

Analisis Statistik Pemilihan Bidang Skripsi Menggunakan *Expectation Maximization*

Yovi Pratama

Sekolah Tinggi Ilmu Komputer Dinamika Bangsa Jambi
Jl.Jendral Sudirman Thehok-Jambi, Indonesia
E-mail: yovi.pratama@gmail.com

Abstract

In a college major in computer science, students are required to apply what they have learned in an implementation or application on a specific area in the form of learning formats, analysis and application form, program or computer system as a condition of graduation (thesis). But with many fields being studied to make students confused determine which fields will be taken. Based on this thesis the researchers conducted recommendation field using Expectation Maximization Clustering algorithms. Cluster generated from data gathered research groups related subjects are taken from the Computer Science Curricula published by ACM and IEEE. From the data subjects related to the field of research, the authors matched subjects with course code contained in Computer Science program. The results of this study prove that recommendation Thesis fields using Expectation Maximization Clustering is able to provide the lowest value on reaching a value of 1% and a highest score of 100% and an average of 67% of the number of students as many as 11167 people.

Keywords: Thesis, Expectation Maximization *Clustering* algorithm, *Data mining*.

Abstrak

Pada sebuah perguruan tinggi jurusan ilmu komputer, mahasiswa dituntut untuk menerapkan apa yang mereka pelajari ke dalam sebuah implementasi atau penerapan pada bidang tertentu baik dalam bentuk format pembelajaran, analisis maupun berbentuk aplikasi, program atau sistem komputer sebagai syarat lulus (skripsi). Namun dengan banyaknya bidang yang dipelajari membuat mahasiswa bingung menentukan bidang yang akan diambil. Berdasarkan hal tersebut peneliti melakukan perekomendasi bidang skripsi menggunakan algoritma *Expectation Maximization Clustering*. *Cluster* yang dihasilkan berasal dari data kelompok mata kuliah terkait penelitian yang diambil dari *Computer Science Curricula* yang diterbitkan oleh ACM dan IEEE. Dari data mata kuliah yang terkait bidang penelitian, peneliti mencocokkan mata kuliah dengan kode mata kuliah yang terdapat pada program studi Teknik Informatika. Hasil dari penelitian ini membuktikan bahwa rekomendasi bidang Skripsi dengan menggunakan *Expectation Maximization Clustering* mampu memberi nilai rekomendasi terendah mencapai nilai 1 % dan nilai tertinggi 100 % serta rata-rata 67 % dari jumlah mahasiswa sebanyak 11.167 orang.

Kata Kunci: Skripsi, algoritma *Expectation Maximization Clustering*, *Data mining*

© 2017 Jurnal MEDIASISFO.

1. Pendahuluan

Skripsi merupakan persyaratan untuk mendapatkan status sarjana di setiap perguruan tinggi negeri maupun swasta yang ada di Indonesia. Untuk skripsi mahasiswa jurusan ilmu komputer, mahasiswa dituntut untuk menerapkan apa yang mereka pelajari ke dalam sebuah implementasi atau penerapan pada bidang tertentu baik dalam bentuk format pembelajaran, analisis maupun berbentuk aplikasi, program atau sistem komputer. Walaupun mahasiswa telah mempelajari banyak materi saat kuliah namun masih terdapat kendala dalam menentukan bidang skripsi yang akan diambil. Terlebih lagi untuk mahasiswa di jurusan ilmu komputer terkadang mengalami hambatan seperti kurangnya minat dan keahlian dalam

pemrograman, sehingga tidak semua mahasiswa jurusan ini mampu menerapkan pemrograman dalam menyelesaikan skripsi. Selain itu, bidang yang dipelajari di dalam ilmu komputer sangat banyak. Bidang tersebut masih merupakan suatu area yang masih luas ruang lingkupnya. Setiap bidang tersebut mempunyai sub area yang membawahi bidang tersebut. Berdasarkan hasil evaluasi terhadap korelasi antara nilai mahasiswa pada matakuliah tertentu dibandingkan dengan bidang skripsi yang diambil maka dapat diambil hipotesis sementara bahwa peminatan terhadap suatu bidang skripsi dapat dilihat berdasarkan kompetensi nilai mata kuliah ilmu komputer yang pernah diambilnya. Berdasarkan hal tersebut peneliti mencoba melakukan penelitian untuk menentukan bidang skripsi yang akan diambil. Dalam penelitian ini dilakukan sebuah perhitungan yang memberikan rekomendasi bidang skripsi mahasiswa berdasarkan nilai mata kuliah yang telah dicapai mahasiswa.

Penelitian ini menggunakan teknik *data mining* untuk mengelompokkan data dalam rekomendasi bidang skripsi kepada mahasiswa. Dalam *data mining*, banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain: Algoritma ID3, CART, dan C4.5[1]. Selain itu, *data mining* juga dapat dilakukan dengan *classification*, *association*, *clustering*, *estimation* dan *prediction* atau *forecasting*[2].

Pada penelitian ini akan digunakan metode *clustering* dalam mengelompokkan data mahasiswa kepada suatu bidang skripsi tertentu berdasarkan nilai. Metode *clustering* digunakan karena teknik ini mampu membagi rangkaian data *non time series* menjadi beberapa kelompok berdasarkan kesamaan-kesamaan yang telah ditentukan oleh metode yang digunakan[2]. *Clustering* terdiri dari jenis *hierarchial* dan *non-hierarchial*. Metode *clustering* yang paling banyak digunakan adalah Agglomerative, K-Means, dan *Expectation Maximization*. Pemilihan algoritma *Expectation Maximization* dikarenakan dapat menangani *missing value*[2]. Selain itu, algoritma tersebut memiliki performansi yang bagus, di mana hasil perbandingan *non hierarchial clustering* antara K-Means dan *Expectation Maximization* mempunyai nilai sama 89%. Namun, ketika diujikan dengan data yang lebih banyak *Expectation Maximization* dapat menghasilkan lebih tinggi yakni 90 %[3] sehingga *Expectation Maximization* mampu mempunyai prediksi yang baik[4]. Berdasarkan hal tersebut, maka penelitian ini menganalisis secara statistic penggunaan *expectation maximization clustering* untuk merekomendasikan bidang skripsi mahasiswa agar mampu menghasilkan nilai yang baik pula.

2. Tinjauan Pustaka/ Penelitian Sebelumnya

Sebelumnya, penentuan atau rekomendasi seperti penelitian ini sudah dilakukan oleh Zhiwu Liu, dkk [5] yang menggunakan Pohon Keputusan. Selain itu ada Rong Cao dan Lizhen Xu[6] menggunakan Algoritma C4.5 untuk analisa penentuan penjualan. Isnawati menentukan kelas modul itdial dengan teknik *clustering* digabung dengan *template matching* menghasilkan nilai kecocokan yang tinggi[7]. Sementara itu, dalam bidang pendidikan[8] telah dilakukan model fuzzy untuk membuat klasifikasi siswa berhasil atau gagal. Terdapat juga, Wen-Chih Chang, dkk[9] yang mengukur kemampuan belajar siswa menggunakan K-Means. Kemudian, penelitian lainnya, mengukur keakuratan pemilihan jurusan SMA menggunakan algoritma Fuzzy C-Means berdasarkan nilai[10], namun penelitian ini berfokus pada penerapan teknik yang digunakan saja. Sampel data yang di uji sebanyak 81 siswa. Selanjutnya, penelitian lainnya merekomendasikan peminatan Skripsi dengan menerapkan Fuzzy C-Means dengan tingkat akurasi sebesar 82 %[11]. Dari berbagai penelitian tersebut, masih terdapat kelemahan seperti sedikitnya data yang digunakan, namun, walau demikian penelitian tersebut telah membuktikan bahwa teknik *clustering* dapat digunakan untuk proses rekomendasi.

2.1 Data Mining

Data mining merupakan suatu proses yang menggunakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis [6][12]. *Data Mining* dapat digunakan untuk mengelompokkan data, memprediksi, mengestimasi, dan menentukan kaidah asosiasi dalam suatu data yang ada. Perlunya *data mining* karena adanya sejumlah besar data yang dapat digunakan untuk menghasilkan informasi dan *knowledge* yang berguna. Informasi dan *knowledge* yang didapat tersebut digunakan untuk mengetahui pola dalam suatu data yang sangat banyak [8].

Dalam penelitian ini, *data mining* dilakukan dengan cara melakukan perhitungan terhadap ribuan *record* data mahasiswa menggunakan algoritma *expectation maximization clustering* secara iterasi hingga nilai sebuah kelas Y mencapai nilai konvergen. Nilai konvergen yang dicapai artinya seorang mahasiswa telah direkomendasikan kepada sebuah bidang skripsi. Keluaran rekomendasi yang dihasilkan berupa persentasi.

2.2 Clustering

Clustering adalah metode yang digunakan untuk membagi rangkaian data menjadi beberapa kelompok berdasarkan kesamaan-kesamaan yang telah ditentukan sebelumnya [1][13]. Pengertian *clustering* keilmuan dalam *data mining* adalah pengelompokan sejumlah obyek ke dalam *cluster* sehingga dalam sebuah *cluster* tersebut berisi data yang mempunyai kemiripan yang tinggi dan berbeda dengan obyek dalam *cluster* yang lainnya. Sampai saat ini, para ilmuwan masih terus melakukan berbagai usaha untuk melakukan perbaikan model *cluster* dan menghitung jumlah *cluster* yang optimal sehingga dapat dihasilkan *cluster* yang paling baik.

2.2.1 Expectation maximization Clustering

Algoritma *expectation maximization* merupakan algoritma *clustering* yang melakukan estimasi *Maximum Likelihood* dari parameter dalam sebuah model probabilistik. Algoritma ini akan melakukan *clustering* terhadap objek data termasuk data yang mengalami *missing value*[14] (dalam penelitian *missing value* berupa data mahasiswa yang kosong). Algoritma *expectation maximization* termasuk algoritma *clustering* yang berbasis perhitungan probabilitas. Secara iteratif algoritma *expectation maximization* melakukan 2 tahapan yaitu [2][14]:

a. *Expectation Step*,

Pada tahap ini, akan menghitung probabilitas objek terhadap kluster.

b. *Maximization Step*

Pada tahap ini, akan dihitung ulang nilai parameter-parameter yang digunakan untuk memaksimalkan nilai probabilitas. Kedua tahap tersebut dilakukan secara berulang-ulang sampai mencapai nilai konvergen. Yang dimaksud dengan nilai konvergen ini adalah nilai *threshold* yang digunakan untuk menghentikan iterasi yang terus menerus dimana iterasi tersebut tidak akan berhenti jika nilai nya tak nol (0). Langkah-langkah algoritma *expectation maximization* dapat dilihat sebagai berikut:

1. Menentukan jumlah *cluster*(k) kemudian data-data yang ada di dalam *cluster*.
2. Menentukan *mean*(μ), *covariance*(Σ) dan peluang *cluster*(π) untuk setiap

cluster dan hitung *log-likelihood*-nya (l^0)

$$\pi_k^0 = \frac{\text{jumlah_point_k}}{\text{Total_point}} \quad (1)$$

$$\mu_k^0 = \sum_i^{n_c} \frac{x_i}{n_c} \quad (2)$$

$$\Sigma_k^0 = \frac{\sum (x_i - \text{mean})^2}{n_c - 1} \quad (3)$$

$$l^0 = \frac{1}{N} \sum_{n=1}^N \log(\sum_{i=1}^k \pi_k^0 \Phi(x_n | \mu_k^0, \Sigma_k^0)) \quad (4)$$

3. Melakukan ekspektasi dengan menghitung $\tau(x_{nk})$:

$$\tau(x_{nk}) = \frac{\pi_k \Phi(x_n | \mu_k, \Sigma_k)}{\sum_{j=1}^k \pi_j \Phi(x_n | \mu_j, \Sigma_j)} \quad (5)$$

$$\Phi(x_n | \mu_k, \Sigma_k) = \frac{1}{\sqrt{2\pi\Sigma_k}} e^{-\frac{(x_n - \mu)^2}{2\Sigma}} \quad (6)$$

Bandingkan tiap nilai $\tau(x_{nk})$, jika $\tau(x_{11}) > \tau(x_{12})$ maka x_1 diletakan di *cluster* 1, Jika $\tau(x_{21}) < \tau(x_{22})$ maka x_2 diletakan di *cluster* 2, Jika $\tau(x_{31}) = \tau(x_{32})$ maka x_3 diletakan di *cluster* 1.

4. Melakukan *maximization* dengan menghitung ulang nilai tiap parameter:

$$\pi_k^{new} = \frac{\text{jumlah_objek_k_baru}}{N} \quad (7)$$

$$\mu_k^{new} = \frac{\sum_{n=1}^{N_c} \tau(x_{nk}) x_n}{N_k} \quad (8)$$

$$\Sigma_k^{new} = \frac{1}{N_k} \sum_{n=1}^{N_c} \tau(x_{nk}) (x_n - \mu_k^{new})^2 \quad (9)$$

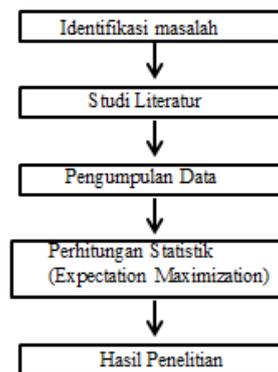
$$N_k = \sum_{n=1}^{N_c} \tau(x_{nk}) \quad (10)$$

5. Mengevaluasi *log-likelihood* baru

$$l^{new} = \frac{1}{N} \sum_{n=1}^N \log(\sum_{k=1}^k \pi_k \Phi(x_n | \mu_k^{new}, \Sigma_k^{new})) \quad (11)$$
6. Memasukkan nilai mean estimasi ke dalam *missing value* sesuai klusternya, cek apakah konvergen. Jika $|l^{new} - l^0| > \xi$ dengan ξ adalah nilai *threshold*, jika ya maka iterasi berhenti, sebaliknya kembali ke tahap 4.

3. Metodologi

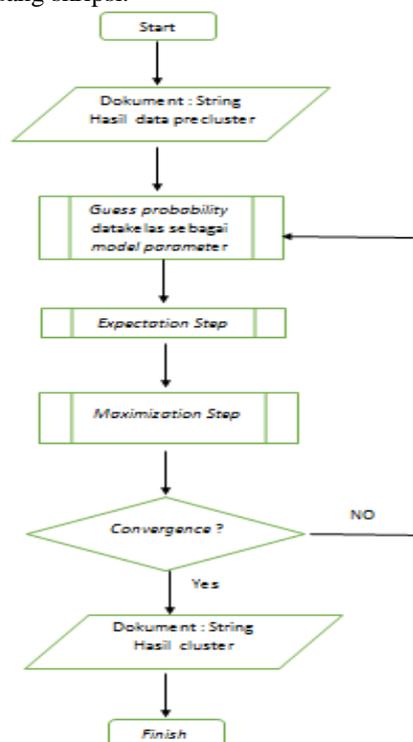
Metodologi penelitian menggambarkan tahapan proses atau metode yang digunakan dalam melakukan penelitian agar penelitian dapat berjalan dengan baik dan tujuan yang telah ditetapkan dapat tercapai. Pada penelitian ini peneliti menggunakan tahapan kegiatan penelitian yang berupa kerangka kerja penelitian yang peneliti lakukan.



Gambar 1. Kerangka Kerja Penelitian

3.1 Alur Algoritma Expectation Maximization Rekomendasi Bidang Skripsi

Berikut ini adalah bagan yang menggambarkan alur kinerja algoritma *Expectation Maximization Clustering* pada rekomendasi bidang skripsi.



Gambar 2 Alur Algoritma Expectation Maximization untuk perhitungan Rekomendasi Bidang Skripsi

Pada bagan tersebut proses *clustering* atau pengelompokan mahasiswa dilakukan berdasarkan bidang ilmu komputer yang direkomendasikan untuk skripsi mahasiswa. *Cluster* yang dihasilkan berasal dari data kelompok mata kuliah terkait penelitian yang diambil dari Computer Science Curricula yang diterbitkan oleh ACM dan IEEE. Dari data mata kuliah yang terkait bidang penelitian, peneliti mencocokkan mata kuliah dengan kode mata kuliah yang terdapat pada program studi Teknik Informatika. Mata kuliah yang terkait dengan nilai maksimal atau nilai A akan menjadi acuan suatu bidang penelitian ilmu Komputer. Hanya mata kuliah terkait yang bernilai A yang akan disesuaikan dengan suatu bidang. Bidang yang terkait tersebut akan menjadi suatu rekomendasi bidang skripsi.

Perhitungan rekomendasi bidang skripsi mahasiswa dengan menggunakan algoritma *ekspectation maximization clustering* ini melakukan cluster mahasiswa teknik informatika menjadi ke dalam beberapa cluster bidang skripsi. Terdapat 18 area bidang ilmu komputer dengan mata kuliah yang terkait bidang tersebut berdasarkan *Computer Science Curricula 2013* yang dikeluarkan oleh ACM dan IEEE. Selanjutnya, peneliti melakukan mencocokkan kode mata kuliah dalam kurikulum Teknik Informatika Sekolah Tinggi Ilmu Komputer Dinamika Bangsa. Untuk daftar bidang terkait mata kuliah dalam *Computer Science Curricula* tersebut dapat dilihat pada tabel berikut:

Tabel 1. Daftar Matakuliah Sesuai 18 Area Bidang Komputer yang tercantum dalam KKNi Bidang Informatika dan Ilmu Komputer 2016 (tabel 4-12, Hal 81, Validasi oleh Aptikom) Berdasarkan Computer Science Curricula 2013 (Validasi Oleh Association for Computing Machinery (ACM)) [15]

Kelas (Y)	Computer Science Curricula 2013 (ACM dan IEEE)		
	Bidang Komputer	Fitur (X)	MATA KULIAH
Y=1	Algoritma Dan Kompleksitas (<i>Algorithms and Complexity (AL)</i>)	x1	Algoritma dan pemrograman
		x2	Lab. Algoritma dan pemrograman
		x3	Struktur Data
		x4	Lab. Struktur data
		x5	Pemrograman Berorientasi Objek
		x6	Lab. Pemrograman Berorientasi Objek
Y=2	(Arsitektur dan Organisasi Komputer) <i>Architecture and Organization (AR)</i>	x7	Pengantar Teknologi Informasi
		x8	Lab. Pengantar Teknologi Informasi
		x9	Organisasi Komputer
		x10	Sistem Operasi
		x11	Kecerdasan Buatan
		x12	Sistem Embedded
Y=3	Ilmu Komputasi (<i>Computational Science or Numerical Methods (CN)</i>)	x13	Kalkulus I
		x14	Kalkulus II
		x15	Matematika Diskrit
Y=4	Struktur Diskrit (<i>Discrete Structures (DS)</i>)	x16	Matematika Diskrit
		x17	Logika matematika
		x18	Probabilitas dan statistik
Y=5	<i>Graphics and Visual Computing (GV)</i>	x19	Algoritma dan pemrograman
		x20	Lab. Algoritma dan pemrograman
		x21	Pengolahan Citra Digital
		x22	Komputer Grafik
		x23	Lab. Komputer Grafik
		x24	Interaksi Manusia dan Komputer.
		x25	Lab. Interaksi Manusia dan Komputer
Y=6	<i>Human-Computer Interaction (HCI)</i>	x26	Interaksi Manusia dan Komputer
		x27	Lab. Interaksi Manusia dan Komputer
		x28	Multimedia
		x29	Lab. Multimedia

Y=7	<i>IAS - Information Assurance and Security</i>	x30	Keamanan Jaringan
		x31	Forensik Komputer
		x32	Jaringan Komputer
Y=8	(Manajemen Informasi) <i>IM - Information Management</i>	x33	Sistem Informasi
		x34	Sistem Informasi Perbankan
		x35	Analisa & Perancangan Sistem
		x36	<i>Data mining</i>
		x37	Basis Data Lanjut
Y=9	Sistem Cerdas (<i>IS - Intelligent System</i>)	x38	Teknologi Basis Data
		x39	Kecerdasan Buatan
Y=10	Jaringan dan Komunikasi (<i>NC - Networking and Communication</i>)	x40	Sistem Pendukung Keputusan
		x41	Jaringan Komputer
Y=11	Sistem Operasi (<i>Operating Systems (OS)</i>)	x42	Keamanan Jaringan
		x43	Sistem Operasi Lanjut
		x44	Organisasi Komputer
		x45	Sistem Digital
		x46	Sistem Operasi
		x47	Sistem Operasi lanjut
Y=12	Pengembangan Berbasis Platform (<i>PBD</i>)	x48	Struktur Data
		x49	Lab. Struktur data
Y=13	Komputasi Paralel dan Terdistribusi (<i>PD - Parallel and Distributed Computing</i>)	x50	Rekayasa Pengembangan Web.
		x51	Sistem Terdistribusi
		x52	Sistem Digital
		x53	Organisasi Komputer
		x54	Komunikasi data
		x55	Lab. Komunikasi data
		x56	Jaringan Komputer
		x57	Lab. Jaringan Komputer
		x58	Management jaringan
		x59	Networking Advanced
Y=14	Bahasa Pemrograman (<i>Programming Languages (PL)</i>)	x60	Keamanan jaringan
		x61	Algoritma dan Pemrograman
		x62	Lab Algoritma dan Pemrograman
		x63	Struktur Data
		x64	Lab. Struktur data
		x65	Pemrograman web
		x66	Lab. Pemrograman web
		x67	Pemrograman desktop (VB)
		x68	Lab. Pemrograman desktop (vb)
		x69	Pemrograman Berorientasi Objek
		x70	Lab. Pemrograman Berorientasi Objek
Y=15	<i>Software Development Fundamentals (SDF)</i>	x71	Assembly Language
		x72	Algoritma dan Pemrograman
		x73	Lab Alpro
		x74	Algorithms and Complexity
Y=16	<i>Software Engineering (SE)</i>	x75	Rekayasa Perangkat Lunak
		x76	Rekayasa Perangkat Lunak
		x77	Teknologi Game
		x78	Manajemen Proyek

		x79	APSI
		x80	Rekayasa Perangkat Lunak Lanjut
Y=17	SF - Systems Fundamentals	x81	Jaringan Komputer
		x82	Arsitektur dan Organisasi Komputer
		x83	Sistem Operasi
		x84	Sistem Operasi lanjut
Y=18	Isu Sosial dan Praktik Profesional (SP - Social Issues and Professional Issues)	x86	Pendidikan Agama dan Etika
		x87	Kerja Praktek
		x88	Pengantar Manajemen
		x89	Kewarganegaraan

*Terdapat kesamaan mata kuliah dengan bidang yang berbeda adalah ketentuan yang sudah tercantum dalam KKNi 2016.

Perhitungan Ekspektasi probabilitas dilakukan dengan contoh 3 kelas Y=1, Y=2 dan Y=3 dengan 4 fitur X=1, X=2, X=3, X=4. Kelas Y adalah kluster bidang skripsi yang akan direkomendasikan kepada mahasiswa berdasarkan fitur X. Fitur X merupakan representasi dari matakuliah dan nilainya. Sebagai contoh jika mata kuliah Kecerdasan Buatan bernilai A, B+, B, C+, C, D dan E. Fitur bernilai 1 pada nilai yang dimiliki mahasiswa dan bernilai 0 jika tidak memiliki nilai. Berikut adalah contoh perhitungan dalam proses rekomendasi bidang skripsi teknik informatika dengan menggunakan algoritma *Expectation-Maximization*. Perhitungan dengan algoritma ini mempunyai dua tahap utama yakni tahap perkiraan (*Expectation*) dan tahap maksimalisasi (*Maximization*), di mana tahap ini akan diulang (*increment*) sampai nilai kelas Y mencapai nilai 0 atau nilai 1 [16]. Secara teori perhitungan *Expectation Maximization* dijelaskan dari paper yang dibuat oleh Arthur Dempster, dan laird dan Donald Rubin dalam jurnal 'Royal Statistical Society' pada tahun 1977 [17]. Algoritma EM termasuk algoritma *clustering* yang berbasis perhitungan probabilitas. Secara iteratif Algoritma EM melakukan 2 tahapan yaitu [17]:

a. *Expectation Step*

Pada tahap ini, akan menghitung probabilitas objek terhadap kluster.

b. *Maximization Step*

Pada tahap ini, akan menghitung ulang nilai parameter-parameter yang digunakan untuk memaksimalkan nilai probabilitas.

Kedua tahap tersebut dilakukan secara berulang-ulang sampai mencapai nilai konvergen [17]. Yang dimaksud dengan nilai konvergen ini adalah nilai *threshold* yang digunakan untuk menghentikan iterasi yang terus menerus dimana iterasi tersebut tidak akan berhenti jika nilai nya tak nol 0.

Tahap *Expectation*:

Tabel 2. Kondisi Awal Probabilitas Kelas Y yang dimiliki setiap fitur X

Y=1	Y=2	Y=3	X1	X2	X3	X4
?	?	?	0	1	0	1
?	?	?	1	0	1	0
?	?	?	0	1	1	1
?	?	?	1	1	1	0
?	?	?	0	1	0	0
?	?	?	0	0	1	1

Pada table 2 di atas, kelas Y=1, Y=2 dan Y=3 adalah data awal yang akan dihitung probabilitas kelas yang dimiliki oleh setiap fitur X1, X2, X3 dan X4. Tabel tersebut menunjukkan bahwa nilai kelas Y belum diketahui, untuk data setiap fitur dapat dilihat bernilai 1 jika mata kuliah yang diambil misalnya bernilai A dan 0 jika mata kuliah tersebut bukan bernilai A.

Tabel 3. Frekuensi Fitur X

Y	X1	X2	X3	X4
1	?	?	?	?
2	?	?	?	?
3	?	?	?	?

Tabel 3 di atas adalah table frekuensi yang dimiliki setiap fitur terhadap kelas Y=1, Y=2 dan Y=3. Pada awal kondisi nilai frekuensi setiap fitur x1, x2, x3 dan x4 belum diketahui. Untuk dapat menghitung nilai frekuensi setiap fitur terhadap setiap kelas, maka digunakan sebuah model parameter. Tabel 4 berikut ini menampilkan nilai model parameter.

Tabel 4. Model Parameter

Y	X1	X2	X3	X4	P(y)
1	0.1	0.3	0.8	0.8	0.7
2	0.2	0.3	0.1	0.1	0.2
3	0.7	0.4	0.1	0.1	0.1

Pada tabel 4 yang berisi model parameter masing-masing nilai X1, X2, X3 dan X4 diperoleh dari secara random dimana jumlah total seluruh fitur yang sama dari seluruh fitur harus berjumlah 1 (satu). Dapat dilihat bahwa $x1 \rightarrow 0.1+0.2+0.7 = 1$, $x2 \rightarrow 0.3+0.3+0.4 = 1$, $x3 \rightarrow 0.8 + 0.1 + 0.1 = 1$, $x4 \rightarrow 0.8+0.1+0.1=1$ dan $\alpha \rightarrow 0.1+0.2+0.1=1$. Dari nilai yang dihasilkan secara random ini berikutnya akan dihitung frekuensi setiap fitur X beserta nilai probabilitas terhadap kelas Y seperti yang ditampilkan pada tabel 5 berikut ini.

Tabel 5. Frekuensi fitur X baru dan Probabilitas kelas Y baru

prob Y=1					
X1	X2	X3	X4	P(y)	N
0.9	0.3	0.2	0.8	0.7	0.03024
0.1	0.7	0.8	0.2	0.7	0.00784
0.9	0.3	0.8	0.8	0.7	0.12096
0.1	0.3	0.8	0.2	0.7	0.00336
0.9	0.3	0.2	0.2	0.7	0.00756
0.9	0.7	0.8	0.8	0.7	0.28224
prob Y=2					
X1	X2	X3	X4	P(y)	N
0.8	0.3	0.9	0.1	0.2	0.00432
0.2	0.7	0.1	0.9	0.2	0.00252
0.8	0.3	0.1	0.1	0.2	0.00048
0.2	0.3	0.1	0.9	0.2	0.00108
0.8	0.3	0.9	0.9	0.2	0.03888
0.8	0.7	0.1	0.1	0.2	0.00112
prob Y=3					
X1	X2	X3	X4	P(y)	N
0.3	0.4	0.9	0.1	0.1	0.00108
0.7	0.6	0.1	0.9	0.1	0.00378
0.3	0.4	0.1	0.1	0.1	0.00012
0.7	0.4	0.1	0.9	0.1	0.00252
0.3	0.4	0.9	0.9	0.1	0.00972
0.3	0.6	0.1	0.1	0.1	0.00018

$$N = X1 * X2 * x3 * x4 * P(y); \tag{12}$$

Perhitungan setiap fitur X baru pada tabel 5 diperoleh dari nilai model parameter yang terdapat di tabel 4. Selanjutnya untuk mengisi frekuensi X baru pada tabel 5 tersebut akan di periksa nilai awal fitur X yang terdapat pada tabel 2. Perhitungan yang digunakan adalah jika nilai fitur X pada data awal bernilai 1, maka ambil nilai frekuensi yang dihasilkan secara random pada tabel model parameter, jika 0 maka

frekuensi X baru adalah 1 dikurangi nilai frekuensi yang dihasilkan secara random pada model parameter tabel 4. Selanjutnya setelah semua nilai frekuensi fitur X baru dihasilkan, maka dapat dihitung nilai probabilitas kelas Y baru.

Tabel 6. Nilai probabilitas baru dari kelas Y

Y=1 baru	Y=2 baru	Y=3 baru
0.85	0.12	0.03
0.55	0.18	0.27
1.00	0.00	0.00
0.48	0.16	0.36
0.13	0.69	0.17
1.00	0.00	0.00

Maka, probabilitas setiap instance/ fitur terhadap sebuah kelas Y adalah sebagai berikut:

$$\prod P(y|k) = P'(y) / \sum P'(y) \quad (13)$$

Di mana nilai $\prod P(y|k)$ adalah nilai probabilitas setiap fitur terhadap sebuah kelas Y, nilai $P'(y)$ adalah nilai probabilitas yang dimiliki sebuah kelas Y, nilai $\sum P'(y)$ adalah jumlah seluruh probabilitas yang dimiliki seluruh kelas Y.

- A. $\prod P(y|k)$ terhadap kelas Y=1
 Instance1 : $0.03024 / (0.03024 + 0.00432 + 0.00108) = 0.85$
 Instance2 : $0.00784 / (0.00784 + 0.00252 + 0.00378) = 0.55$
 Instance3 : $0.12096 / (0.12096 + 0.00048 + 0.00012) = 1.00$
 Instance4 : $0.00336 / (0.00336 + 0.00432 + 0.00252) = 0.48$
 Instance5 : $0.00756 / (0.00756 + 0.03888 + 0.00972) = 0.13$
 Instance6 : $0.28224 / (0.28224 + 0.00112 + 0.00018) = 1.00$
- B. $\prod P(y|k)$ terhadap kelas Y=2
 Instance1 : $0.00432 / (0.03024 + 0.00432 + 0.00108) = 0.12$
 Instance2 : $0.00252 / (0.00784 + 0.00252 + 0.00378) = 0.18$
 Instance3 : $0.00048 / (0.12096 + 0.00048 + 0.00012) = 0.00$
 Instance4 : $0.00432 / (0.00336 + 0.00432 + 0.00252) = 0.16$
 Instance5 : $0.03888 / (0.00756 + 0.03888 + 0.00972) = 0.69$
 Instance6 : $0.00112 / (0.28224 + 0.00112 + 0.00018) = 0.00$
- C. $\prod P(y|k)$ terhadap kelas Y=3
 Instance1 : $0.00108 / (0.03024 + 0.00432 + 0.00108) = 0.03$
 Instance2 : $0.00378 / (0.00784 + 0.00252 + 0.00378) = 0.27$
 Instance3 : $0.00012 / (0.12096 + 0.00048 + 0.00012) = 0.00$
 Instance4 : $0.00252 / (0.00336 + 0.00432 + 0.00252) = 0.36$
 Instance5 : $0.00972 / (0.00756 + 0.03888 + 0.00972) = 0.17$
 Instance6 : $0.00018 / (0.28224 + 0.00112 + 0.00018) = 0.00$

Tabel 6 di atas menunjukkan nilai probabilitas kelas Y yang baru. Nilai tersebut ditunjukkan oleh nilai Y=1 baru, Y=2 baru dan nilai Y=3 baru. Nilai tersebut diperoleh dari hasil perbandingan setiap nilai N dengan total nilai N. Dengan perhitungan sebagai berikut:

$$Y_{\text{baru}} = N_{(\text{prob } Y_1)} / (N_{(\text{prob } Y_1)} + N_{(\text{prob } Y_2)} + N_{(\text{prob } Y_3)}) \quad (14)$$

Selanjutnya nilai baru tersebut akan digunakan pada iterasi pertama sebagai tahap *Maximization*. Berikut tahap *Maximization* nilai probabilitas setiap kelas Y:

Tahap Maximization:

Tabel 7. Data Probabilitas Kelas yang dimiliki setiap fitur dalam iterasi -I

Kelas			Fitur			
Y=1	Y=2	Y=3	X1	X2	X3	X4
0.85	0.12	0.03	0	1	0	1
0.55	0.18	0.27	1	0	1	0
1.00	0.00	0.00	0	1	1	1
0.48	0.16	0.36	1	1	1	0
0.13	0.69	0.17	0	1	0	0
1.00	0.00	0.00	0	0	1	1

Tabel 8. Tabel Frekuensi nilai dalam iterasi -I

Y	X1	X2	X3	X4	Total
1	1.04	2.46	3.03	2.84	4.01
2	0.33	0.97	0.34	0.13	1.15
3	0.63	0.57	0.63	0.03	0.83

Probabilitas seluruh data (X1.....X4) terhadap kelas Y={ 0,1,2}:

a. Y'(X) Untuk Y=1:

$$Y'(x) = \sum P(X|Y') = 0.85 + 0.55 + 1.00 + 0.48 + 0.13 + 1.00 = 4.01$$

b. Y'(X) Untuk Y=2:

$$Y'(x) = \sum P(X|Y') = \{0.12 + 0.18 + 0.00 + 0.16 + 0.69 + 0.00\} = 1.15$$

c. Y'(X) Untuk Y=3:

$$Y'(x) = \sum P(X|Y') = \{0.03 + 0.27 + 0.00 + 0.36 + 0.17 + 0.00\} = 0.83$$

Jadi, masing –masing nilai fitur x adalah hasil jumlah seluruh nilai fitur x dikali dengan nilai probabilitas kelas Y yang dihasilkan oleh hasil tahap sebelumnya. Selanjutnya, pada iterasi ini lakukan kembali perhitungan model parameter.

Tabel 9. Model Parameter dalam iterasi -I

Y	X1	X2	X3	X4	alpha
1	0.26	0.61	0.75	0.71	0.67
2	0.29	0.84	0.30	0.11	0.19

Untuk nilai model parameter yang dihasilkan dalam iterasi I tahap *Maximization* ini dilakukan dengan cara membagi nilai frekuensi setiap fitur x dengan total frekuensi fitur [17][3]. Hingga selanjutnya seluruh tahap yang dilakukan dalam proses *maximization* ini dilakukan secara berulang terus menerus hingga nilai probabilitas kelas yang dihasilkan mencapai nilai 0 atau 1 yang artinya setiap mahasiswa dapat ditentukan akan termasuk kepada bidang skripsi apa yang cocok (kelas Y yang terpilih) berdasarkan nilai mata kuliah yang diperoleh (fitur x). Berikut adalah contoh nilai probabilitas kelas Y yang sudah mencapai nilai ambang 0 atau 1 .

Tabel 10. Nilai probabilitas Iterasi terakhir dari kelas Y (Kelas hanya bernilai 0 atau 1)

Kelas			Fitur			
Y=0	Y=1	Y=2	X1	X2	X3	X4
0.00	1.00	0.00	0	1	0	1
0.00	0.00	1.00	1	0	1	0
1.00	0.00	0.00	0	1	1	1

0.00	0.00	1.00	1	1	1	0
0.00	1.00	0.00	0	1	0	0
1.00	0.00	0.00	0	0	1	1

4. Hasil dan Pembahasan

4.1 Analisis Rekomendasi Bidang Skripsi Mahasiswa

Selama ini proses rekomendasi bidang skripsi mahasiswa dilakukan secara manual. Yakni baik dosen yang memberi masukan atau ide yang diperoleh dari berbagai makalah penelitian. Dalam arti kata, proses yang telah berjalan belum menggunakan metode secara statistik. Metode statistik sendiri dapat memberikan nilai rekomendasi yang terukur dan mampu dilakukan secara otomatis.

Dari penelitian yang peneliti lakukan dapat disimpulkan bahwa secara statistik peminatan terhadap suatu bidang skripsi dapat dilihat berdasarkan nilai mata kuliah menggunakan algoritma *expectation maximization clustering*. Rekomendasi bidang yang diberikan adalah hasil *maximization* dari kumpulan nilai mata kuliah mahasiswa yang mencapai nilai satu (1) atau nol (0) terhadap kelas Y (Bidang Skripsi). Nilai 1 mengindikasikan bahwa nilai yang diperoleh mahasiswa masuk ke kelas Y dan nilai nol (0) mempunyai arti sebaliknya. Subbab berikut menjelaskan hasil persentasi rekomendasi suatu bidang skripsi kepada mahasiswa.

4.2 Pembahasan Hasil

Setelah dilakukan secara statistik menggunakan algoritma *Expectation-Maximization*, seluruh hasil rekomendasi bidang skripsi berdasarkan nilai mata kuliah maka dapat disimpulkan bahwa rekomendasi ini mampu memberikan persentasi hingga 100 % terhadap mahasiswa berdasarkan nilai mata kuliah.

Tabel 11. Nilai probabilitas Iterasi terakhir dari kelas Y (Kelas hanya bernilai 0 atau 1)

Kelas (Y)	Bidang Komputer	Persentasi	Jumlah mahasiswa
Y=1	<i>Algorithms and Complexity (AL)</i>	100%	350
		50%	170
		2%	50
Y=2	<i>Architecture and Organization (AR)</i>	100%	345
		75%	210
		1%	66
Y=3	<i>Computational Science or Numerical Methods (CN)</i>	100%	250
		98%	320
		3%	78
Y=4	<i>Discrete Structures (DS)</i>	100%	250
		98%	320
		3%	78
Y=5	<i>Graphics and Visual Computing (GV)</i>	100%	250
		98%	320
		3%	78
Y=6	<i>Human-Computer Interaction (HCI)</i>	100%	345
		75%	210
		1%	66
Y=7	<i>IAS - Information Assurance and Security</i>	100%	350
		50%	170
		2%	50
Y=8	<i>IM - Information Management</i>	100%	250
		98%	320
		3%	78
Y=9	<i>IS - Intelligent System</i>	100%	350

		50%	170
Y=10	<i>NC - Networking and Communication</i>	2%	50
		100%	350
		50%	170
Y=11	<i>Operating Systems (OS)</i>	100%	250
		98%	320
		3%	78
Y=12	<i>PBD - Platform-based Development</i>	100%	345
		75%	210
		1%	66
Y=13	<i>PD - Parallel and Distributed Computing</i>	100%	250
		98%	320
		3%	78
Y=14	<i>Programming Languages (PL)</i>	100%	250
		98%	320
		3%	78
Y=15	<i>Software Development Fundamentals (SDF)</i>	100%	250
		98%	320
		3%	78
Y=16	<i>Software Engineering (SE)</i>	100%	250
		98%	320
		3%	78
Y=17	<i>SF - Systems Fundamentals</i>	100%	345
		75%	210
		1%	66
Y=18	<i>SP - Social Issues and Professional Issues</i>	100%	345
		75%	210
		1%	66

Tabel 11 di atas menjelaskan hasil rekomendasi dari jumlah mahasiswa sebanyak 11167 orang, di mana nilai rekomendasi terendah mencapai nilai 1 % dan nilai tertinggi 100 % serta rata-rata 67 %. Nilai 100% artinya bahwa seorang mahasiswa direkomendasikan suatu bidang skripsi dengan kadar 100%.

5. Kesimpulan

5.1 Kesimpulan

Adapun kesimpulan dari penelitian ini adalah:

1. Rekomendasi bidang skripsi dengan menggunakan algoritma *Expectation Maximization Clustering* dapat digunakan untuk merekomendasikan bidang skripsi pada mahasiswa dengan nilai rekomendasi hingga 100%.
2. Diharapkan metode ini dapat digunakan untuk membantu mahasiswa yang akan mengambil skripsi dengan memberikan sebuah rekomendasi mengenai bidang ilmu komputer mana yang sesuai dengan kompetensi mereka serta membantu kampus dalam memberikan bimbingan dan arahan kepada mahasiswa dalam hal memilih bidang skripsi yang sesuai dengan kompetensi mahasiswa.

5.2 Saran

1. Perlu pengembangan penelitian ini dengan metode yang berbeda sama sekali, seperti metode algoritma genetika atau teknik pembelajaran mesin lainnya.
2. Teknik ini dapat diimplementasikan di bidang lainnya.
3. Perlu melakukan evaluasi terhadap perhitungan yang telah dibuat.

6. Daftar Rujukan

- [1] Larose DT., 2013. *Discovering Knowledge In Data*. New York.Publisher: Wiley.
- [2] Acock, C. A., 2005. Working With Missing Values. *Journal Marriage Fam*, (67), pp. 1012–28.
- [3] Osama Abu Abbas., 2008. Comparisons Between Data Clustering Algorithms. *International Arab Journal Information And Technology*, (5), pp.320–5.
- [4] Narwati., 2010. Pengelompokan Mahasiswa Menggunakan Algoritma K-Means. *Jurnal Dinamika Informatika*, (2), pp. 1–7.
- [5] Liu Z, Zhang X.,2010. Prediction And Analysis For Students' Marks Based On Decision Tree Algorithm. *3rd International Conference Intelligent Networks Intelligent System*, ICINIS, pp.338–41.
- [6] Xu R cao, 2009. Improved C4.5 Algorithm For The Analysis Of Sales. Proceeding 6th Web Information System Application Conference. Publisher: WISA, pp.173–6.
- [7] Isnawati AF,2013. Studi P, Telekomunikasi T. Klasifikasi Modulasi Digital Menggunakan Kombinasi Teknik Fuzzy Clustering Dan Template Matching Sebagai Pengenalan. *INFOTEL*, (5), pp. 30–8.
- [8] Obbie Kristanto., 2015. Penerapan Algoritma Klasifikasi Data Mining Id3 Untuk Menentukan Penjurusan Siswa SMAN,*Konferensi nasional Sistem & informatika* , (6), pp. 6–7.
- [9] Chang W, Yang H., 2009. Applying IRT to Estimate Learning Ability And K-Means Clustering In Web Based Learning. *Journal of Software*, (4), pp.167–74.
- [10] Sumanto, Wahono RS.,2011. Penerapan Fuzzy C-Means Dalam Pemilihan. *Proceeding Seminar Nasional Inovasi Teknologi*, (1).
- [11] Asroni RA., 2015. Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang. *Ilmiah Semesta Teknika*, (18), pp.76–82.
- [12] Mulyani, 2015. Penerapan Data Mining Classification Untuk Prediksi Perilaku Pola Pembelian Terhadap Waktu Transaksi Menggunakan Metode Naïve Bayes. *Konferensi Nasional Sistem Informasi*, pp. 9–10.
- [13] Sutrisno, Afriyudi, Widiyanto., 2013. Penerapan Data Mining Pada Penjualan Menggunakan Metode Clustering Study Kasus Pt . Indomarco. *Journal Ilmiah Teknologi Informasi Ilmu Komputer*, (X), pp.1–11.
- [14] Sainani Kristin, L., 2015. Dealing With Missing Data. *PM & R*,(7), pp. 990–4.
- [15] Aptikom Tim.2016. Kerangka Kualifikasi Nasional Indonesia (KKNI). Depok. APTIKOM
- [16] Ridwan M, Suyono H, Sarosa M., 2013. Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *EECCIS*,(7), pp. 59–64.
- [17] Dellaert F.,2010. An Expectation-Maximization Algorithm For The Lasso Estimation of Quantitative Trait Locus Effects. *HEREDITY*, (105), pp. 483–94.