

Penerapan Metode KMeans dan Cobweb Terhadap Analisis Prestasi Akademik Mahasiswa yang Mengikuti Kegiatan Kemahasiswaan

Neil Casaandra Sudharmono¹, Mewati Ayub²

Fakultas Teknologi Informasi, Universitas Kristen Maranatha
Jl. Suria Sumantri 65, Bandung

¹neilsudharmono@gmail.com

²mewati.ayub@itmaranatha.org

Abstract— The growth of an university can be seen with student's academic records which can be used for determining students' quality. Besides student's academic records, other component that can be used to determine students' quality is student's participation in non-academic activities which most of universities have to develop their student's softskill during their study period. Based on these data provision, we can analyze how non-academic activities affect student's academic achievement. In this research, the data that used for analysis are students' academic data and students' activities data from Maranatha Christian University as case study. This analysis use one of data mining technique, which is clustering. The clustering method that used for this research are Kmeans and Cobweb. Clustering method groups students and alumni that participate in student activities or not by looking and the student's and alumni's faculty, major, years of study, and if the student or alumni participate in student activities, we can analyze it with how many activities that the student and alumni had.

Keywords— clustering, KMeans, Cobweb, students' quality, university

I. PENDAHULUAN

Tingginya kualitas diri dari setiap pribadi manusia diukur dari berbagai bidang, Salah satu tolak ukurnya adalah dalam bidang pendidikan. Pendidikan di Indonesia khususnya pada perguruan tinggi merupakan suatu wadah yang menempa setiap mahasiswanya agar memiliki nilai-nilai yang baik dalam dirinya. Mahasiswa yang memiliki nilai yang tinggi adalah cerminan keberhasilan sistem manajemen pendidikan dari perguruan tinggi.

Menurut BAN-PT, kualitas perguruan tinggi di Indonesia diukur berdasarkan akreditasi pada setiap program studinya. Terdapat 7 standar akreditasi BAN-PT yang salah satu diantaranya yaitu mahasiswa dan lulusan dimana lama masa studi dan indeks penilaian kumulatif (IPK) menjadi salah satu variabel penilaian standar akreditasi butir tiga [1].

Setiap fakultas dari Universitas Kristen Maranatha (UKM) sudah memiliki pembinaan mahasiswa, baik itu dalam *hardskill* maupun *softskill*. Selain dalam bidang akademik, UKM juga mempunyai wadah lain untuk meningkatkan nilai dari setiap mahasiswanya melalui unit kegiatan. Beban studi yang diambil setiap semesternya, unit kegiatan yang diikuti, IPK yang diraih dan aktivitas akademik dapat mempengaruhi dan menguji kualitas dari mahasiswa yang berdampak pada ketepatan lulus mahasiswa, dimana standar ketepatan waktu lulus mahasiswa di UKM untuk program sarjana adalah empat tahun.

Kualitas diri dari mahasiswa mencerminkan kualitas pembinaan perguruan tinggi tersebut. Berdasarkan permasalahan tersebut, maka diperlukan sebuah aplikasi yang dapat menganalisis data akademik mahasiswa dikaitkan dengan kegiatan kemahasiswaan. Analisis ini ditinjau dari histori kegiatan akademik mahasiswa aktif dan lulusan UKM menggunakan metode *data mining*. Aplikasi ini juga menggunakan aplikasi Weka sebagai alat bantu proses *data mining* untuk membantu dalam proses analisis.

Oleh karena itu tujuan dari penelitian ini adalah sebagai berikut:

1. Menggunakan *data mining* dalam menganalisis karakteristik akademik mahasiswa baik mahasiswa aktif maupun lulusan yang mengikuti kegiatan kemahasiswaan maupun yang tidak mengikuti kegiatan kemahasiswaan.
2. Menggunakan metode *clustering* dalam pembuatan aplikasi analisis karakteristik akademik mahasiswa baik mahasiswa aktif maupun lulusan yang mengikuti kegiatan kemahasiswaan ataupun yang tidak mengikuti kegiatan kemahasiswaan.

II. LANDASAN TEORI

A. Data Mining

Data mining merupakan suatu metode yang digunakan untuk mendapatkan suatu pengetahuan dari sejumlah data berukuran besar yang dapat digunakan untuk suatu

kebutuhan tertentu. *Data mining* merupakan salah satu proses dalam *Knowledge Discovery from Data* (KDD). KDD terdiri dari proses sebagai berikut [2][3]:

1. *Data Cleaning* (membersihkan data yang tidak valid/ lengkap).
2. *Data Integration* (proses integrasi data).
3. *Data Selection* (proses pemilihan data yang layak untuk dianalisis).
4. *Data Transformation* (proses transformasi data supaya lebih mudah dalam melakukan *data mining*).
5. *Data Mining* (proses penggalian informasi dengan algoritma-algoritma yang menunjang sesuai kebutuhan).
6. *Pattern Evaluation* (proses evaluasi hasil penggalian informasi).
7. *Knowledge Presentation* (proses visualisasi dan penyampaian dari hasil penggalian informasi).

B. Clustering

Clustering merupakan salah satu metode analisis *data mining* yang dapat digunakan untuk berbagai bidang salah satunya adalah penggalian pengetahuan baru dalam penyebaran dokumen yang sangat masif baik dalam *web* maupun sistem privat [4][5]. Proses *clustering* mengelompokkan sejumlah data yang dianalisis sesuai dengan kesamaan karakteristik setiap datanya. Data dari satu *cluster* dengan *cluster* lain mempunyai karakteristik yang berbeda. Karakteristik setiap *cluster* yang dianalisis tidak ditentukan sebelumnya. Proses *clustering* harus dilakukan oleh pihak yang memahami karakteristik domain data yang dianalisis.

Proses *clustering* juga dapat digunakan dalam pra pemrosesan data sebelum data diolah dengan menggunakan teknik *data mining* lainnya untuk meningkatkan pemahaman karakteristik domain data. Terdapat empat metode *clustering* yaitu *partitioning method*, *hierarchical method*, *density-based method* dan *grid-based method* [2][4][5].

Pemilihan atribut dalam data set yang akan digunakan dalam clustering, penting dilakukan untuk mendapatkan hasil *clustering* yang lebih baik [6][7].

Metode yang dipakai dalam analisis karakteristik akademik mahasiswa baik mahasiswa aktif maupun lulusan yang mengikuti kegiatan kemahasiswaan ataupun yang tidak mengikuti kegiatan di Universitas Kristen Maranatha yaitu *Partitional Clustering* (*KMeans*) dan *Hierarchical Clustering* (*Cobweb*).

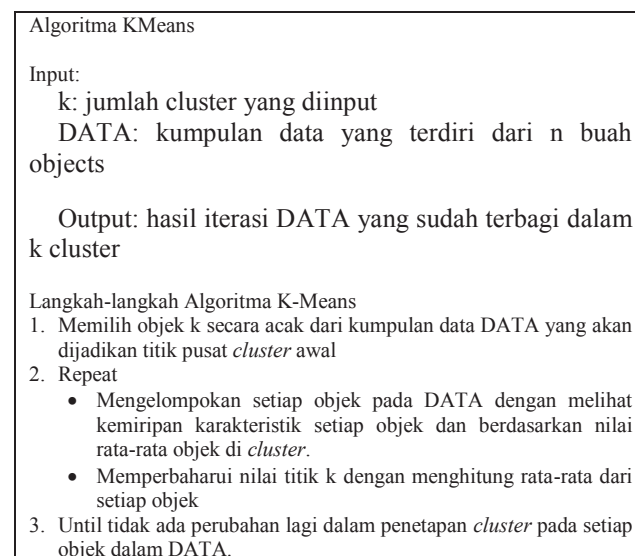
Penelitian ini dilakukan dikarenakan dunia pendidikan saat ini memiliki data mengenai mahasiswa dan lulusan yang dihasilkannya. Karena ketersediaan data yang berlimpah, maka penerapan *data mining* dapat dilakukan untuk pembinaan mahasiswa yang lebih baik.

C. Partitional Clustering

KMeans merupakan salah satu metode *partitional clustering* yang bekerja dengan cara mengelompokkan data-

data yang dianalisis secara iteratif ke dalam salah satu cluster diantara k buah cluster. *KMeans* menggunakan rumus *Euclidean Distance* dalam menghitung jarak antara data dan pusat data (*centroid*). *KMeans* adalah algoritma yang menyelesaikan proses *clustering* dengan cepat. Namun, pemilihan *seeds* yang berbeda menghasilkan jumlah *cluster* yang berbeda [8].

Metode clustering *KMeans* pada Weka diterapkan menggunakan algoritma *SimpleKMeans*. Parameter *SimpleKMeans* yang digunakan adalah dengan menentukan jumlah *cluster* (*numCluster*) dan *user* dapat menentukan iterasi maksimal (*max iteration*) yang akan dilakukan. Algoritma *K-means* dijelaskan pada Gambar 1.



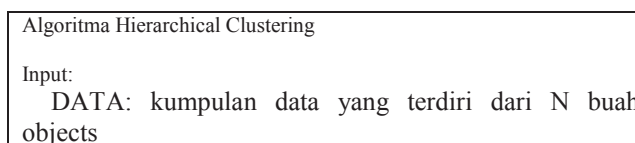
Gambar 1 Algoritma KMeans [2]

D. Hierarchical Clustering

Metode hirarki (*hierarchical method*) mengelompokkan sejumlah data objek menjadi sebuah hirarki atau cluster dalam bentuk *tree*. Metode hirarki ini sangat berguna untuk penyimpulan dan visualisasi data, metode ini ada dua jenis yaitu *Agglomerative* dan *Divisive* [2].

Metode *Agglomerative Hierarchical Clustering* pada Weka diterapkan dalam algoritma *Cobweb* [9][10]. *Cobweb* menerapkan algoritma *Cobweb* untuk atribut nominal dan algoritma *classit* untuk atribut numerik.

Parameter yang digunakan pada *Cobweb* adalah *acuity* dan *cutoff*. *Acuity* digunakan untuk menentukan standar deviasi minimum pada atribut numerik. *Cutoff* digunakan untuk memangkas node yang terbuat untuk menekan banyaknya *node* yang terbentuk dalam *tree* [9]. Algoritma *hierarchical clustering* akan dijelaskan pada Gambar 2.



Output: hasil iterasi DATA yang sudah terbagi dalam cluster

Langkah-langkah Hierarchical Clustering

3. Setiap *objects* menjadi satu *cluster*. *Cluster* yang terbentuk terdapat N buah *cluster*
4. Repeat
 - Cari jarak *cluster* terdekat/ paling mirip untuk setiap *cluster* dengan *cluster* lainnya
 - Gabungkan *cluster* terdekat
5. Until semua *cluster* telah bergabung menjadi satu *cluster*.

Gambar 2 Algoritma Cobweb

E. Penelitian mengenai clustering

Penelitian Liu dan George [11] menggunakan metode *clustering* KMeans berbasis fuzzy untuk menganalisis data cuaca (*spatio temporal*) di Amerika Serikat.

Al-Shboul dan Myaeng [8] melakukan penelitian untuk menentukan inisialisasi data dalam KMeans dengan menggunakan algoritma genetik.

Penelitian yang dilakukan Marghny, Abd El-Aziz, dan Taloba [12] mengevaluasi dataset hepatitis C menggunakan metode KMeans.

Penelitian dalam makalah ini mengeksplorasi dua metode, yaitu KMeans dan Cobweb dalam menganalisis prestasi akademik mahasiswa dalam dua kelompok, yaitu kelompok yang mengikuti kegiatan kemahasiswaan dan kelompok yang tidak mengikuti kegiatan kemahasiswaan.

III. METODE PENELITIAN

Penelitian ini dilakukan dengan langkah-langkah sebagai berikut.

A. Data Preparation

Terdapat empat data set yang akan digunakan sebagai bahan analisis, yaitu data set mahasiswa aktif yang mengikuti kegiatan kemahasiswaan, data set lulusan yang mengikuti kegiatan kemahasiswaan, data set mahasiswa aktif yang tidak mengikuti kegiatan kemahasiswaan, dan data set lulusan yang tidak mengikuti kegiatan kemahasiswaan. Data yang diperoleh merupakan data mentah, baik dalam bentuk *softcopy* maupun *hardcopy*.

B. Data Preprocessing

Preprocessing terhadap data yang dikumpulkan meliputi proses ekstraksi data, pembersihan data, dan transformasi data untuk mendapatkan data yang layak untuk dilakukan proses *clustering*.

Pada Tabel I, data mahasiswa aktif dan lulusan dengan beberapa atribut telah dipilih sebagai data set mahasiswa aktif yang mengikuti kegiatan di seluruh fakultas yang akan digunakan dalam proses *clustering*. Data mahasiswa yang digunakan adalah data fakultas yang terdiri dari Fakultas E, P, T, I, R, K, H, S. Unit yang diikuti dikelompokkan menjadi Unit Kegiatan Sosial dan Organisasi, Unit Bela Diri, Unit Olahraga, dan Unit Seni. Lama studi dengan nilai

diantara 1 tahun sampai 7 tahun. IPK dikelompokkan menjadi status IPK dengan nilai Sangat Memuaskan, Memuaskan, Kurang Memuaskan dan Dengan Pujian. Sedangkan keaktifan dikelompokkan menjadi aktif dan pasif.

TABEL I
DATA SET MAHASISWA AKTIF DAN LULUSAN YANG MENGIKUTI KEGIATAN KEMAHASISWAAN

Atribut	Deskripsi Atribut	Value	Jumlah mahasiswa aktif	Jumlah lulusan
Fakultas	Fakultas Mahasiswa	Fakultas E	210	67
		Fakultas P	139	21
		Fakultas T	73	29
		Fakultas I	56	17
		Fakultas R	59	19
		Fakultas K	17	8
		Fakultas H	23	0
Unit	Unit Kegiatan Mahasiswa	Unit Kegiatan Sosial dan Organisasi	13	0
		Unit Bela Diri	97	41
		Unit Olahraga	76	39
		Unit Seni	454	96
Status IPK	Status IPK Mahasiswa	Sangat Memuaskan (SM)	300	111
		Dengan Pujian (DP)	97	49
		Memuaskan (M)	181	14
		Kurang Memuaskan (KM)	62	2
Lama Studi	Lama Studi	Minimum	1	1
		Maximum	7	7
Keaktifan	Keaktifan Mahasiswa dalam Kegiatan	Aktif	445	120
		Pasif	195	56

Pada Tabel II beberapa atribut telah dipilih sebagai data set mahasiswa aktif dan data set lulusan yang tidak mengikuti kegiatan di Fakultas I yang akan digunakan dalam proses *clustering*. Atribut yang digunakan adalah data program studi, data IPK yang dikelompokkan menjadi status ipk, dan lama studi dalam tahun.

TABEL II
DATA SET MAHASISWA AKTIF DAN LULUSAN FAKULTAS I YANG TIDAK MENGIKUTI KEGIATAN

Atribut	Deskripsi Atribut	Value	Jumlah mahasiswa aktif	Jumlah lulusan
Prodi	Prodi Mahasiswa	Prodi F	1147	283
		Prodi S	432	108
Status IPK	Status IPK	Sangat	395	163

	Mahasiswa	Memuaskan (SM)		
		Dengan Pujian (DP)	125	218
		Memuaskan (M)	432	9
		Kurang Memuaskan (KM)	626	0
Lama Studi	Lama Studi	Minimum	1	1
		Maximum	7	7

Fakultas	E	E	P	E
Unit	Unit Seni	Unit Seni	Unit Seni	Unit Olahraga
Status IPK	SM	SM	SM	SM
Lama Studi	3,11	2,68	3,08	3,94
Keaktifan	Aktif	Pasif	Aktif	Aktif
Instances	100 %	35%	46%	19%

C. Proses Clustering dengan KMeans dan Cobweb

Setelah data set melalui proses *preprocessing*, proses *clustering* akan dilakukan dengan metode KMeans dan Cobweb dengan menggunakan *library* Weka pada aplikasi yang telah dikembangkan. Proses *clustering* pada setiap metode akan dibagi menjadi dua kelompok yaitu kelompok yang mengikuti kegiatan kemahasiswaan yang meliputi semua fakultas di Universitas Kristen Maranatha, serta kelompok yang tidak mengikuti kegiatan yang berasal dari Fakultas I di Universitas Kristen Maranatha. Data set pada setiap kelompok dibagi menjadi data set mahasiswa aktif dan lulusan.

IV. HASIL PENELITIAN

A. Metode KMeans

Berikut ini adalah hasil *clustering* dengan metode KMeans dari mahasiswa aktif dan lulusan yang mengikuti kegiatan kemahasiswaan yang meliputi seluruh fakultas, dan yang tidak mengikuti kegiatan meliputi Fakultas I.

1) *Mahasiswa Aktif yang Mengikuti Kegiatan Kemahasiswaan Seluruh Fakultas dengan 3 Cluster*: Hasil analisis Pada Tabel III dengan jumlah tiga cluster menunjukkan bahwa pada *cluster 1*, atribut Fakultas yang mendominasi adalah Fakultas E, atribut unit yang mendominasi adalah Unit Seni, rata-rata mahasiswanya berada di tahun ke 2 dan sebagian besar memiliki status IPK Sangat Memuaskan serta aktif mengikuti kegiatan di dalam unit. Pada *cluster 2*, atribut Fakultas yang mendominasi adalah Fakultas P, atribut unit yang mendominasi adalah Unit Seni, sebagian besar mahasiswa berada di tahun ke 3, memiliki status IPK Sangat Memuaskan, dan tidak terlalu aktif mengikuti kegiatan unit.

Pada *cluster 3*, atribut Fakultas yang mendominasi adalah Fakultas E, atribut unit yang mendominasi adalah Unit Olahraga, sebagian besar mahasiswa memiliki status IPK sangat memuaskan, aktif dalam unit.

TABEL III
HASIL CLUSTERING KMEANS MAHASISWA YANG MENGIKUTI KEGIATAN KEMAHASISWAAN DENGAN 3 CLUSTER

Atribut	Cluster			
	Seluruh Data	1	2	3

2) *Mahasiswa Aktif Fakultas I yang Tidak Mengikuti Kegiatan Kemahasiswaan dengan 3 Cluster*: Hasil analisis pada Tabel IV dengan jumlah 3 *cluster* menunjukkan bahwa pada *cluster 1*, atribut prodi yang mendominasi adalah prodi F, sebagian besar mahasiswa pada *cluster 1* memiliki status IPK sangat memuaskan dan sudah menempuh studi rata-rata 4,5 – 5 tahun.

Pada *cluster 2*, atribut prodi yang mendominasi adalah prodi F, sebagian besar mahasiswa memiliki status IPK memuaskan. Lama studi yang sudah ditempuh rata-rata 3,5 tahun. Pada *cluster 3*, atribut prodi yang mendominasi adalah prodi F, sebagian besar mahasiswa pada *cluster 3* memiliki status IPK Memuaskan. Lama Studi yang ditempuh rata-rata lebih dari 6,5 tahun.

TABEL IV
HASIL CLUSTERING KMEANS MAHASISWA AKTIF FAKULTAS I YANG TIDAK MENGIKUTI KEGIATAN KEMAHASISWAAN DENGAN 3 CLUSTER

Atribut	Cluster			
	Seluruh Data	1	2	3
Prodi	F	F	F	F
Status IPK	M	SM	M	M
Lama Studi	4,75	4,79	3,41	6,63
Instances	100%	43%	31%	26%

3) *Mahasiswa Aktif yang Mengikuti Kegiatan Kemahasiswaan Seluruh Fakultas dengan 5 Cluster*: Hasil analisis pada Tabel V dengan jumlah 5 *cluster* menunjukkan bahwa pada *cluster 1*, atribut Fakultas yang mendominasi adalah Fakultas T, atribut unit yang mendominasi adalah Unit Seni. Keaktifan mahasiswa tergolong Aktif dan *cluster 1* memiliki status IPK Kurang memuaskan.

Pada *cluster 2*, atribut Fakultas yang mendominasi adalah Fakultas E, atribut unit yang mendominasi adalah Unit Olahraga, sebagian besar mahasiswa pada *cluster 2* memiliki status IPK Memuaskan.

Pada *cluster 3*, atribut Fakultas yang mendominasi adalah Fakultas E, atribut unit yang mendominasi adalah Unit Olahraga, sebagian besar mahasiswa pada *cluster 3* memiliki status IPK sangat memuaskan. Keaktifan dalam unit termasuk Aktif.

Pada *cluster 4*, atribut Fakultas yang mendominasi adalah Fakultas E, atribut unit yang mendominasi adalah

Unit Seni, sebagian besar mahasiswa pada *cluster* 4 memiliki status IPK Memuaskan. Keaktifan mahasiswa dalam unit kegiatan menunjukkan bahwa mahasiswa yang berada di *cluster* 4 pasif jika dibandingkan dengan *cluster* 1, 2, dan 3.

Pada *cluster* 5, atribut Fakultas yang mendominasi adalah Fakultas E, atribut unit yang mendominasi adalah Unit Seni, sebagian besar mahasiswa pada *cluster* 5 memiliki status IPK sangat memuaskan. Keaktifan mahasiswa dalam unit kegiatan menunjukkan bahwa mahasiswa yang berada di *cluster* 5 cukup pasif dalam unit kegiatan.

Perbedaan hasil analisis pada Tabel V tersebut dengan jumlah 5 *cluster* jika dibandingkan dengan hasil analisis 3 *cluster* pada Tabel IV adalah penambahan jumlah *cluster* memunculkan 1 *cluster* dengan jumlah 17% dari seluruh mahasiswa aktif yang sebagian besar mahasiswanya memiliki Status IPK Kurang memuaskan dan mahasiswa yang berasal dari Fakultas T yang mendominasi pada *cluster* tersebut.

Walaupun muncul satu *cluster* dengan hasil tersebut, sebanyak 83% dari mahasiswa aktif yang mengikuti kegiatan masih dapat memiliki IPK yang memuaskan bahkan 55% mahasiswa memiliki Status IPK Sangat Memuaskan, kecuali untuk kelompok mahasiswa tertentu.

TABEL V
HASIL CLUSTERING KMEANS MAHASISWA YANG MENGIKUTI KEGIATAN KEMAHASISWAAN DENGAN 5 CLUSTER

Atribut	Cluster					
	Seluruh Data	1	2	3	4	5
Fakultas	E	T	E	E	E	E
Unit	Unit Seni	Unit Seni	Unit Olahraga	Unit Olahraga	Unit Seni	Unit Seni
Status IPK	SM	KM	M	SM	M	SM
Lama Studi	3,11	2,97	2,63	3,49	3,12	3,72
Keaktifan	Aktif	Aktif	Aktif	Aktif	Pasif	Pasif
Jumlah (%)	100%	17%	15%	35%	13%	20%

4) *Mahasiswa Aktif Fakultas I yang Tidak Mengikuti Kegiatan Kemahasiswaan dengan 5 Cluster*: Hasil analisis pada Tabel VI dengan jumlah 5 *cluster* menunjukkan hal yang menarik dengan munculnya dua *cluster* dimana sebanyak 34% dari seluruh mahasiswa aktif Fakultas I yang berada di antara tahun kedua dan ketiga mendapatkan IPK Memuaskan. Bila dibandingkan dengan mahasiswa aktif Fakultas I yang mengikuti kegiatan kemahasiswaan, prestasi akademik yang diraih lebih baik daripada mahasiswa yang tidak mengikuti kegiatan.

TABEL VI
HASIL CLUSTERING KMEANS MAHASISWA FAKULTAS I YANG TIDAK MENGIKUTI KEGIATAN KEMAHASISWAAN DENGAN 5 CLUSTER

Atribut	Cluster					
	Seluruh Data	1	2	3	4	5
Prodi	F	F	F	F	S	F
Status IPK	M	M	SM	M	SM	M
Lama Studi	4,75	6,10	4,67	2,87	4,6	3,22
Instances	100%	11%	37%	20%	18%	14%

5) *Lulusan yang Mengikuti Kegiatan Kemahasiswaan Seluruh Fakultas dengan 3 Cluster*: Hasil analisis pada Tabel VII dengan jumlah 3 *cluster* menunjukkan bahwa pada *cluster* 1, atribut Fakultas yang mendominasi adalah Fakultas E, atribut unit yang mendominasi adalah Unit Seni. Sebagian besar lulusan pada *cluster* 1 memiliki status IPK sangat memuaskan dan aktif dalam unit kegiatan selama masa studinya. Rata-rata masa studi lulusan pada *cluster* 1 adalah 4,5 – 5 tahun.

Pada *cluster* 2, atribut Fakultas yang mendominasi adalah Fakultas E, atribut unit yang mendominasi adalah Unit Olahraga, sebagian besar lulusan pada *cluster* 2 memiliki status IPK sangat memuaskan dan tidak terlalu aktif dalam unit kegiatan selama masa studinya. Lulusan pada *cluster* 2 rata-rata menempuh masa studi selama 4 – 4,5 tahun.

Pada *cluster* 3, atribut Fakultas yang mendominasi adalah Fakultas S, atribut unit yang mendominasi adalah unit bela diri, sebagian besar lulusan pada *cluster* 3 memiliki status IPK sangat memuaskan dan aktif dalam unit kegiatan. Rata-rata masa studi yang ditempuh oleh lulusan adalah 5 tahun.

TABEL VII
HASIL CLUSTERING KMEANS LULUSAN YANG MENGIKUTI KEGIATAN KEMAHASISWAAN DENGAN 3 CLUSTER

Atribut	Cluster			
	Seluruh Data	1	2	3
Fakultas	E	E	E	S
Unit	Unit Seni	Unit Seni	Unit Olahraga	Unit Bela Diri
Status IPK	SM	SM	SM	SM
Lama Studi	4,70	4,82	4,33	4,99
Keaktifan	Aktif	Aktif	Pasif	Aktif
Instances	100%	51%	27%	22%

6) *Lulusan Fakultas I yang Tidak Mengikuti Kegiatan Kemahasiswaan dengan 3 cluster*: Hasil analisis pada Tabel VIII dengan jumlah 3 *cluster* menunjukkan bahwa pada *cluster* 1, atribut prodi yang mendominasi adalah prodi F, sebagian besar lulusan pada *cluster* 1 memiliki status IPK sangat memuaskan. Lama studi yang ditempuh rata-rata 4,5 – 5 tahun.

Pada *cluster 2*, atribut prodi yang mendominasi adalah prodi F, sebagian besar lulusan pada *cluster 2* memiliki status IPK Dengan Pujian. Lama studi yang sudah ditempuh kira-kira 4,08 tahun. Pada *cluster 3*, atribut prodi yang mendominasi adalah prodi S, sebagian besar lulusan pada *cluster 3* memiliki status IPK Sangat Memuaskan. Lama studi yang ditempuh adalah sekitar 4 – 4,5 tahun.

TABEL VIII
HASIL CLUSTERING KMEANS LULUSAN I YANG TIDAK MENGIKUTI KEGIATAN KEMAHASISWAAN DENGAN 3 CLUSTER

Atribut	Cluster			
	Seluruh Data	1	2	3
Prodi	F	F	F	S
Status IPK	SM	SM	DP	SM
Lama Studi	5,26	4,98	4,08	4,32
Instances	100%	39%	34%	27%

7) *Lulusan yang Mengikuti Kegiatan Kemahasiswaan Seluruh Fakultas dengan 5 Cluster*: Pada Tabel IX, setelah hasil *cluster* ditambahkan menjadi 5, dapat disimpulkan bahwa hasil analisis didominasi oleh lulusan yang memiliki IPK sangat memuaskan. Walaupun rata-rata lama studi adalah 4,5 – 5 tahun, 15 % dari lulusan dapat meraih IPK Dengan pujian dengan masa studi 4 tahun. Hal ini menunjukkan bahwa dengan mengikuti unit kegiatan, mahasiswa masih dapat meraih IPK yang sangat baik.

TABEL IX
HASIL CLUSTERING KMEANS LULUSAN YANG MENGIKUTI KEGIATAN KEMAHASISWAAN DENGAN 5 CLUSTER

Atribut	Cluster					
	Seluruh Data	1	2	3	4	5
Fakultas	E	E	E	T	R	E
Unit	Unit Seni	Unit Seni	Unit Olahraga	Unit Bela Diri	Unit Olahraga	Unit Seni
Status IPK	SM	SM	SM	SM	DP	SM
Lama Studi	4,70	5,17	4,53	4,9	3,92	4,87
Keaktifan	Aktif	Pasif	Pasif	Aktif	Aktif	Aktif
Instances	100%	10%	17%	17%	15%	41%

8) *Lulusan Fakultas I yang Tidak Mengikuti Kegiatan Kemahasiswaan dengan 5 Cluster*: Hasil analisis pada Tabel X dengan jumlah 5 *cluster* menunjukkan bahwa lulusan Fakultas I masih dapat meraih IPK yang baik walaupun tidak mengikuti unit kegiatan dengan rata-rata lama studi 5 tahun. Sebanyak 27% dari lulusan meraih IPK Dengan Pujian dengan lama studi sekitar 3-4 tahun, 63% dari lulusan lulus dengan IPK Sangat Memuaskan dan 10% lulus dengan IPK Memuaskan.

TABEL X
HASIL CLUSTERING KMEANS LULUSAN FAKULTAS I YANG TIDAK MENGIKUTI KEGIATAN KEMAHASISWAAN DENGAN 5 CLUSTER

Atribut	Cluster					
	Seluruh Data	1	2	3	4	5
Jurusan	F	F	F	S	F	F
Status IPK	SM	SM	DP	SM	M	DP
Lama Studi	5,26	4,96	4,00	4,32	5,37	3,00
Instances	100%	38%	22%	25%	10%	5%

B. Metode Cobweb

Berikut ini adalah hasil *clustering* dengan metode *Cobweb* dari mahasiswa aktif dan lulusan yang mengikuti kegiatan kemahasiswaan yang meliputi seluruh fakultas, dan yang tidak mengikuti kegiatan meliputi Fakultas I. Dalam *Cobweb* terdapat parameter *acuity* dan *cutoff* yang digunakan akan menentukan jumlah *cluster* yang akan terbentuk. Nilai *acuity* dan *cutoff* yang digunakan dalam analisis ini adalah nilai yang dapat membuat jumlah *cluster* sekecil mungkin dan lebih besar dari nilai 1.

1) *Mahasiswa Aktif yang Mengikuti Kegiatan Kemahasiswaan Seluruh Fakultas*: Analisis seluruh mahasiswa aktif yang mengikuti dengan *Cobweb* dilakukan per fakultas dan hasil analisis yang ditinjau adalah atribut program studi, unit dan Status IPK yang dominan dari hasil setiap *cluster* yang terbentuk.

Tabel XI menunjukkan hasil analisis mahasiswa aktif seluruh fakultas dengan menggunakan metode *Cobweb*. Fakultas E membentuk 4 *cluster* dengan nilai *acuity* 0.9 dan *cutoff* 0.2. Hasil analisis yang terbentuk dari Fakultas E menunjukkan bahwa program studi dominan yang mengikuti kegiatan adalah A, dengan Unit Kegiatan yang paling banyak diikuti adalah Unit Seni. Pada *cluster* tersebut pula, sebagian besar mahasiswa memiliki status IPK Sangat Memuaskan.

Dari hasil analisis untuk fakultas P, terdapat 4 *cluster* yang terbentuk dengan atribut-atribut yang mendominasi adalah P untuk program studi, Unit Seni untuk unit kegiatan dan Memuaskan untuk atribut Status IPK.

Hasil analisis untuk fakultas R, terdapat 2 *cluster* yang terbentuk dengan atribut-atribut yang mendominasi adalah DK untuk program studi, Unit Seni untuk unit kegiatan dan Sangat Memuaskan untuk atribut Status IPK.

Untuk Fakultas H jumlah *cluster* yang terbentuk adalah 7 dengan nilai-nilai atribut yang mendominasi adalah H untuk prodi, Unit Seni untuk unit kegiatan dan Sangat Memuaskan untuk atribut Status IPK.

Pada Fakultas I jumlah *cluster* yang terbentuk adalah 6 dengan atribut yang mendominasi adalah F untuk atribut prodi, Unit Seni untuk unit kegiatan dan Sangat Memuaskan untuk Status IPK.

Dari hasil analisis untuk fakultas K, terdapat 5 *cluster* yang terbentuk dengan atribut-atribut yang mendominasi adalah K untuk program studi, Unit Seni untuk unit kegiatan dan Sangat Memuaskan untuk atribut Status IPK.

Hasil analisis untuk fakultas T, terdapat 3 *cluster* yang terbentuk dengan atribut-atribut yang mendominasi adalah TE untuk program studi, Unit Seni untuk unit kegiatan dan Sangat Memuaskan untuk atribut Status IPK.

Pada hasil analisis untuk fakultas S, terdapat 7 *cluster* yang terbentuk dengan atribut yang mendominasi adalah SI untuk program studi, Unit Seni untuk unit kegiatan dan Memuaskan untuk atribut Status IPK.

Berdasarkan hasil analisis ini, dapat disimpulkan bahwa mahasiswa aktif masih dapat meraih IPK yang baik bila mengikuti kegiatan, bahkan sebagian besar *cluster* yang terbentuk menunjukkan bahwa mahasiswa mampu mendapatkan IPK Sangat Memuaskan. Hanya 3 *cluster* dari seluruh fakultas yang menghasilkan status IPK Kurang Memuaskan.

TABEL XI
HASIL DETAIL CLUSTERING COBWEB MAHASISWA AKTIF SELURUH
FAKULTAS YANG MENGIKUTI KEGIATAN

Fakultas	Nilai		Cluster	Jumlah	Atribut Dominan		
	A	C			Prodi	Unit	Status IPK
E	0.9	0.2	1	22	M	Unit Seni	SM
			2	36	A	Unit Olahraga	SM
			3	51	A	Unit Seni	SM
			4	104	A	Unit Seni	SM
P	1.0	0.16	1	51	P	Unit Seni	SM
			2	9	P	Unit Seni	DP
			3	17	P	Unit Seni	KM
			4	3	P	Unit Kegiatan Sosial dan Organisasi	DP
			5	59	P	Unit Seni	M
R	0.1	0.17	1	58	DK	Unit Seni	SM
			2	25	DI	Unit Seni	SM
H	0.2	0.07	1	11	H	Unit Seni	SM
			2	3	H	Unit Seni	M
			3	1	H	Unit Bela Diri	DP
			4	1	H	Unit Olahraga	DP
			5	4	H	Unit Seni	DP
			6	1	H	Unit Kegiatan Sosial dan Organisasi	DP
			7	2	H	Unit Seni	KM
I	1.0	0.17	1	6	F	Unit Bela Diri	SM
			2	5	S	Unit Bela Diri	M
			3	6	F	Unit Olahraga	SM
			4	7	S	Unit Olahraga	M
			5	22	F	Unit Seni	SM

Fakultas	Nilai		Cluster	Jumlah	Atribut Dominan		
	A	C			Prodi	Unit	Status IPK
			6	10	S	Unit Seni	M
K	1.0	0.17	1	7	K	Unit Seni	SM
			2	1	K	Unit Bela Diri	M
			3	2	K	Unit Seni	DP
			4	2	K	Unit Seni	M
			5	6	K	Unit Seni	KM
T	1.0	0	1	36	TE	Unit Seni	SM
			2	19	TI	Unit Seni	DP
			3	18	TS	Unit Seni	SM
S	0.6	0.2	1	9	D3	Unit Bela Diri	SM
			2	9	SI	Unit Seni	SM
			3	2	SJ	Unit Seni	DP
			4	1	SJ	Unit Seni	M
			5	3	SM	Unit Seni	DP
			6	1	D3	Unit Seni	DP
			7	38	SI	Unit Seni	M

2) *Lulusan yang Mengikuti Kegiatan Kemahasiswaan Seluruh Fakultas*: Tabel XII menunjukkan hasil analisis lulusan untuk seluruh fakultas kecuali Fakultas H dengan menggunakan metode *Cobweb*. Hasil analisis yang terbentuk dari Fakultas E dengan 3 *cluster*, menunjukkan bahwa program studi dominan yang mengikuti kegiatan adalah M, dengan Unit Kegiatan yang paling banyak diikuti adalah Unit Olahraga. Pada *cluster* tersebut pula, sebagian besar mahasiswa memiliki status IPK Sangat Memuaskan.

Dari hasil analisis untuk fakultas P, terdapat 2 *cluster* yang terbentuk dengan atribut-atribut yang mendominasi adalah P untuk program studi, Unit Seni untuk unit kegiatan dan Sangat Memuaskan untuk atribut Status IPK.

Hasil analisis untuk fakultas R, terdapat 3 *cluster* yang terbentuk dengan atribut-atribut yang mendominasi adalah DI untuk program studi, Unit Seni untuk unit kegiatan dan Sangat Memuaskan untuk atribut Status IPK.

Pada Fakultas I jumlah *cluster* yang terbentuk adalah 7, dengan nilai-nilai atribut yang mendominasi adalah F untuk atribut prodi, Unit Bela Diri untuk unit kegiatan dan Sangat Memuaskan untuk Status IPK.

Dari hasil analisis untuk fakultas K, terdapat 3 *cluster* yang terbentuk dengan atribut-atribut yang mendominasi adalah K untuk program studi, Unit Seni untuk unit kegiatan dan Sangat Memuaskan untuk atribut Status IPK.

Hasil analisis untuk fakultas T, terdapat 4 *cluster* yang terbentuk dengan atribut-atribut yang mendominasi adalah TE untuk program studi, Unit Seni untuk unit kegiatan dan Sangat Memuaskan untuk atribut Status IPK.

Pada hasil analisis untuk fakultas S, terdapat 5 *cluster* yang terbentuk dengan atribut-atribut yang mendominasi

adalah SI untuk program studi, Unit Bela Diri dan Olahraga untuk unit kegiatan dan Sangat Memuaskan dan Dengan Pujian untuk atribut Status IPK.

Hasil analisis lulusan ini menunjukkan bahwa mengikuti unit kegiatan tidak mempengaruhi prestasi akademik lulusan. Lulusan cenderung memiliki Status IPK Sangat Memuaskan dan hanya satu cluster yang menghasilkan Status IPK Memuaskan.

TABEL XII
HASIL DETAIL CLUSTERING COBWEB LULUSAN SELURUH FAKULTAS YANG MENGIKUTI KEGIATAN

Fakultas	Nilai		Cluster	Jumlah	Atribut Dominan		
	A	C			Prodi	Unit	Status IPK
E	1.0	0.17	1	32	M	Unit Olahraga	SM
			2	22	M	Unit Seni	SM
			3	13	A	Unit Seni	DP
R	1.0	0.24	1	5	DK	Unit Olahraga	DP
			2	10	DI	Unit Seni	SM
			3	4	DI	Unit Seni	DP
I	0.5	0.19	1	1	F	Unit Seni	DP
			2	4	F	Unit Seni	SM
			3	1	F	Unit Olahraga	SM
			4	1	S	Unit Olahraga	SM
			5	7	F	Unit Bela Diri	SM
			6	2	S	Unit Bela Diri	SM
			7	1	S	Unit Bela Diri	DP
K	1.0	0.19	1	5	K	Unit Seni	SM
			2	2	K	Unit Olahraga	DP
			3	1	K	Unit Bela Diri	SM
P	1.0	0.14	1	12	P	Unit Seni	SM
			2	8	P	Unit Seni	M
S	1.0	0.19	1	2	SI	Unit Bela Diri	SM
			2	1	SJ	Unit Bela Diri	SM
			3	1	SM	Unit Seni	SM
			4	2	SI	Unit Olahraga	DP
			5	1	SJ	Unit Bela Diri	DP
T	1.0	0.10	1	4	TI	Unit Olahraga	DP
			2	4	TI	Unit Olahraga	SM
			3	11	TE	Unit Seni	SM
			4	10	TS	Unit Bela Diri	SM

3) Mahasiswa Aktif dan Lulusan Fakultas I yang Tidak Mengikuti Kegiatan Kemahasiswaan:

Tabel XIII menunjukkan hasil analisis mahasiswa aktif dan lulusan untuk Fakultas I dengan menggunakan metode Cobweb dilihat dari atribut program studi dan status IPK.

Hasil analisis mahasiswa aktif menghasilkan 3 cluster dengan program studi yang mendominasi adalah F dan Status IPK yang mendominasi adalah Memuaskan. Untuk hasil analisis lulusan Fakultas I, atribut program studi yang mendominasi adalah F dan Sangat Memuaskan untuk atribut Status IPK.

Hasil analisis ini menunjukkan mahasiswa dan lulusan yang tidak mengikuti kegiatanpun dapat meraih IPK yang baik. Jika dibandingkan dengan mahasiswa dan lulusan yang mengikuti kegiatan, sebagian besar Status IPK yang dihasilkan sama yaitu Sangat Memuaskan. Hal ini menunjukkan bahwa unit kegiatan tidak mempengaruhi prestasi akademik mahasiswa.

TABEL XIII
HASIL DETAIL CLUSTERING COBWEB MAHASISWA AKTIF DAN LULUSAN FAKULTAS I

Fakultas	Status	Nilai		Cluster	Jumlah	Atribut Dominan	
		A	C			Prodi	Status IPK
I	Aktif	1.0	0.17	1	432	F	M
				2	372	F	SM
				3	150	S	M
	Lulusan	1.0	0.17	1	108	S	SM
				2	134	F	DP
				3	143	F	SM

V. SIMPULAN

Makalah ini berfokus pada penerapan metode clustering terhadap analisis prestasi akademik bagi mahasiswa dan lulusan yang mengikuti kegiatan kemahasiswaan. Selain menganalisis prestasi akademik mahasiswa dan lulusan di seluruh fakultas Universitas Kristen Maranatha, analisis juga dilakukan berdasarkan setiap fakultas yang ada untuk mahasiswa dan lulusan yang mengikuti kegiatan kemahasiswaan. Sebagai pembandingan, dilakukan juga analisis prestasi akademik bagi mahasiswa dan lulusan yang tidak mengikuti kegiatan kemahasiswaan, khususnya untuk mahasiswa dari Fakultas I.

Berikut ini adalah hasil simpulan analisis yang telah dilakukan.

1. Mahasiswa aktif yang mengikuti unit kegiatan kemahasiswaan dan Mahasiswa aktif Fakultas I yang tidak mengikuti unit kegiatan. Dari hasil analisis mahasiswa aktif yang ikut unit kegiatan berdasarkan metode KMeans dan Cobweb, rata-rata Status IPK mahasiswa aktif yang mengikuti unit kegiatan sekitar 55% adalah Sangat Memuaskan, sekitar 25% meraih IPK Memuaskan dan 20% meraih IPK Kurang Memuaskan. Pada hasil analisis mahasiswa Fakultas I yang tidak ikut unit kegiatan menunjukkan bahwa sekitar 57% dari seluruh mahasiswa aktif memiliki IPK yang Memuaskan dan sekitar 33% memiliki IPK

Sangat Memuaskan. Mahasiswa aktif yang mengikuti Unit Kegiatan di Fakultas I sebagian besar mendapatkan IPK Sangat Memuaskan sedangkan mahasiswa yang tidak ikut unit kegiatan di Fakultas I sebagian besar meraih IPK Memuaskan.

2. Lulusan yang mengikuti unit kegiatan kemahasiswaan dan Lulusan Fakultas I yang tidak mengikuti unit kegiatan. Dari hasil analisis lulusan yang mengikuti unit kegiatan untuk metode KMeans dan Cobweb menunjukkan bahwa lulusan seluruh fakultas sebanyak 95% meraih IPK Sangat Memuaskan dan 15% meraih IPK Dengan Pujian. Menurut hasil analisis Kmeans 60% dari lulusan seluruh fakultas rata-rata menempuh lama studi 4,5 – 5 tahun. Pada hasil analisis mahasiswa dan lulusan Fakultas I yang tidak ikut kegiatan menunjukkan bahwa sebanyak sekitar 70% dari seluruh lulusan memiliki IPK yang Sangat Memuaskan sedangkan sekitar 30% memiliki IPK Dengan Pujian dengan rata-rata lama studi 4 – 5 tahun. Untuk Lulusan dari mahasiswa yang mengikuti unit kegiatan sebagian besar memperoleh IPK yang Sangat Memuaskan dan adapula yang meraih IPK Dengan Pujian. Jika dibandingkan dengan lulusan Fakultas I yang tidak mengikuti kegiatan, persentase lulusan yang memiliki IPK Dengan Pujian memang lebih besar dibandingkan dengan persentase lulusan yang mengikuti unit kegiatan. Menurut metode KMeans, baik lulusan yang mengikuti kegiatan maupun tidak mengikuti kegiatan sebagian besar mempunyai lama studi 4 – 5 tahun. Hal ini membuktikan bahwa dengan mengikuti kegiatan, lama studi lulusan tidak terpengaruh.

Untuk penelitian selanjutnya, dapat menggunakan data mahasiswa yang lebih banyak, baik mahasiswa aktif maupun lulusan yang mengikuti unit kegiatan maupun yang tidak mengikuti unit kegiatan. Akan lebih baik jika terdapat data mahasiswa dan lulusan dari berbagai fakultas bahkan dari berbagai universitas untuk mengetahui sejauh mana pengaruh keaktifan dalam kegiatan kemahasiswaan terhadap prestasi akademik mahasiswa.

DAFTAR PUSTAKA

- [1] BAN-PT, Buku 3A, Borang Akreditasi Sarjana, BAN-PT, 2010.
- [2] J. Han, M. Kamber and J. Pei, *Data Mining Concepts and Techniques Third Edition*, San Fransisco: Morgan Kauffman, 2012.
- [3] H. Olson and Y. Shi, *Introduction to Bussiness Data Mining*, New York: McGraw-Hill, 2007.
- [4] P.N. Tan, M. Steinbach and V. Kumar, *Introduction to Data Mining*. Pearson, 2005.
- [5] C. Anggarwal, K. Chandan, and Reddy, *Data Clustering: Algorithms and Applications*, CRC Press, 2014.
- [6] J.L. Andrews, P.D. McNicholas. Variable Selection for Clustering and Classification. *Journal of Classification*, Volume 31, no. 2, pp 136-153, July 2014.
- [7] D. T.Larose, *Discovering Knowledge In Data*, New Jersey: John Wiley & Sons, Inc., 2005.
- [8] B. Al-Shboul, and S. H. Myaeng. *Initializing K-Means Using GeneticAlgorithm*. World Academy of Science, Engineering and Technology, 2010.
- [9] I.H. Witten, E.Frank, M.A.Hall. *Data Mining: Practical Machine Learning Tools and Techniques*, Third Edition, San Francisco: Elsevier Inc, 2011.
- [10] N.Sharma, A. Bajpai, M.R. Litoriya. Comparison the various clustering algorithms of weka tools. *International Journal of Emerging Technology and Advanced Engineering*, Volume 2, no. 5, Mei 2012.
- [11] Z. Liu, R. George. *Mining Weather Data Using Fuzzy Cluster Analysis*. *Fuzzy Modeling with Spatial Information for Geographic Problems*, Springer Link, 2005.
- [12] M. H. Marghny, R. M. Abd El-Aziz, A.I. Taloba. An Effective Evolutionary Clustering Algorithm:Hepatitis C case study, *International Journal of Computer Applications*, Volume 34,No.6, November 2011.