikasi, Yogyakarta: Universitas Gajah Mada, 2010.
- [2] Andina Budiarti, Yudho Giri Suchyo, dan Yova Ruldeviyani, *Studi karakteristik kelulusan peserta didik dengan teknik clustering*”, Seminar Nasional Sistem dan Informatika, Bali, 2006.
- [3] Andi Syafrianto, “Perancangan aplikasi K-means untuk pengelompokan mahasiswa STMIK EL-ERHAMA Yogyakarta berdasarkan frekuensi kunjungan ke perpustakaan dan IPK””, Yogyakarta: STMIK EL-ERHAMA, 2005.
- [4] Efraim T., *Decision support systems and intelligent systems*, Edisi ke-7, Andi, Yogyakarta, 2005.
- [5] Pramudyo, A. S., Case base reasoning untuk klasifikasi mahasiswa baru berdasarkan prediksi Indeks Prestasi semester I (studi kasus Program Studi Teknik Informatika Universitas Bina Darma Palembang), Tesis Tidak Terpublikasi, Yogyakarta: Universitas Gajah Mada, 2008.
- [6] Heribertus Giyanto, , *Penerapan algoritma clustering K-Means, K-Medoid, Gath Geva*, Tesis Tidak Terpublikasi, Yogyakarta: Universitas Gajah Mada, 2008.
- [7] Jiawei Han and Micheline Kamber, *Data mining concepts and techniques second edition*, San Francisco: Morgan Kauffman, 2001.
- [8] Oyelade, Oladipupo, Obagbuwa, “ Application of K-Menas clustering algorithm for prediction of students acaddecic performance””, *International Journal of Computer Science and Information Security*, Volume 7, 2010.



GAMBAR 5. FLOWCHART METODE K-MENAS