

## **RANCANG BANGUN APLIKASI DATA MINING UNTUK MENAMPILKAN INFORMASI TINGKAT KELULUSAN MAHASISWA DENGAN ALGORITMA APRIORI**

**Harton Rohul Meisa Tambun<sup>1</sup>, Anofrizen<sup>2</sup>**

Jurusan Sistem Informasi Fakultas Sains dan Teknologi UIN Suska Riau

Email : <sup>1</sup>harton.rohul@yahoo.co.id, <sup>2</sup>anofrizen@gmail.com

### **ABSTRAK**

Pertumbuhan yang pesat dari akumulasi data telah menciptakan kondisi kaya akan data tapi minim informasi. Data mining merupakan penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data dalam jumlah besar yang diharapkan dapat mengatasi kondisi tersebut. Dengan memanfaatkan data induk mahasiswa dan data kelulusan mahasiswa, diharapkan dapat menghasilkan informasi tentang tingkat kelulusan dengan data induk mahasiswa melalui teknik data mining. Kategori tingkat kelulusan di ukur dari lama studi dan IPK. *Algoritma* yang digunakan adalah *algoritma apriori*, informasi yang ditampilkan berupa nilai *support* dan *Confidence* dari masing-masing kategori tingkat kelulusan.

Kata kunci : *algoritma apriori*, data induk mahasiswa, data mining, tingkat kelulusan

### **I. PENDAHULUAN**

Fakultas Sains dan Teknologi adalah salah satu fakultas yang ada di Universitas Islam Negeri Sultan Syarif Kasim Riau, Fakultas Sains dan Teknologi didirikan dengan tujuan untuk menghasilkan Sarjana muslim yang berakhlak mulia yang memiliki kemampuan akademik dan profesional dibidang Sains dan Tekhnologi yang berwawasan keislaman dalam mengembangkan dan memanfaatkan ilmu pengetahuan dan teknologi dalam kehidupan keseharian baik di masyarakat maupun bangsa. Fakultas Sains dan Teknologi dipimpin oleh seorang Dekan dan dibantu oleh Pembantu dekan I, Pembantu dekan II dan Pembantu dekan III. dimana fakultas Sains dan Teknologi terdiri dari lima jurusan yaitu : Sitem Informasi, Teknik Informatika, Teknik Industry, Teknik Elektro dan Matematika Terapan. Dengan jumlah mahasiswa semester ganjil tahun ajaran 2013/2014 3827 orang, terdiri dari mahasiswa Sistem Informasi 884 orang, mahasiswa Teknik Informatika 1256 orang, mahasiswa Teknik industri 646 orang, mahasiswa Matematika Terapan 427 orang dan mahasiswa Teknik Elektro 614 orang.

Dengan jumlah mahasiswa yang mencapai 3827 orang dengan tingkat kelulusa yang berbeda-beda, maka dibutuhkan suatu metode atau teknik yang dapat menganalisis tingkat kelulusan mahasiswa dan faktor yang mempengaruhi kelulusan mahasiswa didalam data

induk dan data kelulusan mahasiswa yang akan menjadi sebuah informasi berharga atau pengetahuan ( Knowledge) yang bermanfaat untuk mendukung pengambilan keputusan dalam menerima mahasiswa baru dan juga untuk menyusun strategi perkuliahan agar disesuaikan dengan metode belajarnya. Dengan cara mengelompokkan data mahasiswa kedalam kategori-kategori yang dapat menghasilkan suatu pola yang bisa dijadikan sebagai acuan untuk menemukan informasi yang tidak diketahui selama ini di dalam data induk dan data kelulusan mahasiswa.

Suatu teknologi yang dapat digunakan untuk mewujudkannya adalah data mining. data minig adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. salah satu metode yang sering kali digunakan dalam teknologi data mining adalah metode asosiasi atau asosiation rule mining. didalam data kelulusan dan data induk mahasiswa. yang dimaksud dengan asosiation rule adalah pola-pola keterkaitan data dalam basis data. atribut data mahasiswa yang akan digunakan sebagai acuan didalam penentuan hubungan asosiasi adalah sebagai berikut:

1. Hubungan tingkat kelulusan dengan jalur masuk ke UIN SUSKA RIAU yang terdiri dari jalur masuk Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN), Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBM-PTN),

Seleksi Penerimaan Mahasiswa Baru Perguruan Tinggi Agama Islam Negeri (SPMB-PTAIN), Undangan Jalur Mandiri (PBUD) dan Ujian Masuk Jalur Mandiri (UMJM), dengan harapan dapat mengetahui sejauh mana tingkat keberhasilan jalur masuk yang ada di UIN SUSQA RIAU agar pimpinan membuka peluang yang lebih besar terhadap jalur dengan lulusan tepat waktu terbanyak.

2. Hubungan tingkat kelulusan dengan asal sekolah dan jalur masuk, dengan harapan dapat mengetahui tingkat keberhasilan mahasiswa dari sebuah sekolah dengan jalur masuk yang ada agar keputusan untuk menerima lulusan yang tepat waktu di beri pulang lebih besar terhadap sekolah dengan jalur masuk yang ada.

3. Hubungan tingkat kelulusan dengan kota asal sekolah, Hubungan tingkat kelulusan dengan asal kota bermanfaat untuk mengetahui daerah-daerah mana yang mempunyai tingkat keberhasilan tinggi ataupun rendah.

4. Hubungan tingkat kelulusan dengan program studi, untuk mengetahui tingkat kelulusan program studi yang ada di Fakultas Sains dan Teknologi UIN SUSKA RIAU sebagai bahan evaluasi bagi pimpinan fakultas untuk meningkatkan jumlah lulusan yang tepat waktu terhadap jurusan yang ada di Fakultas Sains dan Teknologi.

Data mining adalah suatu konsep yang digunakan untuk menemukan pengetahuan yang tersembunyi didalam database. Data mining merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang tersimpan di dalam database besar. (Turban dkk, 2005 ). Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. (Pramudiono, 2006).

Dalam penelitian ini akan dicari tentang itemset, *support* dan *Confidence* dari data induk dan data kelulusan mahasiswa dengan *algoritma Apriori*. *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* atau *threshold*.

Dengan dilakukannya penelitian ini, diharapkan dapat memberikan hasil berupa informasi yang bermanfaat bagi pihak-pihak yang terkait dalam melakukan proses pengambilan keputusan manajerial, terutama yang berkaitan

dengan penerimaan mahasiswa baru, serta strategi perkuliahan.

## II. RUMUSAN MASALAH

Dari latar belakang yang telah di uraikan sebelumnya, maka permasalahan yang akan dibahas dalam penelitian ini adalah, “Bagaimana Rancang Bangun Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa Dengan *Algoritma Apriori*” di Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.

## III. BATASAN MASALAH

Masalah yang akan dikaji pada penelitian ini dibatasi pada :

1. Algoritma yang digunakan pada teknik data mining ini ialah *algoritma apriori* untuk menemukan pola frekuensi tinggi.
2. Dalam penulisan tugas akhir ini tidak membahas pada sistem pendukung keputusan maupun sistem informasi akademik.
3. Dalam tugas akhir ini tidak membahas pada hasil proses data mining dan analisis hasil yang keluar.

## IV. TUJUAN PENELITIAN

Adapun tujuan dari penelitian ini adalah:

1. Untuk merancang dan membangun aplikasi data mining yang dapat menghasilkan informasi yang berguna tentang hubungan tingkat kelulusan dengan data induk mahasiswa.
2. Untuk membantu mengetahui tingkat kelulusan mahasiswa dan faktor yang mempengaruhi tingkat kelulusan mahasiswa.
3. Untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) didalam data induk dan data kelulusan mahasiswa

## V. MANFAAT PENELITIAN

Adapun beberapa manfaat yang diharapkan pada pembuatan tugas akhir ini adalah:

1. Bagi Penulis  
Penulis dapat lebih mengetahui cara menerapkan ilmu-ilmu yang telah dipelajari selama ini dalam merancang dan membuat aplikasi sistem dengan teknik data mining, serta sebagai syarat dalam memperoleh gelar sarjana komputer.
2. Bagi Fakultas Sains dan Teknologi  
Diharapkan dengan adanya aplikasi ini dapat membantu menyajikan informasi tentang hubungan tingkat kelulusan

dengan data induk mahasiswa. Pihak fakultas dapat mengetahui tingkat kelulusan mahasiswanya dan mengetahui faktor yang mempengaruhi tingkat kelulusan mahasiswa.

## VI. METODE DATA MINING

Dengan definisi data mining yang luas, ada banyak jenis metode analisis yang dapat digolongkan dalam data mining.

### A. Association rules

*Association rules* (aturan asosiasi) atau *affinity analysis* (analisis afinitas) berkenaan dengan studi tentang “apa bersama apa”. Sebagai contoh dapat berupa studi transaksi di supermarket, misalnya seseorang yang membeli susu bayi juga membeli sabun mandi. Pada kasus ini berarti susu bayi bersama dengan sabun mandi. Karena awalnya berasal dari studi tentang database transaksi pelanggan untuk menentukan kebiasaan suatu produk dibeli bersama produk apa, maka aturan asosiasi juga sering dinamakan market basket analysis.

Aturan asosiasi ingin memberikan informasi tersebut dalam bentuk hubungan “*if-then*” atau “jika-maka”. Aturan ini dihitung dari data yang sifatnya *probabilistik* (Santoso, 2007).

Analisis asosiasi dikenal juga sebagai salah satu metode data mining yang menjadi dasar dari berbagai metode data mining lainnya. Khususnya salah satu tahap dari analisis asosiasi yang disebut analisis pola frekuensi tinggi (*frequent pattern mining*) menarik perhatian banyak peneliti untuk menghasilkan *algoritma* yang efisien. Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, *support* (nilai penunjang) yaitu prosentase kombinasi item tersebut. dalam database dan *Confidence* (nilai kepastian) yaitu kuatnya hubungan antar item dalam aturan asosiatif. Analisis asosiasi didefinisikan suatu proses untuk menemukan semua aturan asosiatif yang memenuhi syarat minimum untuk *support* (minimum *support*) dan syarat minimum untuk *Confidence* (minimum confidence) (Pramudiono, 2007).

Ada beberapa *algoritma* yang sudah dikembangkan mengenai aturan asosiasi, namun ada satu *algoritma* klasik yang sering dipakai yaitu *algoritma apriori*. Ide dasar dari *algoritma* ini adalah dengan mengembangkan *frequent itemset*. Dengan menggunakan satu item dan secara rekursif mengembangkan *frequent itemset* dengan dua item, tiga item dan seterusnya hingga *frequent itemset* dengan semua ukuran.

Untuk mengembangkan *frequent set* dengan dua item, dapat menggunakan *frequent set* item. Alasannya adalah bila set satu item tidak melebihi

*support* minimum, maka sembarang ukuran itemset yang lebih besar tidak akan melebihi *support* minimum tersebut. Secara umum, mengembangkan set dengan *fc-item* menggunakan *frequent set* dengan  $k - 1$  item yang dikembangkan dalam langkah sebelumnya. Setiap langkah memerlukan sekali pemeriksaan ke seluruh isi database.

Dalam asosiasi terdapat istilah *antecedent* dan *consequent*, *antecedent* untuk mewakili bagian “jika” dan *consequent* untuk mewakili bagian “maka”. Dalam analisis ini, *antecedent* dan *consequent* adalah sekelompok item yang tidak punya hubungan secara bersama (Santoso, 2007).

Dari jumlah besar aturan yang mungkin dikembangkan, perlu memiliki aturan-aturan yang cukup kuat tingkat ketergantungan antar item dalam *antecedent* dan *consequent*. Untuk mengukur kekuatan aturan asosiasi ini, digunakan ukuran *support* dan confidence. *Support* adalah rasio antara jumlah transaksi yang memuat *antecedent* dan *consequent* dengan jumlah transaksi. *Confidence* adalah rasio antara

Jumlah transaksi yang meliputi semua item dalam *antecedent* dan *consequent* dengan jumlah transaksi yang meliputi semua item dalam *antecedent*.

$$S = \frac{\sum(Ta+Tc)}{\sum(T)} \quad (1)$$

Keterangan :

$S = Support$

$\sum(Ta+Tc)$  = Jumlah transaksi yang mengandung *antecedent* dan consequent

$\sum(T)$  = Jumlah transaksi

$$C = \frac{\sum(Ta+Tc)\sum(Ta)}{\sum(Ta)} \quad (2)$$

Keterangan :

$C = Confidence$

$\sum(Ta+Tc)$  = Jumlah transaksi yang mengandung *antecedent* dan consequent

$\sum(Ta)$  = Jumlah transaksi yang mengandung *antecedent*

Langkah pertama *algoritma apriori* adalah, *support* dari setiap item dihitung dengan men-scan database. Setelah *support* dari setiap item didapat, item yang memiliki *support* lebih besar dari minimum *support* dipilih sebagai pola frekuensi tinggi dengan panjang 1 atau sering disingkat 1-itemset. Singkatan k-itemset berarti satu set yang terdiri dari k item.

*Iterasi* kedua menghasilkan 2-itemset yang tiap set-nya memiliki dua item. Pertama dibuat kandidat 2-itemset dari kombinasi semua 1-itemset. Lalu untuk tiap kandidat 2-itemset ini dihitung *support*-nya dengan men-scan database. *Support* artinya jumlah transaksi dalam database yang mengandung kedua item dalam kandidat 2-

itemset. Setelah *support* dari semua kandidat 2-itemset didapatkan, kandidat 2-itemset yang memenuhi syarat minimum *support* dapat ditetapkan sebagai 2-itemset yang juga merupakan pola frekuensi tinggi dengan panjang 2. (Pramudiono, 2007)

Untuk selanjutnya *iterasi iterasi* ke-k dapat dibagi lagi menjadi beberapa bagian :

1. Pembentukan kandidat itemset

Kandidat k-itemset dibentuk dari kombinasi (k-1)-itemset yang didapat dari *iterasi* sebelumnya. Satu ciri dari *algoritma apriori* adalah adanya pemangkasan kandidat k-itemset yang subset-nya yang berisi k-1 item tidak termasuk dalam pola frekuensi tinggi dengan panjang k-1.

2. Penghitungan *support* dari tiap kandidat k-itemset

*Support* dari tiap kandidat k-itemset didapat dengan men-scan database untuk menghitung jumlah transaksi yang memuat semua item di dalam kandidat k-itemset tersebut. Ini adalah juga ciri dari *algoritma apriori* yaitu diperlukan penghitungan dengan scan seluruh database sebanyak k-itemset terpanjang.

3. Tetapkan pola frekuensi tinggi

Pola frekuensi tinggi yang memuat k item atau k-itemset ditetapkan dari kandidat k-itemset yang *support*-nya lebih besar dari minimum *support*. Kemudian dihitung *Confidence* masing-masing kombinasi item.

*Iterasi* berhenti ketika semua item telah dihitung sampai tidak ada kombinasi item lagi. (Pramudiono, 2007) Secara ringkas *algoritma apriori* sebagai berikut :

```
Create L1 = set of supported itemsets of
cardinality one
Set k to 2
while (Lk-1 ≠ ∅) {
    Create Ck from Lk-1
    Prune all the itemsets in Ck that are not
supported, to create Lk
    Increase k by 1
}
The set of all supported itemsets is L1 ∪ L2 ∪ ... ∪ Lk
```

Selain *algoritma apriori*, terdapat juga *algoritma* lain seperti *FP-Grwoth*. Perbedaan *algoritma apriori* dengan *FP-Growth* pada banyaknya scan database. *Algoritma apriori* melakukan scan database setiap kali *iterasi* sedangkan *algoritma* *FP-Growth* hanya melakukan sekali di awal (Bramer, 2007).

## VII. ANALISA DAN PERANCANGAN APLIKASI DATA MINING

Bab ini menjelaskan tentang analisis dan perancangan dalam membangun Aplikasi Data Mining. Analisis meliputi analisis data mining, analisis lingkungan sistem serta analisis dalam membangun aplikasi.

### A. Analisis Data Mining

Dalam penulisan tugas akhir ini akan dicari nilai *support* dan *Confidence* dari hubungan tingkat kelulusan dengan data induk mahasiswa. Tidak semua data induk mahasiswa akan dicari hubungannya dengan data kelulusan, hanya beberapa atribut yang kira-kira berguna dan sebarannya tidak terlalu acak. Karena data yang terlalu acak akan membuat proses mining memakan waktu lama dan tingkat hubungannya pun rendah. Data induk mahasiswa yang akan dicari hubungannya meliputi proses masuk, asal sekolah, kota asal sekolah, dan program studi. Adapun yang akan diproses mining meliputi :

1. Hubungan tingkat kelulusan dengan proses masuk

Hasil dari proses mining ini dapat membantu untuk mengetahui sejauh mana tingkat keberhasilan SMPTN, PBUD, UMJM dan PTAIN.

2. Hubungan tingkat kelulusan dengan asal sekolah dan proses masuk

Dari atribut proses masuk dan asal sekolah dicari hubungan tingkat kelulusan dengan asal sekolah yang melalui proses masuk UMJM dan PBUD dengan harapan dapat mengetahui tingkat keberhasilan mahasiswa dengan sekolah tertentu.

3. Hubungan tingkat kelulusan dengan kota asal sekolah

Hubungan tingkat kelulusan dengan asal kota bermanfaat untuk mengetahui daerah-daerah mana yang mempunyai tingkat keberhasilan tinggi ataupun rendah. Diasumsikan bahwa kota asal sekolah merupakan kota tempat asal mahasiswa.

4. Hubungan tingkat kelulusan dengan program studi

Dari atribut program studi dapat diketahui hubungan tingkat kelulusan dan program studi untuk mengetahui tingkat kelulusan program studi.

### B. Data Yang Digunakan

Dalam penulisan tugas akhir kali ini dicari hubungan beberapa atribut dari data induk mahasiswa dengan tingkat kelulusan. Karena tidak semua tabel digunakan maka perlu dilakukan pembersihan data agar data yang akan diolah benar-benar relevan dengan yang

dibutuhkan. Pembersihan ini penting guna meningkatkan performa dalam proses mining. Cara pembersihan dengan menghapus atribut yang tidak terpakai dan menghapus data-data yang tidak lengkap isiannya. atribut yang digunakan terdiri dari atribut pada data kelulusan dan pada data induk mahasiswa. Atribut yang digunakan dalam data induk mahasiswa meliputi :

1. Atribut NIM digunakan sebagai primary key untuk menghubungkan dengan data kelulusan
2. Atribut proses masuk digunakan untuk proses mining guna mengetahui hubungan antara tingkat kelulusan dengan jalur masuk yang digunakan mahasiswa.
3. Atribut nama asal sekolah digunakan untuk proses mining guna mengetahui hubungan antara tingkat kelulusan dengan asal sekolah.
4. Atribut kota asal sekolah digunakan untuk proses mining guna mengetahui hubungan tingkat kelulusan dengan kota asal mahasiswa.

Atribut yang digunakan dalam data kelulusan meliputi :

1. NIM digunakan sebagai primary key untuk menghubungkan dengan data induk mahasiswa.
2. Indeks Prestasi Kumulatif (IPK) digunakan sebagai ukuran tingkat kelulusan mahasiswa
3. Lama studi digunakan sebagai ukuran tingkat kelulusan mahasiswa.
4. Program studi digunakan untuk proses mining guna mengetahui hubungan tingkat kelulusan dengan program studi.

#### C. Integrasi Data

Dalam penulisan tugas akhir kali ini diasumsikan bahwa data yang diambil sudah berupa tabel-tabel dalam satu server. Untuk proses mining, data kelulusan dan data induk mahasiswa digabungkan dengan primary key NIM. Setelah itu baru dilakukan proses mining. Proses integrasi data dilakukan ketika proses ETL (*extract, transform, and Load*) ketika membangun data warehouse, dalam proses ETL data dalam data source digabungkan menjadi satu dalam data warehouse dengan key NIM.

#### D. Transformasi Data

Transformasi data merupakan proses perubahan atau penggabungan data ke dalam format yang sesuai untuk diproses dalam data mining. Seringkali data yang akan digunakan dalam proses data mining mempunyai format

yang belum langsung bisa digunakan, oleh karena itu perlu dirubah formatnya.

Dalam penulisan tugas akhir ini penulis mencari keterkaitan antara tingkat kelulusan dengan data induk mahasiswa. Tingkat kelulusan mahasiswa dapat dilihat dari lama studi dan IPK (Indeks Prestasi Kumulatif). Dari dua parameter tersebut data diubah menjadi tipe data yang memudahkan untuk diproses. Tingkat kelulusan diukur dari lama studi dan IPK, Di dalam paduan informasi akademik UIN SUSKA Riau tahun 2011 pada BAB II menerangkan bahwa “Program Sarjana (S1) reguler adalah program pendidikan, yang memiliki beban studi sekurang-kurangnya 144 (seratus empat puluh empat) sks (satuan kredit semester) dan sebanyak-banyaknya 160 (seratus enam puluh ) sks yang dijadwalkan untuk 8 (delapan) semester dan dapat ditempuh dalam waktu minimal 8 (delapan) semester dan maksimal 14 (empat belas) semester” (Paduan Informasi Akademik, 2011). sedangkan IPK dikategorikan berdasarkan predikat kelulusan yang diatur dalam peraturan akademik pada BAB II menerangkan bahwa “ Predikat kelulusan yang diberlakukan di UIN SUSKA Sebagai berikut :

Tabel 1. Predikat Kelulusan

Indeks Prestasi Kumulatif	Predikat
3.50 - 4.00	Pujian ( <i>Cum Lode</i> )
3.00 - 3.49	Sangat Memuaskan
2.50 - 2.99	Memuaskan
2.00 - 2.49	Cukup

Dari tabel 1 data kelulusan berdasarkan IPK dapat dikategorikan menjadi empat kategori yaitu:

1. IPK Pujian (*Cum Lode*) dengan IPK 3.50 - 4.00
2. IPK Sangat Memuaskan dengan IPK 3.00 - 3.49
3. IPK Memuaskan dengan IPK 2.50 - 2.99
4. IPK Cukup dengan IPK 2.00 - 2.49

Pengkategorian data kelulusan berdasarkan lama studi yaitu :

1. Sesuai jadwal, bila lama studi 4 tahun atau kurang dari 4 tahun
2. Tidak sesuai jadwal, bila lama studi lebih dari 4 tahun

Dari dua pengkategorian tersebut dapat dibuat kategori berdasarkan kombinasi keduanya, seperti yang dapat dilihat pada Tabel 2 :

Tabel 2. Transformasi Data

Kategori	Keterangan
A1	lama studi 4 tahun atau kurang dari 4 tahun dan IPK 3.50 - 4.00
A2	lama studi 4 tahun atau kurang dari 4 tahun dan IPK 3.00 - 3.49
A3	lama studi 4 tahun atau kurang dari 4 tahun dan IPK 2.50 - 2.99
A4	lama studi 4 tahun atau kurang dari 4 tahun dan IPK 2.00 - 2.49
B1	lama studi lebih dari 4 tahun dan 3.50 - 4.00
B2	lama studi lebih dari 4 tahun dan 3.00 - 3.49
B3	lama studi lebih dari 4 tahun dan 2.50 - 2.99
B4	lama studi lebih dari 4 tahun dan 2.00 - 2.49

Dari kombinasi yang terdapat di Tabel 2 terdapat delapan tingkatan untuk mengukur tingkat kelulusan mahasiswa.

#### E. Penggunaan *Algoritma Apriori*

*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* atau *threshold*. *Threshold* adalah batas minimum transaksi. Jika jumlah transaksi kurang dari *threshold* maka item atau kombinasi item tidak akan diikutkan perhitungan selanjutnya. Penggunaan *threshold* dapat mempercepat perhitungan.

*Algoritma apriori* dibagi menjadi beberapa tahap yang disebut *iterasi*. Tiap *iterasi* menghasilkan pola frekuensi tinggi dengan panjang yang sama dimulai dari *pass* pertama yang menghasilkan pola frekuensi tinggi dengan panjang satu. Di *iterasi* pertama ini, *support* dari setiap item dihitung dengan men-scan database. Setelah *support* dari setiap item didapat, item yang memiliki *support* lebih besar dari minimum *support* dipilih sebagai pola frekuensi tinggi dengan panjang 1 atau sering disingkat 1-itemset. Singkatan k-itemset berarti satu set yang terdiri dari k item.

*Iterasi* kedua menghasilkan 2-itemset yang tiap set-nya memiliki dua item. Pertama dibuat kandidat 2-itemset dari kombinasi semua 1-itemset. Lalu untuk tiap kandidat 2-itemset ini dihitung *support*-nya dengan men-scan database. *Support* artinya jumlah transaksi dalam database yang mengandung kedua item dalam kandidat 2-

itemset. Setelah *support* dari semua kandidat 2-itemset didapatkan, kandidat 2-itemset yang memenuhi syarat minimum *support* dapat ditetapkan sebagai 2-itemset yang juga merupakan pola frekuensi tinggi dengan panjang 2.

Contoh proses mining untuk mengetahui hubungan tingkat kelulusan dengan proses masuk. Misal data seperti pada Tabel 3 :

Table 3. Data Awal

NIM	Kategori Kelulusan	Jalur Masuk
11053102994	A1	PBUD Lokal
11053102993	A2	SNMPTN
11053102992	A1	PBUD Nasional
11053102991	A3	PTAIN
11053102980	B2	UMJM
11053102981	A3	PBUD Nasional
11053102982	A3	PBUD Lokal
11053102983	A2	SNMPTN
11053102984	A2	UMJM
11053102985	A2	UMJM
11053102986	B2	SNMPTN
11053102994	A1	PBUD Lokal
11053102993	A2	SNMPTN
11053102992	A1	PBUD Nasional
11053102991	A3	PTAIN
11053102980	B2	UMJM
11053102981	A3	PBUD Nasional
11053102982	A3	PBUD Lokal
11053102983	A2	SNMPTN
11053102984	A2	UMJM
11053102985	A2	UMJM
11053102986	B2	SNMPTN

Dari data awal tersebut didapat kandidat pertama (C1) seperti pada Table 4 :

Tabel 4. Kandidat Pertama (C1) Itemset Count

Itemset	Count
A1	2
A2	4
A3	3
B2	2
SNMPTN	3
PBUD Lokal	2
PBUD Nasional	2
PTAIN	2
UMJM	3

Ditetapkan *threshold* = 3, maka kandidat yang nilainya kurang dari 3 akan dihapus. Sehingga, didapat hasil seperti pada Tabel 5 (L1) :

Tabel 5. hasil setelah threshold ditetapkan (L1)

Itemset	Count
A2	4
A3	3
SMPTN	3
UMJM	3

Dari tabel Tabel (L1) didapat kandidat kedua (C2) seperti pada Tabel 6 :

Tabel 6 Kandidat kedua (C2)

Itemset	Count
A2, SMPTN	2
A2, UMJM	2
A3, SMPTN	0
A3, UMJM	0

Setelah ditetapkan threshold menghasilkan data seperti pada Tabel 7 :

Tabel 7 Hasil kedua (L2)

Itemset	Count
A2, UMJM	2
A2, SMPTN	2

Dari Tabel 7 dapat diambil hasil sebagai berikut :

$Support\ A2,\ UMJM = Count(A2,\ UMJM) / \text{jumlah transaksi} = 2/11$

$Support\ A2,\ SMPTN = Count(A2,\ SMPTN) / \text{jumlah transaksi} = 2/11$

$Confidence\ A2,\ UMJM = Count(A2,\ UMJM) / Count(A2) = 2/4$

$Confidence\ A2,\ SMPTN = Count(A2,\ SMPTN) / Count(A2) = 2/4$

Dapat dilihat bahwa proses mining hubungan tingkat kelulusan dengan proses masuk mahasiswa dengan threshold 3 menghasilkan hubungan A2, UMJM mempunyai nilai  $support = 2/11$   $Confidence = 2/4$  dan hubungan A2, SMPTN mempunyai nilai  $support = 2/11$   $Confidence = 2/4$  mempunyai UMJM mempunyai tingkat kelulusan A2 dan SMPTN mempunyai tingkat kelulusan A2 sehingga dapat disimpulkan bahwa mahasiswa yang melalui proses masuk UMJM mempunyai tingkat kelulusan sama dengan mahasiswa yang melalui proses masuk SMPTN.

## VIII. PENUTUP

### A. Kesimpulan

Kesimpulan yang dapat diambil dalam pengerjaan tugas akhir ini adalah Aplikasi Data Mining ini dapat digunakan untuk menampilkan informasi tingkat kelulusan. Informasi yang ditampilkan berupa nilai *support* dan *Confidence* hubungan antara tingkat kelulusan dengan data induk mahasiswa. Semakin tinggi nilai *Confidence* dan *support* maka semakin kuat nilai hubungan antar atribut. Data induk mahasiswa yang diproses mining meliputi data proses masuk, data asal sekolah, data kota mahasiswa, dan data program studi. Hasil dari proses data mining ini dapat digunakan sebagai pertimbangan dalam mengambil keputusan lebih lanjut tentang faktor yang mempengaruhi tingkat kelulusan khususnya faktor dalam data kelulusan dan data induk mahasiswa.

### B. Saran

Untuk pengembangan Aplikasi Data Mining lebih lanjut, dapat menggunakan *algoritma* lain, misal *algoritma* FP-Growth. Perbedaannya adalah *algoritma apriori* harus melakukan scan database setiap kali *iterasi*, sedangkan *algoritma* FP-Growth hanya melakukan satu kali scan database diawal.

## REFERENSI

- 1) Bramer, Max. "Principles of Data Mining",
- 2) Springer, London. Chintakayala. 2007.
- 3) Davies, and Paul Beynon. "Database Systems Third Edition", Palgrave Macmillan, New York. 2004.
- 4) Elmasri, Ramez and Shamkant B. Navathe. "Fundamentals of Database Systems. Third Edition", Addison Wesley Publishing Company, New York. 2000.
- 5) Han, J. and Kamber, M. "Data Mining Concepts and Techniques Second Edition". Morgan Kaufman, San Francisco. 2006.
- 6) Harianto Kristanto, Ir. "Konsep dan Perancangan Database", Penerbit Andi Yogyakarta. 1993.
- 7) Kusriani, dan Emha Taufik Luthfi, 2009, "Algoritma Data Mining", Penerbit Andi, Yogyakarta.
- 8) Khair Hayatul. "Paduan Informasi Akademik UIN SUSKA Riau 2010-2011", Pekanbaru. 2011.

- 17) Rossa, A. S. Dkk, Modul Pembelajaran Rekayasa
- 18) Perangkat Lunak (Terstruktur dan Berorientasi
- 19) Objek). Halaman 26-28, 86-90, 118, 122, 124, 130,
- 20) 134, 137-138. Modula, Bandung. 2011.
- 21) Pramudiono, I. "Pengantar Data Mining :
- 22) Menambang Permata Pengetahuan di
- 23) Gunung Data " .  
[http://www.ilmukomputer.org/wp-](http://www.ilmukomputer.org/wp-content/uploads/2006/08/iko-datamining.zip)
- 24) [content/uploads/2006/08/iko-datamining.zip](http://www.ilmukomputer.org/wp-content/uploads/2006/08/iko-datamining.zip)  
Diakses
- 25) pada tanggal 03 Oktober 2013 jam 08.54. 2007.
- 26) Pramudiono, I. " *Algoritma Apriori*"
- 27) [http://datamining.japati.net/cgi-](http://datamining.japati.net/cgi-bin/indodm.cgi?bacaarsip&1172210143)
- 28) [bin/indodm.cgi?bacaarsip&1172210143](http://datamining.japati.net/cgi-bin/indodm.cgi?bacaarsip&1172210143)  
Diakses
- 29) pada tanggal 03 Oktober jam 10.00. 2007.