

**Laporan Tugas Akhir MK. Data Mining (KOM332), Semester Genap 2014/2015**

***Penerapan Data Mining Menggunakan Teknik Association Rule Pada Keanggotaan BEM FMIPA IPB***

KYULENE BUNLYA FALLAH (G64120022), WIDURI WULANSARI PUTRI (G64120047), RAMDHAN ABDUL GHIFARI (G64120117)

Kelompok: 12, Lab: 1

**PENDAHULUAN**

**Latar Belakang**

Organisasi merupakan sekumpulan orang-orang yang bekerja sama untuk mencapai suatu tujuan bersama. Menurut Peraturan Pemerintah Republik Indonesia No. 60 Tahun 1999, organisasi kemahasiswaan adalah suatu wadah yang dibentuk untuk melaksanakan peningkatan kepemimpinan, penalaran, minat, kegemaran, dan kesejahteraan mahasiswa dalam kehidupan kemahasiswaan di perguruan tinggi. Berikutnya, organisasi kemahasiswaan dalam kehidupan intra perguruan tinggi juga dipahami sebagai wahana dan sarana pengembangan diri mahasiswa ke arah perluasan wawasan dan peningkatan kecendekiawanan serta integritas kepribadian untuk mencapai tujuan pendidikan tinggi (Surat Keputusan Mendikbud No. 155/U/1998, pasal 1 ayat 1).

Ada beberapa bentuk organisasi mahasiswa di kampus, diantaranya dapat di golongan menjadi dua yaitu: organisasi intra kampus seperti Senat Mahasiswa atau saat ini lebih dikenal dengan Badan Eksekutif Mahasiswa (BEM), Unit Kegiatan Mahasiswa (UKM), Himpunan Mahasiswa Jurusan/Program Studi, dan organisasi ekstra kampus seperti HMI, GMNI, GMKI, PMKRI, PMII, KAMMI, dan sejenisnya. Dengan bervariasinya bentuk organisasi tersebut mahasiswa memiliki keleluasaan untuk dapat memilih organisasi mana yang sesuai dengan minat dan bakat mereka masing-masing.

Setiap tahun pastinya sebuah organisasi melakukan regenerasi untuk melakukan pembaruan anggota maupun sistem kepemimpinan untuk menjamin kelangsungan organisasi tersebut. Tidak dapat dipungkiri jika Badan Eksekutif Mahasiswa merupakan salah satu organisasi yang menyedot banyak perhatian bagi mahasiswa yang ingin mengikuti sebuah organisasi di tingkat perguruan tinggi. Melalui penelitian ini, kelompok kami ingin melihat aturan-aturan yang dapat memperlihatkan pola keikutsertaan mahasiswa FMIPA IPB dalam BEM FMIPA IPB.

## **Tujuan**

Tujuan dari tugas akhir yang berjudul “Penerapan Data Mining Menggunakan Teknik Aturan Asosiasi dengan Algoritma Apriori Pada Keanggotaan BEM FMIPA IPB” adalah :

1. Menggunakan teknik data mining *association rule mining* dengan untuk mengetahui pola keikutsertaan mahasiswa FMIPA IPB dalam Badan Eksekutif Mahasiswa FMIPA (BEM FMIPA) berdasarkan jenis kelamin, departemen dan semester ketika masuk sebagai anggota.

## **Ruang Lingkup**

Ruang lingkup tugas akhir ini difokuskan pada :

1. Data seluruh anggota BEM FMIPA IPB pada tahun 2013-2015 yang didapatkan dari database anggota yang dimiliki sekretaris umum setiap kabinet.
2. Atribut yang digunakan, yaitu kode departemen, jenis kelamin dan semester.
3. Penerapan teknik data mining *association rule mining*.

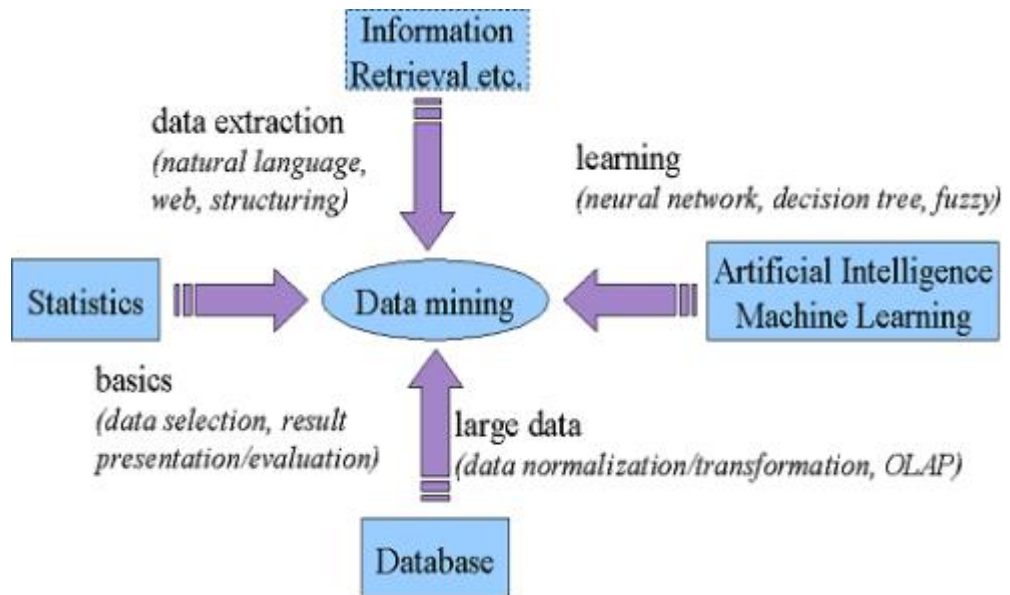
## **Manfaat**

Hasil asosiasi yang diperoleh diharapkan dapat digunakan untuk memrediksi seseorang berdasarkan departemen, jenis kelamin, dan semester untuk masuk ke dalam BEM FMIPA dengan melihat nilai confidence.

## **TINJAUAN PUSTAKA**

### ***Data Mining***

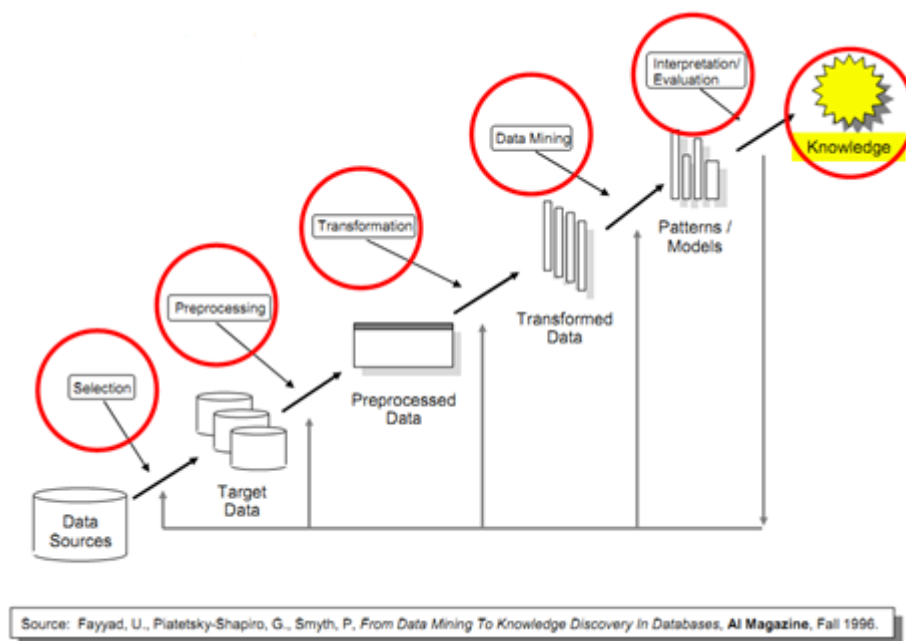
Data Mining (penggalian data) merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data (Larose 2005). Hubungan yang dicari dalam data mining dapat berupa hubungan antara dua atau lebih dalam satu dimensi, misalnya dalam dimensi produk, kita dapat melihat keterkaitan pembelian suatu produk dengan produk yang lain. Selain itu hubungan juga dapat dilihat antara 2 atau lebih atribut dan 2 atau lebih obyek (Pramudiono 2006). Data mining adalah bagian dari proses KDD (Knowledge Discovery in Databases) yang terdiri dari beberapa tahapan seperti pemilihan data, pra pengolahan, transformasi, data mining, dan evaluasi hasil (Maimon dan Last 2000).



Gambar 1. Bidang Ilmu Data Mining

### ***Knowledge Discovery in Database***

Knowledge Discovery in Database atau KDD secara umum juga dikenal sebagai pangkalan data. Di bawah ini merupakan gambar dari tahapan KDD :



Gambar 2. Tahapan KDD

## ***Association Rule Mining***

Analisis asosiasi atau *association rule mining* adalah teknik mining untuk menemukan aturan asosiatif antara suatu kombinasi item dalam suatu data set yang ditentukan (Han 2004). Contoh dari aturan asosiatif dari analisa pembelian di suatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seorang pelanggan membeli roti bersamaan dengan susu. Dengan pengetahuan tersebut pemilik pasar swalayan dapat mengatur penempatan barangnya atau merancang kampanye pemasaran dengan memakai kupon diskon untuk kombinasi barang tertentu. Karena analisis asosiasi menjadi terkenal karena aplikasinya untuk menganalisa keranjang belanja di pasar swalayan, analisis asosiasi juga sering disebut dengan istilah *market basket analysis* (Berry dan Linoff 2004)

Aturan asosiasi berkenaan dengan studi tentang “apa bersama apa”. Aturan asosiasi juga ingin memberikan informasi tersebut dalam bentuk hubungan “*if-then*” atau “jika-maka”. Aturan ini dihitung dari data yang sifatnya probabilistik (Santoso 2007: 225-226).

Association Rule Mining meliputi dua tahap:

- a. Mencari kombinasi yang paling sering terjadi dari suatu itemset (*frequent itemset*).
- b. *Generate Association Rule* dari *frequent itemset* yang telah dibuat sebelumnya.

Umumnya ada dua ukuran kepercayaan (*interestingness measure*) yang digunakan dalam menentukan suatu association rule, yaitu: (Han 2004)

- a. *Support*, suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu item atau itemset dari keseluruhan transaksi.
- b. *Confidence*, suatu ukuran yang menunjukkan hubungan antar dua item secara *conditional*.

## ***Apriori Algorithm***

Algoritma Apriori adalah algoritma paling terkenal untuk menemukan pola frekuensi tinggi. Pola frekuensi tinggi adalah pola-pola item di dalam suatu database yang memiliki frekuensi atau support di atas ambang batas tertentu yang disebut dengan istilah minimum support.

Pola frekuensi tinggi ini digunakan untuk menyusun aturan asosiatif dan juga beberapa teknik data mining lainnya. Walaupun akhir-akhir ini dikembangkan banyak algoritma yang lebih efisien dari Apriori seperti FP-growth, LCM dsb, tetapi Apriori tetap menjadi algoritma yang paling banyak diimplementasikan dalam produk komersial untuk data mining karena dianggap algoritma yang paling mapan.

## METODE

### Data

Data yang digunakan dalam tugas akhir ini adalah Data Anggota BEM FMIPA IPB Tahun 2013- 2015.

### Tahapan Kegiatan

Tahapan kegiatan mengacu pada proses knowledge discovery in database (KDD). Proses KDD secara garis besar dapat dijelaskan sebagai berikut (Fayyad 1996).

1. Data Selection

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. Pre-processing/Cleaning

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuat duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses enrichment, yaitu proses "memperkaya" data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

3. Transformation

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

Transformation atau transformasi adalah proses menggabungkan atau mengubah format data sesuai dengan proses yang digunakan dalam data mining. Karena beberapa format data mining membutuhkan format data khusus dalam pemrosesannya.

4. Data mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat tergantung pada tujuan dan proses KDD secara keseluruhan.

Proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik *association rule mining* pada data anggota BEM FMIPA IPB 2015.

#### 5. Interpretation/ Evaluation

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

Tahap ini merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Termasuk dalam langkah akhir dari data mining adalah mengimplementasikan analisis yang didapat. Sehingga akan memperoleh kesimpulan real (Han 2006).

### **Lingkungan Pengembangan**

Perangkat lunak yang digunakan dalam penelitian ini adalah:

- Sistem operasi Microsoft Windows 8 (32-bit)
- WEKA 3.6 sebagai proses pengolahan data
- Microsoft Excel 2013 untuk mengolah data sistem lahan dalam format .xlsx

Perangkat keras yang digunakan dalam penelitian ini computer personal dengan spesifikasi:

- Prosesor : Intel® Core™ i3 CPU 350M @2.26 GHz
- Memory : 5 GB RAM
- VGA : Intel® HD Graphics (Core i3)

## **HASIL DAN PEMBAHASAN**

### **Data Selection**

Data mentah yang diperoleh dalam penelitian ini masih belum mengandung beberapa atribut yang diperlukan dan mengandung atribut yang tidak diperlukan dalam penelitian. Sehingga diperlukan pra-proses data untuk mendapatkan dan mengisi atribut yang diperlukan juga menghapus atribut yang tidak diperlukan tersebut. Dari tahapan pra-proses data inilah diperoleh jumlah 12 atribut yaitu kode departemen dari G1 sampai G8, jenis kelamin berupa P dan L, serta semester 4 dan semester 6. Selain itu didapatkan pula 275 records data dari 276. Satu

record data yang telah dihapus secara manual adalah data yang tidak dapat diakuisisi nilainya.

### **Pre-processing/Cleaning**

Untuk tahapan keseluruhan tahapan ini dilakukan secara manual dalam microsoft excel, melihat jumlah data dalam data set yang kami miliki juga tidak terlalu banyak.

- Data cleaning

Data cleaning dilakukan untuk mengisi data yang terdapat missing value pada atribut jenis kelamin dan NIM, sehingga data dapat diolah dengan baik. Pengisian missing value untuk atribut jenis kelamin yang tadinya tidak ada dengan melihat nama dan memastikan kebenarannya melalui wawancara. Untuk mengatasi missing value pada atribut NIM dilakukan pencarian dengan kata kunci nama lengkap. Sedangkan untuk yang NIMnya tidak dapat diketahui setelah dilakukan berbagai pencarian, diputuskan untuk dihilangkan satu record datanya. Menambahkan atribut semester masuk bem yang ditentukan berdasarkan tahun kepengurusan dan NIM. Contoh : Jika seseorang dengan NIM berkode tahun 12 dan menjadi anggota BEM di tahun 2015 maka orang tersebut masuk BEM di semester 6.

- Data integration

Data integration dilakukan untuk menggabungkan seluruh data dari tahun 2013 hingga 2015. Karena data yang didapat tidak dalam satu file, sehingga data harus digabung terlebih dahulu.

- Data transformation

Data transformation yang dilakukan adalah normalisasi data. Terdapat duplikasi pada data setelah digabung seluruhnya, sehingga record data yang terduplikasi tersebut harus dihapus salah satu agar tidak terjadi kesalahan.

- Data reduction

Data reduction tidak dilakukan, karena dilihat dataset yang didapat sudah cukup, yakni sebesar 276 dengan 3 atribut, NIM (hanya diambil 2 karakter awal sebagai ID departemen), jenis kelamin dan semester masuk BEM.

- Data discretization

Data discretization tidak dilakukan karena dataset yang didapat sudah berupa data diskret.

## Transformation

Data ditransformasi ke bentuk yang sesuai untuk proses association rule mining. Pada tahap ini, dilakukan konversi format dari file.xlsx kemudian di konversi menjadi file.arff agar dapat diolah dalam Weka.

	A	B	C	D	E
1	No	Departemen	NIM	Jenis Kelamin	Semester
2	1	G6	G64100004	L	6
3	2	G8	G84110004	L	4
4	3	G5	G54100096	P	6
5	4	G2	G24110008	P	4
6	5	G3	G34100017	P	6
7	6	G8	G84110055	P	4
8	7	G5	G54100011	L	6
9	8	G8	G84100049	P	6
10	9	G8	G84100018	L	6
11	10	G1	G14110044	P	4
12	11	G2	G24110012	P	4
13	12	G4	G44110065	P	4
14	13	G1	G14110041	P	4
15	14	G5	G54110062	P	4
16	15	G2	G24110026	L	4
17	16	G6	G64110011	L	4
18	17	G6	G64110013	L	4
19	18	G6	G64110065	L	4
20	19	G7	G74100045	P	6
21	20	G1	G141100007	P	6

Gambar 3. Data dalam format xlsx

```

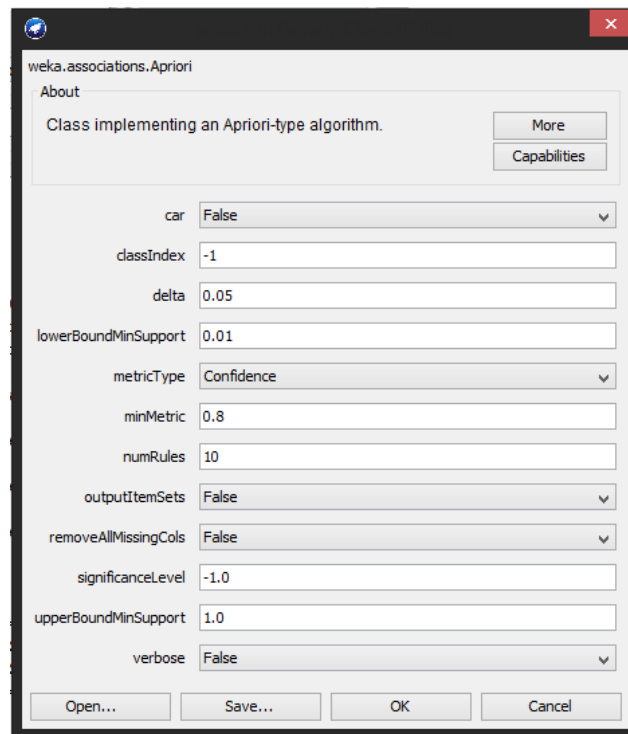
1 @relation 'BEM_FMIPA'
2
3 @attribute G1 {t}
4 @attribute G2 {t}
5 @attribute G3 {t}
6 @attribute G4 {t}
7 @attribute G5 {t}
8 @attribute G6 {t}
9 @attribute G7 {t}
10 @attribute G8 {t}
11 @attribute P {t}
12 @attribute L {t}
13 @attribute Sem4 {t}
14 @attribute Sem6 {t}
15
16 @data
17
18 ?,?,?,t,?,t,?,t
19 ?,?,t,?,t,?,t,?,t
20 ?,t,?,t,?,t,?,t,?,t
21 ?,t,?,t,?,t,?,t,?,t
22 ?,t,?,t,?,t,?,t,?,t
23 ?,t,?,t,?,t,?,t,?,t
24 ?,t,?,t,?,t,?,t,?,t
25 ?,t,?,t,?,t,?,t,?,t
26 ?,t,?,t,?,t,?,t,?,t
27 t,?,t,?,t,?,t,?,t,?

```

Gambar 4. Data ditransformasi menjadi .arff



## Data mining



Gambar 5. Pemilihan min.sup dan min.cof

Dalam asosiasi yang kami lakukan, diambil nilai minimum support sebesar 0,01 dan minimum confidence sebesar 0,8. Karena jika diambil nilai minimum support sebesar 1, tidak ada rule yang dihasilkan, dan jika diambil nilai minimum confidence 1, tidak ada rule yang dihasilkan juga. Sehingga setelah kami coba beberapa kali, dengan nilai minimum support 0,01 dan minimum support 0,8 dihasilkan sebanyak 6 rule. Berikut adalah gambar dari rule yang didapat.

```

Apriori
=====

Minimum support: 0.01 (3 instances)
Minimum metric <confidence>: 0.8
Number of cycles performed: 20

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12

Size of set of large itemsets L(2): 36

Size of set of large itemsets L(3): 30

Best rules found:

1. G3=t Sem6=t 10 ==> P=t 9      conf:(0.9)
2. G3=t L=t 8 ==> Sem4=t 7      conf:(0.88)
3. G7=t L=t 6 ==> Sem4=t 5      conf:(0.83)
4. G1=t Sem4=t 27 ==> P=t 22    conf:(0.81)
5. G1=t 43 ==> P=t 35          conf:(0.81)
6. G1=t Sem6=t 16 ==> P=t 13    conf:(0.81)

```

Gambar 6. Hasil asosiasi dengan rule yang didapatkan

Penjelasan dari rule di atas adalah sebagai berikut:

- 1) Jika mahasiswa tersebut adalah mahasiswa G3 (Biologi) dan berada di semester 6, maka jenis kelaminnya adalah perempuan, dengan confidence 0,9.
- 2) Jika mahasiswa tersebut adalah mahasiswa G3 (Biologi) dan jenis kelaminnya adalah laki-laki, maka mahasiswa tersebut berada di semester 4 dengan confidence 0,88.
- 3) Jika mahasiswa tersebut adalah mahasiswa G7 (Fisika) dan jenis kelaminnya adalah laki-laki, maka mahasiswa tersebut berada di semester 4 dengan confidence 0,83.
- 4) Jika mahasiswa tersebut adalah mahasiswa G1 (Statistika) dan berada di semester 4, maka jenis kelaminnya adalah perempuan dengan confidence 0,81.
- 5) Jika mahasiswa tersebut adalah mahasiswa G1 (Statistika), maka jenis kelaminnya adalah perempuan, dengan confidence 0,81.
- 6) Jika mahasiswa tersebut adalah mahasiswa G1 (Statistika) dan berada di semester 6, maka jenis kelaminnya adalah perempuan, dengan confidence 0,81.

## **KESIMPULAN DAN SARAN**

### **Kesimpulan**

Dari hasil penelitian yang telah dilakukan maka dapat diambil kesimpulan antara lain :

1. Pola keikutsertaan organisasi mahasiswa FMIPA berdasarkan departemen, jenis kelamin dan semester dari data anggota BEM FMIPA IPB dapat dilihat dengan menerapkan metode association rule mining menggunakan algoritma apriori.
2. Algoritma apriori cocok diterapkan untuk item set dengan jumlah data record yang sedikit.
3. Dengan menggunakan nilai minimum support sebesar 0,01 dan minimum confidence sebesar 0,8 didapatkan 6 rule terbaik dengan nilai maksimum confidence sebesar 0,9.

### **Saran**

Dari penelitian yang telah dilakukan dapat direkomendasikan saran-saran sebagai berikut:

1. Pada pengembangan yang akan datang, dapat gunakan pula data dari mahasiswa non anggota BEM sebagai data tandingan. Selain itu dapat juga digunakan data anggota BEM dari periode yang lebih panjang.
2. Dalam pencarian pola dalam association rule mining, dapat dicoba menggunakan algoritma lain ketika jumlah data yang digunakannya besar.
3. Untuk menambah keakuratan referensi, bisa pula ditambahkan atribut-atribut lain yang mungkin dapat menjadi dasar pertimbangan seseorang dalam mengikuti suatu organisasi, misal IPK atau golongan darah.

## DAFTAR PUSTAKA

- Amiruddin, I Ketut Eddy Purnama dan Purnomo, Mauridhi Hery. 2010. *Penerapan Association Rule mining pada Data Nomor Unik Pendidik dan Tenaga Kependidikan untuk Menemukan Pola Sertifikasi guru*. Digilib ITS.
- Berry, Michael J.A dan Linoff , Gordon S. 2004. *Data Mining Techniques For Marketing, Sales, Customer Relationship Management Second Editon*. Wiley Publishing, Inc.
- Han, Jiawei, dan Kamber, Micheline. 2004. *Data Mining: Concepts and Techniques*. 2nd Edition. Morgan Kaufmann.
- Kusrini dan Luthfi, Emha Taufiq. 2009. *Algoritma Data Mining*. Yogyakarta (ID): C.V Andi.
- Larose, Daniel T. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining*. John Wiley & Sons, Inc.
- Pramudiono, I. 2003. *Pengantar Data mining : Menambang Permata Pengetahuan di Gunung Data*.