

Implementasi Teknik Data Mining untuk Evaluasi Kinerja Mahasiswa Berdasarkan Data Akademik

Gita Indah Marthasari^{1)*}

Teknik Informatika, Universitas Muhammadiyah Malang
gita@umm.ac.id^{1)*}

Abstrak

Salah satu indikator efisiensi perguruan tinggi adalah masa studi mahasiswa. Masa studi merupakan salah satu indikator penting dan menjadi salah satu komponen penilaian dalam akreditasi program studi. Penting bagi pengelola program studi untuk meningkatkan rasio mahasiswa yang lulus tepat waktu. Penelitian ini bertujuan menganalisis karakteristik yang mempengaruhi masa studi mahasiswa dari data akademik. Metode yang digunakan adalah association rule mining (ARM) dan clustering. Sebuah framework analisis berbasis ARM dan clustering diusulkan dalam penelitian ini. Metode ARM merupakan sebuah metode untuk menemukan aturan-aturan asosiasi yang memenuhi nilai minimum support dan minimum confidence. Algoritma yang digunakan adalah Apriori. Sedangkan clustering menggunakan algoritma Simple Expectation-Maximization (EM-clustering). Simple EM adalah algoritma berbasis model yang mencari nilai maximum likelihood estimation dalam model probabilitas. Variabel yang dianalisis adalah indeks prestasi mahasiswa, provinsi asal, dan asal sekolah. Analisis dilakukan menggunakan perangkat lunak WEKA. Penelitian dimulai dengan pengumpulan data dari sumber primer yaitu Biro Administrasi Akademik (BAA) Universitas Muhammadiyah Malang (UMM). Langkah selanjutnya adalah pembersihan dan transformasi data. Selanjutnya, dilakukan pencarian aturan menggunakan algoritma Apriori. Parameter yang diatur adalah nilai minimum support dan minimum confidence. Berikutnya dilakukan analisis cluster menggunakan Simple EM. Uji coba dilakukan untuk menemukan hasil clustering dengan nilai log likelihood terbesar. Berdasarkan pengujian, metode yang digunakan berhasil menggambarkan karakteristik berdasarkan masa studi.

Kata kunci: masa studi, algoritma apriori, algoritma simple expectation maximization, data mining untuk pendidikan, WEKA

Abstract

[Technical Implementation of Data Mining to Evaluate Student Performance Based on Academic Data] One indicator of college efficiency is the study period of the students. It is important for university managers to improve the ratio of students who graduate on time. This research aims to analyze the characteristics that affect the study period of students from academic data. The methods used are association rule mining (ARM) and clustering. We propose a framework to analyze academic data using ARM dan clustering method. ARM method is a method to find association rules that meet minimum support and minimum confidence. The algorithm used is Apriori. While clustering using Simple Expectation-Maximization (EM-clustering) algorithm. Simple EM is a model-based algorithm that searches for maximum likelihood estimation in the probability model. The variables analyzed were student achievement index, province of the students, and type of high school. The analysis is done using WEKA. Research begins with the collection of data from the primary source of the Biro Administrasi Akademik (BAA) Universitas Muhammadiyah Malang (UMM). Then, we do the data cleaning and transformation. Analyzing process is done in two step. First, do a rule search using Apriori algorithm. The regulated parameters are the minimum support and minimum confidence value. Second, we use Simple EM algorithm for the clustering process. The experiments were conducted to find the clustering result with the largest log likelihood value. Based on the experiment, the method used successfully describes the characteristics based on the study period.

Keywords: study period, apriori algorithm, simple expectation maximization algorithm, education data mining, WEKA

1. PENDAHULUAN

Pendidikan merupakan kunci kemajuan sebuah bangsa. Seluruh jenjang pendidikan berperan besar dalam hal itu termasuk pendidikan tinggi. Indonesia memiliki lebih dari 4000 perguruan tinggi yang tersebar pada 35 propinsi dengan jumlah mahasiswa mendekati angka 300.000 orang (Sumber: Forlap DIKTI). Universitas Muhammadiyah Malang (UMM) merupakan salah satu perguruan tinggi swasta di Indonesia yang memiliki mahasiswa dengan jumlah cukup besar yaitu 33.744 orang pada tahun 2015 (Sumber: Forlap DIKTI) yang tersebar dalam 52 program studi.

Pengelolaan yang profesional menjadi tuntutan yang harus dipenuhi oleh para pengelola perguruan tinggi. Meningkatnya persaingan dalam hal penyelenggaraan proses pendidikan di dunia pendidikan tinggi menjadi salah satu alasan perlunya pengelola untuk selalu memperhatikan fenomena-fenomena yang terjadi baik di lingkungan internal maupun eksternalnya. Perguruan tinggi dituntut untuk semakin efisien dalam pengelolaannya. Tuntutan ini juga dirasakan oleh pengelola program studi yang menjadi ujung tombak pelaksanaan proses belajar mengajar. Salah satu indikator efisiensi perguruan tinggi adalah masa studi mahasiswa. Masa studi merupakan salah satu indikator penting dan menjadi salah satu komponen penilaian dalam akreditasi program studi [1]. Berdasarkan data akademik mahasiswa program studi Teknik Informatika UMM, dari sekitar 570 data mahasiswa angkatan 2010 hingga 2012 terdapat lebih dari 400 orang (77,4%) yang masa studinya lebih dari 8 semester. Hal ini menjadi persoalan tersendiri bagi pengelola program studi yang selalu diharapkan untuk meningkatkan kinerja akademiknya.

Dalam beberapa tahun terakhir, UMM telah menggunakan sistem informasi yang memungkinkan penghimpunan data dalam jumlah besar, baik data pribadi mahasiswa, data hasil ujian calon mahasiswa baru, data keuangan, dan data hasil studi. Data dalam jumlah besar ini menjadi modal cukup penting untuk memperoleh pengetahuan-pengetahuan yang dapat menjawab pertanyaan-pertanyaan pengelola perguruan tinggi terkait permasalahan yang dihadapi. Salah satu teknik yang dapat digunakan adalah penambangan data (*data mining*). Teknik penambangan data mampu mengekstraksi pengetahuan yang tersembunyi dari data dimana pengetahuan tersebut selanjutnya disajikan dalam format yang mudah dimengerti oleh pengguna.

Metode penambangan data telah digunakan dalam analisis data pendidikan yang bidangnya dikenal dengan istilah *education data mining* (EDM). EDM menerapkan berbagai metode antara lain *decision tree*, algoritma apriori, k-means, neural network, dan metode lainnya. Luaran dari EDM dapat digunakan oleh para pemangku kepentingan di dunia pendidikan, termasuk pendidikan tinggi, dalam pengambilan

keputusan. Data yang dianalisis oleh EDM dapat berasal dari data historis dan operasional dari basis data perguruan tinggi, berupa data pribadi maupun akademik mahasiswa [2][3][4]. Selain itu, data dapat bersumber dari sistem e-learning yang digunakan di lingkungan perguruan tinggi [5][6][7].

Beberapa penelitian dalam bidang EDM menerapkan teknik *association rule mining* (ARM) dan klusterisasi terhadap data. Ahmed, dkk [3] menggunakan teknik ARM untuk menganalisis data pribadi dan akademik. Talavera, dkk [7] mengolah data dari sistem e-learning menggunakan teknik klusterisasi. Beberapa penelitian menggabungkan metode ARM dengan clustering untuk memperoleh pengetahuan yang lebih baik [2][8]. Pada penelitian Marthasari [9], digunakan metode ARM untuk mengidentifikasi faktor-faktor yang mempengaruhi status akademik mahasiswa berdasarkan data-data akademik yaitu jenis kelamin, usia saat mendaftar, kota asal, status domisili, agama, marital, asal sekolah, status kerja, asal biaya, pekerjaan orang tua, penghasilan orang tua, indeks prestasi semester 1 sampai dengan 4, dan jumlah SKS diambil pada semester 1 sampai dengan 4.

Analisis terhadap masa studi mahasiswa pernah dilakukan sebelumnya [10]–[12]. Pada [10] yang dilakukan adalah membuat model artificial neural network (ANN) dari data akademik dengan terlebih dahulu menentukan variabel prediktor yang mempengaruhi masa studi. Sedangkan pada [11] menggunakan variabel indeks prestasi per semester dan data induk mahasiswa sebagai prediktor masa studi untuk dianalisis menggunakan algoritma Naïve Bayes. Pada [12] data akademik yang meliputi data induk mahasiswa, nilai indeks prestasi per semester, dan jumlah SKS diolah menggunakan metode pohon keputusan C4.5. Belum ada penelitian yang menggunakan variabel domisili asal dan jenis sekolah asal mahasiswa sebagai prediktor. Selain itu, pada penelitian ini diusulkan sebuah framework untuk menganalisis data akademik menggunakan teknik aturan asosiasi dan *clustering*. Variabel domisili asal dianggap berpengaruh karena lingkungan dan asal daerah seseorang akan membentuk karakter dan perilakunya [13], [14]. Sedangkan variabel jenis sekolah asal menjadi latar belakang pendidikan seseorang mahasiswa yang dapat mempengaruhi kinerjanya di perguruan tinggi [15][15], [16].

2. BAHAN DAN METODE

Penelitian dalam bidang *Educational Data Mining* (EDM) telah banyak dilakukan. El-Haales [2] menggunakan beberapa metode data mining untuk menganalisis data antara lain *association rule mining*, *decision tree*, algoritma *clustering Expectation-Maximization* (EM), dan analisis *outlier*. Data yang dianalisis adalah data akademik dan data dari sistem e-learning. Teknik *association rule* digunakan untuk menemukan faktor-faktor yang mempengaruhi nilai

akhir mahasiswa. *Decision tree* digunakan untuk membentuk aturan logika yang dapat mengklasifikasikan nilai akhir mahasiswa. Sedangkan algoritma EM digunakan dalam proses klasterisasi mahasiswa berdasarkan beberapa variabel.

Talavera, dkk [7] menerapkan teknik klasterisasi dalam proses EDM untuk mengevaluasi proses pembelajaran berbasis kolaborasi. Data yang akan diklaster adalah diambil dari data statistik sistem e-learning yang merekam seluruh aktivitas siswa pada saat pembelajaran. Algoritma klasterisasi yang digunakan adalah algoritma berbasis model yaitu *Expectation-Maximization* karena kemampuannya untuk mengolah data diskrit. Melalui penelitian ini, dapat diperoleh kaitan antara aktivitas siswa dengan nilai akhir pada mata kuliah tertentu.

Ahmed, dkk [3] menggunakan teknik *association rule mining* untuk mempelajari profil mahasiswa. Data yang diolah meliputi data personal dan akademis. Proses-proses yang diterapkan yaitu membuat basis data universal, transformasi data, dan analisis menggunakan aplikasi WEKA. Berdasarkan analisis, diperoleh aturan-aturan penting antara lain korelasi variabel jenis kelamin, tempat tinggal, dan mekanisme penilaian terhadap nilai akhir.

Manvar, dkk [17] menerapkan EDM untuk mempelajari profil siswa. Algoritma yang digunakan adalah k-Means dan Apriori. Penerapan Algoritma Apriori bertujuan untuk memperkirakan hasil ujian akhir. Algoritma k-means digunakan untuk membagi siswa ke dalam kelompok-kelompok berdasarkan hasil ujian akhirnya.

Marthasari [9] melakukan analisis terhadap data pribadi dan akademik mahasiswa Jurusan Teknik Informatika UMM. Metode yang diterapkan adalah *association rule mining*. Beberapa variabel dianalisa untuk dicari korelasinya terhadap status akademik mahasiswa. Uji coba dilakukan dengan mengatur nilai *minimal support* dan *minimal confidence*.

Meinanda, dkk [10] melakukan penelitian untuk memprediksi masa studi mahasiswa. Data-data yang diolah yaitu masa studi, kode mata kuliah, nama mata kuliah, jumlah pengambilan mata kuliah tertentu, dan nilai mata kuliah. Data mentah tersebut kemudian dilakukan pra-proses antara lain menggunakan *cross-tabulation* untuk memperoleh data turunan yang dibutuhkan, dilanjutkan dengan penghapusan *missing value* dan *outlier*. Peneliti membuat hipotesis *a priori* yang menyatakan adanya hubungan antara Indeks Prestasi Kumulatif, jumlah mata kuliah yang diambil, dan jumlah mata kuliah mengulang. Data yang telah mengalami pra-proses selanjutnya dianalisis menggunakan salah satu model *Artificial Neural Network* (ANN) yaitu *Multi Layer Perceptron* (MLP).

Metode yang akan digunakan dalam penelitian ini adalah algoritma Apriori dan *Simple Expectation-Maximization* (EM). Apriori merupakan salah satu algoritma dalam salah satu teknik dalam *data mining* yaitu *association rule mining* (ARM). Metode ARM merupakan sebuah metode untuk menemukan aturan-

aturan asosiasi yang memenuhi nilai minimal *support* dan minimal *confidence* [18]. Mekanisme ARM terbagi menjadi dua: (1) menemukan himpunan item yang kemunculannya memenuhi nilai *support* tertentu dimana himpunan item ini selanjutnya disebut sebagai frequent item set, dan (2) membangkitkan aturan-aturan asosiasi dari seluruh himpunan frequent item set yang memenuhi nilai *confidence* tertentu. Apriori adalah algoritma yang diusulkan oleh R. Agarwal dan R. Srikant pada tahun 1994 [19]. Penamaan ini berkaitan dengan cara kerja dari Apriori yang sesuai dengan namanya, 'prior' yang artinya 'awal', memanfaatkan pengetahuan awal dari himpunan frequent item. Berikut adalah langkah-langkah Apriori:

Jika C_k adalah himpunan kandidat item berukuran k dan L_k adalah himpunan frequent item berukuran k , maka proses iterasi yang berlangsung adalah:

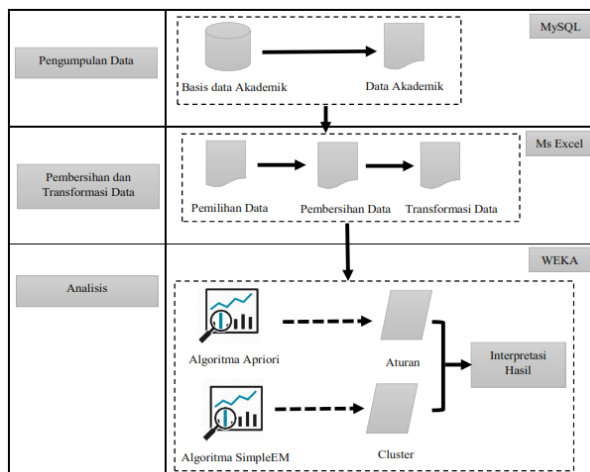
- Cari frequent set L_{k-1}
- Langkah penggabungan: C_k dibangkitkan dengan menggabungkan L_{k-1} dengan dirinya sendiri ($L_{k-1} \times L_{k-1}$)
- Langkah pemangkasan (properti Apriori) : setiap himpunan item berukuran $(k-1)$ yang kemunculannya tidak sering tidak dapat menjadi himpunan bagian dari himpunan item berukuran k yang kemunculannya sering, maka harus dihapus
- Frequent set L_k telah dihasilkan.

Clustering adalah cara untuk mengelompokkan data sehingga data-data dalam cluster yang sama akan serupa dan berbeda dengan data dalam cluster lainnya [2]. Algoritma *clustering Expectation-Maximization* (EM-*clustering*) adalah algoritma berbasis model yang mencari nilai *maximum likelihood estimation* dalam model probabilitas. Algoritma ini kerap digunakan untuk mengolah baik data numerik maupun diskrit [20]. Algoritma EM bersifat iteratif dan tersusun atas dua langkah: langkah *Expectation* (E) dan langkah *Maximization* (M). Secara umum, proses algoritma EM adalah sebagai berikut:

- a) Inisialisasi
Parameter awal dari distribusi k dipilih baik secara acak atau diberikan
- b) Iterasi
Langkah E: hitung nilai probabilitas $P(C_i|x)$ untuk seluruh data menggunakan nilai parameter distribusi. Ubah label tiap data berdasarkan hasil perhitungan probabilitas.
Langkah M: estimasi ulang parameter distribusi untuk memaksimalkan *likelihood* tiap data
- c) Kondisi berhenti
Ketika tercapai konvergen yaitu perubahan kecil terhadap log-likelihood tiap iterasi.

Framework penelitian secara lengkap dapat dilihat pada gambar 1. Penelitian dimulai dengan pengumpulan data dari sumber primer yaitu Biro

Administrasi Akademik (BAA) Universitas Muhammadiyah Malang (UMM). Data yang diterima dalam bentuk file .xls selanjutnya dipersiapkan (pra-proses) sebelum dianalisa. Pada proses penyiapan data dilakukan pembersihan dan transformasi data. Data yang telah mengalami pra-proses selanjutnya dianalisis menggunakan aplikasi WEKA. Dalam proses analisis dilakukan beberapa uji coba dengan mengatur nilai beberapa parameter. Untuk analisis menggunakan metode *association rule mining* (ARM) parameter yang diuji coba adalah nilai minimal *support* dan minimal *confidence*. Algoritma yang akan digunakan adalah Apriori. Luarannya adalah aturan-aturan. Aturan terkuat adalah aturan yang memiliki nilai *confidence* tinggi. Dari aturan-aturan kuat ini akan dicari aturan yang menarik. Aturan yang menarik adalah aturan yang memberikan pengetahuan baru dan penting kepada pengguna. Sedangkan untuk analisis menggunakan metode *clustering*, parameter yang diatur adalah jumlah *cluster*. Algoritma yang digunakan adalah Expectation-Maximization (EM). Hasil klusterisasi yang berupa pengelompokan data mahasiswa akan dianalisis untuk didapatkan deskripsi karakteristik mahasiswa per kluster.



Gambar 1 Framework Penelitian

3. HASIL DAN PEMBAHASAN

A. Pengumpulan Data (*Data Collection*)

Data mentah diperoleh dari basis data akademik mahasiswa Universitas Muhammadiyah Malang. Tabel-tabel sumber yang digunakan adalah tabel IN_KOTA, IN_SEKOLAH_MENENGAH, MASTER_SISWA dan TRANSAKSI_TUGAS_AKHIR. Selanjutnya beberapa query SQL dieksekusi untuk memperoleh data-data mahasiswa Jurusan Teknik Informatika tahun masuk 2010 hingga 2012 yang telah menyelesaikan studinya. Data-data ini kemudian diekspor ke format Microsoft Excel. Data akademik tersebut berisi beberapa variabel yaitu nomor induk mahasiswa (NIM), tahun ajar, semester ajar, judul, IPK, nama, kota asal, dan sekolah asal. Jumlah keseluruhan data adalah sebanyak 573 data.

Penjelasan mengenai masing-masing *field* dapat dilihat pada tabel 1. Sebelum dianalisis, akan dilakukan pembersihan dan transformasi data.

Tabel 1
DESKRIPSI VARIABEL DARI DATA MENTAH

Field	Deskripsi	Tipe Data
NIM	Nomor induk mahasiswa	String
Tahun ajar	Tahun ajaran dimana mahasiswa tersebut lulus	String
Semester Ajar	Semester lulus (1=ganjil, 2=genap)	Numerik
Judul	Judul tugas akhir	String
IPK	Indeks prestasi kumulatif	Numerik
Nama	Nama mahasiswa	String
Kota Asal	Kota asal mahasiswa berdasarkan KTP	String
Sekolah Asal	Sekolah asal (setaraf SMA)	String

B. Pembersihan dan Transformasi Data

Sebelum dianalisis, dilakukan pembersihan data terhadap data mentah karena beberapa alasan antara lain data tidak lengkap dan data yang salah. Contoh data tidak lengkap adalah data yang tidak memiliki nilai IPK. Sedangkan contoh data yang salah adalah data IPK yang tertulis dalam variabel judul tugas akhir. Terdapat 3 item data yang termasuk dalam kategori data salah atau data tidak lengkap. Ketiga data ini selanjutnya dihapus dari keseluruhan data untuk menghindari hasil analisis yang tidak akurat. Setelah proses penghapusan, jumlah keseluruhan data sebanyak 571 item.

Selanjutnya ditentukan variabel-variabel untuk dianalisis. Peneliti menentukan secara *a priori* variabel-variabel yang berpotensi menjadi prediktor variabel masa studi yaitu predikat IPK, jenis sekolah asal, dan asal provinsi. Langkah selanjutnya adalah transformasi data. Variabel-variabel yang ditransformasi adalah status masa studi, predikat IPK, jenis sekolah asal, dan wilayah domisili. Variabel pertama yaitu status masa studi diturunkan dari variabel NIM, semester, dan tahun lulus. Dari variabel NIM diperoleh tahun masuk yang selanjutnya dihitung lama masa studi dalam semester berdasarkan semester dan tahun lulusnya. Nilai masa studi dalam bentuk numerik ini kemudian dikategorisasikan menggunakan aturan seperti pada tabel 2. Variabel kedua, predikat IPK, diperoleh dengan mengkategorisasikan nilai variabel IPK menggunakan aturan pada tabel 3. Variabel ketiga adalah jenis sekolah asal yang diturunkan dari variabel sekolah asal. Terakhir, variabel asal provinsi yang diturunkan dari variabel kota asal.

Tabel 2
ATURAN KATEGORISASI MASA STUDI

Masa Studi	Status Masa Studi
masa_studi \leq 8 semester	TRUE (tepat waktu)
masa_studi $>$ 8 semester	FALSE

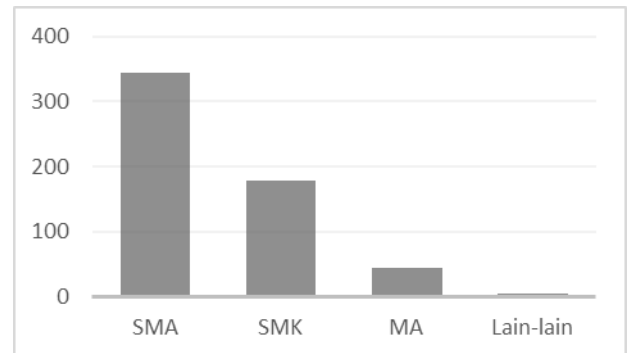
Tabel 3
ATURAN KATEGORISASI IPK

IPK	Predikat IPK
$3.5 < ipk$	Dengan pujian (DP)
$2.75 < ipk \leq 3.5$	Sangat memuaskan (SM)
$ipk \leq 2.75$	Memuaskan (M)

Dari data yang diperoleh, jumlah mahasiswa yang lulus tepat waktu adalah 128 orang (22,4%), sedangkan sebanyak 443 orang sisanya (77,6%) lulus lebih dari 8 semester. Untuk variabel predikat IPK, terdapat 327 data (57.3%) yang memiliki predikat IPK sangat memuaskan (SM), 229 data (40.1%) memiliki predikat dengan pujian (DP), dan 15 data (2.6%) berpredikat memuaskan (M). Untuk variabel jenis sekolah asal mayoritas adalah dari SMA sebanyak 344 data (60.2%), dari SMK sejumlah 178 data (31.2%), dari MA sebanyak 44 data (7,7%), dan lain-lain sebanyak 5 data (0.9%). Komposisi penyebaran data berdasarkan jenis sekolah asal ditunjukkan pada gambar 2. Sedangkan penyebaran data berdasarkan lokasi provinsi dapat dilihat pada tabel 4 dimana sebagian besar data memiliki provinsi asal Jawa Timur (66%).

Tabel 4
KOMPOSISI DATA BERDASARKAN PROVINSI ASAL

Provinsi	Jumlah	%
Kalimantan Timur	41	7.180385289
Kalimantan Barat	4	0.700525394
NTB	31	5.429071804
Kalimantan Selatan	49	8.581436077
Sulawesi Tengah	1	0.175131349
Jawa Timur	377	66.02451839
Kalimantan Tengah	12	2.101576182
Kalimantan Utara	6	1.050788091
Maluku	11	1.926444834
Aceh	4	0.700525394
Sulawesi Selatan	3	0.525394046
Papua	3	0.525394046
Banten	1	0.175131349
Bali	6	1.050788091
Jawa Tengah	4	0.700525394
Sulawesi Tenggara	3	0.525394046
NTT	3	0.525394046
Jawa Barat	5	0.875656743
Sumatera Selatan	2	0.350262697
Lampung	1	0.175131349
Riau	2	0.350262697
DKI	2	0.350262697



Gbr 2 Komposisi data berdasarkan asal sekolah

C. Analisis menggunakan Algoritma Apriori

Analisis menggunakan teknik aturan asosiasi bertujuan menemukan variabel-variabel yang berkaitan dengan masa studi. Pada penelitian ini algoritma yang digunakan adalah Apriori. Akan dilakukan 3 kali uji coba dengan mengubah nilai *minimum confidence*. Ketiga pengujian menggunakan nilai *minimum support* = 0.1. Pada pengujian pertama, nilai *minimum confidence* (min_conf) diatur sebesar 0.9 dan pengujian kedua diubah menjadi 0.7. Pada pengujian terakhir, nilai *minimum confidence* diturunkan menjadi 0.5 dimana hasilnya dapat dilihat pada tabel 5. Meskipun aturan yang kuat ditunjukkan oleh tingginya nilai *minimum confidence*, namun pada penelitian ini ingin dicari aturan yang menarik dan bermanfaat.

Tabel 5
HASIL UJI COBA DENGAN NILAI MIN_CONF = 0.5

No.	Aturan	Nilai Confidence
1	PredikatIPK=SM Provinsi=JATIM Sekolah=SMK ==> Masa Studi=FALSE	0.95
2	PredikatIPK=SM Sekolah=SMK ==> Masa Studi=FALSE	0.89
3	PredikatIPK=SM Provinsi=JATIM ==> Masa Studi=FALSE	0.88
4	PredikatIPK=SM ==> Masa Studi=FALSE	0.85
5	PredikatIPK=SM Provinsi=JATIM Sekolah=SMA ==> Masa Studi=FALSE	0.85
6	PredikatIPK=SM Sekolah=SMA ==> Masa Studi=FALSE	0.83
7	Provinsi=JATIM Sekolah=SMK ==> Masa Studi=FALSE	0.82
8	Sekolah=SMK ==> Masa Studi=FALSE	0.79
9	Provinsi=JATIM ==> Masa Studi=FALSE	0.79
10	Sekolah=SMA ==> Masa Studi=FALSE	0.76
11	Masa Studi=FALSE PredikatIPK=SM Sekolah=SMK ==> Provinsi=JATIM	0.76

12	Masa Studi=FALSE Sekolah=SMK ==> Provinsi=JATIM	0.76
13	Provinsi=JATIM Sekolah=SMA ==> Masa Studi=FALSE	0.76
14	Sekolah=SMK ==> Provinsi=JATIM	0.73
15	PredikatIPK=SM Sekolah=SMK ==> Provinsi=JATIM	0.72
16	Masa Studi=FALSE PredikatIPK=SM ==> Provinsi=JATIM	0.68
17	PredikatIPK=SM Sekolah=SMK ==> Masa Studi=FALSE Provinsi=JATIM	0.68
18	PredikatIPK=DP Sekolah=SMA ==> Masa Studi=FALSE	0.68
19	PredikatIPK=DP ==> Masa Studi=FALSE	0.68
20	Masa Studi=FALSE ==> Provinsi=JATIM	0.67
21	PredikatIPK=DP ==> Provinsi=JATIM	0.67
22	PredikatIPK=DP Provinsi=JATIM ==> Masa Studi=FALSE	0.67
23	Masa Studi=FALSE PredikatIPK=DP ==> Provinsi=JATIM	0.66
24	PredikatIPK=SM ==> Provinsi=JATIM	0.66
25	Masa Studi=FALSE Provinsi=JATIM Sekolah=SMK ==> PredikatIPK=SM	0.65
26	Masa Studi=FALSE Sekolah=SMK ==> PredikatIPK=SM	0.65
27	Masa Studi=FALSE Provinsi=JATIM Sekolah=SMA ==> PredikatIPK=SM	0.65
28	Masa Studi=FALSE PredikatIPK=SM Sekolah=SMA ==> Provinsi=JATIM	0.65
29	Masa Studi=FALSE Provinsi=JATIM ==> PredikatIPK=SM	0.64
30	Masa Studi=TRUE ==> Provinsi=JATIM	0.63
31	Masa Studi=TRUE ==> Sekolah=SMA	0.63
32	PredikatIPK=SM Sekolah=SMA ==> Provinsi=JATIM	0.63
33	Masa Studi=FALSE ==> PredikatIPK=SM	0.63
34	Sekolah=SMA ==> Provinsi=JATIM	0.63
35	Masa Studi=FALSE Sekolah=SMA ==> PredikatIPK=SM	0.62

Dari pengujian yang telah dilakukan, diperoleh sejumlah aturan. Dari sejumlah aturan tersebut dipilih aturan-aturan yang menarik yaitu yang menunjukkan variabel-variabel yang berasosiasi dengan masa studi. Diperoleh beberapa variabel yang berasosiasi dengan masa studi baik yang tepat waktu maupun yang tidak. Sesuai dengan aturan transformasi data, masa studi = TRUE artinya mahasiswa lulus tepat waktu,

sedangkan nilai FALSE berarti sebaliknya. Pertama, dari aturan-aturan yang dihasilkan oleh algoritma Apriori, variabel yang berasosiasi dengan masa studi tidak tepat waktu adalah predikat IPK = Sangat Memuaskan, asal provinsi = Jawa Timur, dan jenis sekolah = SMK (nilai *confidence* = 0.95). Kedua, variabel yang berasosiasi dengan masa studi tepat waktu adalah predikat IPK = Dengan Pujian, asal provinsi = Jawa Timur, dan jenis sekolah = SMA (nilai *confidence* = 0.63). Dari hasil analisis menggunakan algoritma Apriori, diperoleh pengetahuan hubungan antara jenis sekolah asal dengan predikat lulusan. Hal ini menunjukkan adanya perbedaan kemampuan dasar mahasiswa yang berasal dari SMA dengan SMK mengingat secara kurikulum mata pelajaran di SMA bersifat lebih teoritis sedangkan di SMK lebih praktis sehingga siswa SMK cenderung memiliki tingkat penalaran yang rendah [15]. Oleh karena itu, dapat dinyatakan bahwa variabel jenis sekolah asal mahasiswa (SMA/SMK) turut mempengaruhi masa studi di universitas. Hasil ini yang membedakan dengan penelitian sebelumnya [3].

D. Analisis menggunakan Algoritma Simple EM

Dalam EM, model yang paling merepresentasikan data ditunjukkan dengan nilai *log likelihood* yang besar. Untuk itu dilakukan beberapa kali uji coba dengan mengubah jumlah *cluster* dari 1 hingga 10. Dari kelima pengujian akan ditentukan model *cluster* yang memiliki nilai *log likelihood* terbesar. Uji coba dilakukan menggunakan perangkat WEKA. Parameter yang digunakan adalah *default* dengan hanya mengubah variabel *numClusters* menjadi berturut-turut 1, 2, 3, 4, sampai dengan 6. Hasil pengujian ditunjukkan pada tabel 9. Berdasarkan hasil tersebut, pengujian dengan jumlah *cluster* = 4 memiliki nilai *log likelihood* terbesar yaitu -3.6164. Berdasarkan hasil *clustering*, diperoleh beberapa gambaran karakteristik mahasiswa antara lain sebagai berikut:

- Mayoritas data *instance* yang memiliki masa studi lebih dari 8 semester berada pada *cluster* 2 dan 4 sedangkan, sebaliknya, *cluster* 1 dan 3 mayoritas berisi data mahasiswa dengan masa studi kurang dari sama dengan 8 semester.
- Pengelompokan mahasiswa berdasarkan predikat IPK juga mengikuti kecenderungan yang sama dengan pemetaan berdasar masa studi, dimana pada *cluster* 1 dan 3 mayoritas memiliki predikat 'Dengan Pujian', sedangkan *cluster* 2 dan 4 mayoritas berpredikat 'Sangat Memuaskan'. Hasil ini memberikan gambaran bahwa ada kecenderungan mahasiswa yang lulus tepat waktu memiliki IPK dengan predikat "Dengan Pujian".
- Berdasarkan pengamatan pada variabel 'Provinsi', beberapa provinsi memiliki kecenderungan berada di *cluster* tertentu seperti ditunjukkan dalam tabel 7. Dari tabel tersebut,

mahasiswa yang lulus tepat waktu sebagian besar berasal dari Kalimantan Barat, Kalimantan Selatan, Bali, dan Jawa Tengah.

Tabel 6
HASIL UJI COBA UNTUK SIMPLE EM

Jumlah Cluster	Log likelihood
1	-3.66551
2	-3.63498
3	-3.63574
4	-3.6164
5	-3.62858
6	-3.635
7	-3.63616
8	-3.63607
9	-3.63599
10	-3.64115

Tabel 7
PEMETAAN CLUSTER UNTUK VARIABEL ASAL PROVINSI

Provinsi	Cenderung ke Cluster
Kalimantan Timur	4
Kalimantan Barat	1
NTB	4
Kalimantan Selatan	3
Sulawesi Tengah	3
Jawa Timur	4
Kalimantan Tengah	2
Kalimantan Utara	2
Maluku	4
Maluku	4
Sulawesi Selatan	4
Papua	4
Banten	2
Bali	1
Jawa Tengah	3
Sulawesi Tenggara	2
NTT	2
Jawa Barat	4
Lampung	2
Riau	4

4. KESIMPULAN

Penelitian ini bertujuan menganalisis karakteristik mahasiswa berdasarkan masa studi menggunakan dua teknik dalam data mining yaitu *association rule mining* (ARM) dan *clustering*. Hasil yang didapatkan dari proses ARM adalah meskipun seluruh variabel yang mempengaruhi masa studi ditentukan secara *a priori*, dapat ditunjukkan bahwa ada hubungan asosiatif antara variabel masa studi dengan variabel predikat IPK, provinsi asal, dan jenis sekolah asal. Sedangkan hasil yang diperoleh dari implementasi teknik *clustering* mampu memberikan gambaran karakteristik mahasiswa berkaitan dengan masa studi, provinsi asal, dan jenis sekolah asal. Hasil uji coba menggunakan kedua teknik data mining ini memberikan gambaran yang saling mendukung mengenai profil mahasiswa berdasarkan masa studi.

Untuk penelitian selanjutnya, dapat ditambahkan variabel-variabel lain antara lain jumlah mata kuliah mengulang, jumlah mata kuliah yang diambil per semester, dan nilai ujian masuk universitas.

5. DAFTAR PUSTAKA

- [1] B. A. N. Indonesia, "Peraturan Badan Akreditasi Nasional PT No 4/2017," 2017.
- [2] A. El-Halees, "Mining Students Data To Analyze Learning Behavior: a Case Study Educational Systems," *Work*, no. February, 2008.
- [3] S. Ahmed, R. Paul, A. Sayed, and L. Hoque, "Knowledge Discovery from Academic Data using Association Rule Mining," no. December, pp. 22–23, 2014.
- [4] M. I. Al-Twijri and A. Y. Noaman, "A New Data Mining Model Adopted for Higher Institutions," in *Procedia Computer Science*, 2015, vol. 65, pp. 836–844.
- [5] S. Baher and L. L.M.R.J., "Data Preparation Strategy in E-Learning System using Association Rule Algorithm," *Int. J. Comput. Appl.*, vol. 41, no. 3, pp. 35–40, 2012.
- [6] K. Kularbphetpong and C. Tongsir, "Mining Educational Data to Analyze the Student Motivation Behavior," *World Acad. Sci. Eng. Technol.*, vol. 6, no. 8, pp. 1036–1040, 2012.
- [7] L. Talavera and E. Gaudioso, "Mining student data to characterize similar behavior groups in unstructured collaboration spaces," *Work. Artif. Intell. CSCL. 16th Eur. Conf. Artif. Intell.*, pp. 17–23, 2004.
- [8] M. M. A. Tair and A. M. El-Halees, "Mining Educational Data to Improve Students' Performance: A Case Study," *Int. J. Inf. Commun. Technol. Res.*, vol. 2, no. 2, pp. 140–146, 2012.
- [9] G. I. Marthasari, "Identifikasi Faktor Ketidaktifan Mahasiswa menggunakan Teknik Data Mining," 2016.
- [10] M. H. Meinanda, M. Annisa, N. Muhandri, and K. Suryadi, "Prediksi Masa Studi Sarjana dengan Artificial Neural Network," *Internetworking Indones. J.*, vol. 1, no. 2, pp. 31–35, 2009.
- [11] A. Jananto, "Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa," *J. Teknol. Inf. Din.*, vol. 18, no. 1, pp. 9–16, 2013.
- [12] S. Haryati, A. Sudarsono, and E. Suryana, "IMPLEMENTASI DATA MINING UNTUK MEMREDIKSI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA C4 . 5," *Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.
- [13] S. H. Hasan, N. Aziz, and M. Adam, "PENGARUH LINGKUNGAN KERJA TERHADAP KINERJA AKTIVIS PADA

- LEMBAGA SWADAYA MASYARAKAT DI KOTA BANDA ACEH,” *Ilmu Manaj.*, vol. 1, no. 1, pp. 2–22, 2012.
- [14] W. E. Pratiwi, “PENGARUH BUDAYA JAWA TERHADAP ASERTIVITAS PADA REMAJA SISWA KELAS XDI SMA NEGERI 3,” *Psikologi*, vol. 3, no. 1, pp. 348–357, 2015.
- [15] N. Yuliana, “Pengaruh Pendekatan Differentiated Instruction (DI) Terhadap Kecemasan Matematika (Match Anxiety), Peningkatan Kemampuan Pemahaman Dan Penalaran Matematis Siswa SMK,” Universitas Pendidikan Indonesia, 2013.
- [16] N. Azizah, “Perilaku Moral dan Religiusitas Siswa Berlatar Belakang Pendidikan Umum dan Agama,” *Psikologi*, vol. 33, no. 2, pp. 1–8, 2003.
- [17] M. Manvar and M. Rao, “Predicting students performance in higher education: A Data Mining Approach,” *Ijser.Org*, vol. 5, no. 2, pp. 1024–1027, 2014.
- [18] S. Kotsiantis and D. Kanellopoulos, “Association Rules Mining: A Recent Overview,” *GESTS Int. Trans. Comput. Sci. Eng.*, vol. 32, no. 1, pp. 71–82, 2006.
- [19] V. Kumar and A. Chadha, “Mining association rules in student’s assessment data,” *Int. J. Comput. Sci. Issues*, vol. 9, no. 5, pp. 211–216, 2012.
- [20] M. Meila and D. Heckerman, “An Experimental Comparison of Model-Based Clustering Method,” *Mach. Learn.*, vol. 1225, no. January 2001, pp. 41–42, 2014.