

PENERAPAN DATA MINING UNTUK ANALISIS PENGARUH LAMA STUDI MAHASISWA TEKNIK INFORMATIKA UIN SUNAN KALIJAGA YOGYAKARTA MENGGUNAKAN METODE APRIORI

Gathut Cakra Sutradana ¹⁾

M. Didik R. Wahyudi²⁾

^{1,2}Teknik Informatika FST UIN Sunan Kalijaga

Email : m.didik@uin-suka.ac.id

ABSTRACT

The accuracy of a long study of college students at a university becomes very important in demonstrating the quality of the learning process in college. There are many things that affect a student's study time. Data Mining offers a way to know the various aspects that may affect a student's study time. To know the various aspects that influence the duration of the study based on data graduation students are available, then the implementation of a Data Mining algorithms can be used. In this study, Data Mining algorithms used to find aspects that affect student study duration is Apriori algorithm.

Keywords: *graduation analysis, long studying, data mining, apriori algorithms*

Ketepatan lama studi mahasiswa pada suatu perguruan tinggi menjadi hal yang sangat penting dalam menunjukkan kualitas proses pembelajaran di perguruan tinggi. Ada banyak hal yang mempengaruhi lama studi mahasiswa. Data Mining menawarkan suatu cara untuk mengetahui berbagai aspek yang dapat berpengaruh terhadap lama studi mahasiswa. Untuk mengetahui berbagai aspek yang mempengaruhi lama studi mahasiswa berdasarkan data kelulusan yang tersedia, maka implementasi suatu algoritma Data Mining dapat dipergunakan. Dalam penelitian ini, algoritma Data Mining yang dipergunakan untuk menemukan aspek yang mempengaruhi lama studi mahasiswa adalah algoritma Apriori.

Katakunci : *analisis kelulusan, lama studi, data mining, algoritma apriori*

1. PENDAHULUAN

Perkembangan teknologi informasi yang begitu maju saat ini, menyebabkan tingkat akurasi suatu data sangat dibutuhkan dalam kehidupan sehari-hari. Setiap informasi yang ada menjadi suatu hal penting untuk menentukan setiap keputusan dalam situasi tertentu. Hal ini menyebabkan penyediaan informasi menjadi sarana untuk dianalisa dan diringkas menjadi suatu pengetahuan dari data yang bermanfaat ketika pengambilan suatu keputusan dilakukan.

Dengan menggunakan *data mining*, setiap kumpulan atau gudang data dapat memberikan pengetahuan penting yang menjadi informasi yang sangat berharga bagi suatu organisasi, seperti pada organisasi pendidikan. *Data mining* adalah suatu konsep penggalian data yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam database.

Data-data alumni khususnya Teknik Informatika di UIN Sunan Kalijaga misalnya di tumpuk dan tidak ada tindak lanjut manfaat dari data-data yang tersedia. Padahal data - data tersebut dapat di manfaatkan dan di olah kembali untuk menjadi sebuah pengetahuan dan informasi yang bermanfaat sebagai bahan pertimbangan dalam sebuah keputusan.

Pada buku pedoman akademik di bagian beban studi menerangkan “Program Sarjana (S1) merupakan jenjang pertama program akademik yang mempunyai beban studi 144-160 SKS, di jadwalkan sekurang-kurangnya 8 semester, dapat di tempuh dalam waktu kurang dari 8 semester dan selama-lamanya 14 semester, setelah pendidikan menengah”(pedoman akademik, 2012). Berdasarkan data alumni pada kenyataannya pola studi mahasiswa berbeda-beda setiap mahasiswa. Mahasiswa menempuh studi kurang dari 8 semester, tepat di semester 8 bahkan lebih dari smester 8. Hal inilah yang membuat penulis mengkaji penelitian analisis pengaruh lama studi mahasiswa TIF UIN Sunan Kalijaga Yogyakarta menggunakan metode apriori.

1. METODE PENELITIAN

Data mining merupakan proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Data minig juga dapat di artikan sebagai pengekstraksi informasi baru yang di ambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah *data mining* kadang juga disebut juga *Knowledge Discovery* (Tan, 2006).

Menurut Gartner Group *data mining* adalah suatu proses menemukan hubungan yang berarti pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Larose, 2005).

Data mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database besar (larose, 2005).

2.1 Tahap *Data mining*

Tahap-tahap *data mining* yaitu (Han, 2006) :

1. Pembersihan data (*data cleaning*)
Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang di peroleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa *data mining* yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performasi dari teknik *data mining* karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.
2. Integrasi data (*data integration*)
Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Tidak jarang data yang diperlukan untuk *data mining* tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi data dilakukan pada atribut-aribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.
3. Seleksi Data (*Data Selection*)
Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus market basket analysis, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.
4. Transformasi data (*Data Transformation*)
Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*. Beberapa metode *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka

numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

5. *Proses Mining*
Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.
6. *Evaluasi pola (pattern evaluation)*
Untuk mengidentifikasi pola-pola menarik kedalam knowledge based yang ditemukan. Dalam tahap ini hasil dari teknik *data mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses *data mining*, mencoba metode *data mining* lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.
7. *Presentasi pengetahuan (knowledge presentation)*
Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses *data mining* adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami *data mining*. Karenanya presentasi hasil *data mining* dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses *data mining*. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil *data mining*.

2.2 Algoritma Apriori

Algoritma *apriori* adalah suatu algoritma dasar yang diusulkan oleh Agrawal & Srikant pada tahun 1994 untuk menentukan *Frequent itemsets* untuk aturan asosiasi Boolean. Algoritma *Apriori* termasuk jenis Aturan Asosiasi pada *data mining*. Aturan yang menyatakan asosiasi antara beberapa atribut sering disebut affinity analysis atau market basket analysis. Analisis asosiasi atau *association rule mining* adalah teknik *data mining* untuk menemukan aturan suatu kombinasi item. Salah satu tahap analisis asosiasi yang menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien adalah analisis pola frekuensi tinggi (*frequent pattern mining*). Penting tidaknya suatu asosiasi dapat diketahui dengan dua tolok ukur, yaitu : *support* dan *confidence*. *Support* (nilai penunjang) adalah persentase kombinasi item tersebut dalam database, sedangkan *confidence* (nilai kepastian) adalah kuatnya hubungan antara-item dalam aturan asosiasi. (Kusrini, Emha Taufiq Luthfi, 2009 : 149)

Algoritma *apriori* dibagi menjadi beberapa tahap yang disebut narasi (Devi dinda setiawan, 2009).

1. Pembentukan kandidat *itemset*. Kandidat *k-itemset* dibentuk dari kombinasi (*k-1*) *itemset* yang didapat dari *iterasi* sebelumnya. Satu cara dari algoritma *apriori* adalah pemangkasan kandidat *k-itemset* yang subsetnya 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 menscan database untuk menghitung jumlah transaksi yang memuat semua item didalam kandidat *k-itemset* tersebut. Ini adalah juga ciri dari algoritma *apriori* dimana diperlukan penghitungan dengan cara 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 *supportnya* lebih besar dari minimum *support*.
4. Bila tidak didapat pola frekuensi tinggi baru maka seluruh proses dihentikan.

Association rules (aturan asosiasi) atau affinity analysis (analisis afinitas) berkenaan dengan studi tentang “apa bersama apa”. Sebagai contoh dapat berupa 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).

Metodologi dasar analisis asosiasi terbagi menjadi dua tahap (Kusrini & Luthfi, 2009):

1. Analisis pola frekuensi tinggi

Tahap ini mencari kombinasi yang memenuhi syarat minimum dari nilai *support* dalam database. Nilai *support* sebuah item di peroleh dengan rumus berikut.

$$\text{Support}(A) = \frac{\text{jumlah transaksi mengandung } A}{\text{Total transaksi}}$$

Sementara itu , nilai *support* dari 2 item di peroleh dari rumus berikut :

$$\text{Support}(A,B) = P(A \cap B)$$

$$\text{Support}(A,B) = \frac{\sum \text{Transaksi mengandung } A \text{ dan } B}{\sum \text{Total transaksi yang terjadi}}$$

2. Pembentukan aturan asosiasi

Setelah semua pola frekuensi tinggi di temukan , barulah dicari aturan asosiasi yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan asosiatif $A \rightarrow B$.

Nilai dari *confidence* dari aturan $A \rightarrow B$ di peroleh dari rumus berikut :

$$\text{Confidence} = P(B | A) = \frac{\sum \text{Transaksi mengandung } A \text{ dan } B}{\sum \text{Transaksi mengandung } A}$$

2.3 Bahasa R

R merupakan system analisis statistika yang relative lengkap, sebagai hasil dari kolaborasi riset berbagai statistikawan diseluruh dunia. Saat ini R dapat dikatakan merupakan *lingua franca* (bahasa standar) untuk keperluan komputasi statistika modern. Versi pertama R dibuat tahun 1992 oleh di Universitas Auckland, New Zeland oleh Ross Ihaka dan Robert Gentleman (yang menjadi asal muasal akronim R untuk perangkat lunak ini).

Kelebihan menggunakan bahasa R

1. Probilitas. Jika menggunakan perangkat lunak ini, pengguna (user) bebas mempelajari dan menggunakannya sampai kapanpun.
2. Multiplatform. R lebih compatible dari perangkat lunak stastistika manapun yang pernah ada. Dengan demikian jika pengguna memutuskan untuk berpindah system operasi, penyesuaiannya akan relative lebih mudah untuk dilakukan.
3. Umum dan berada dibarisan terdepan. Berbagai metode analisis statistik telah diprogramkan kedalam bahasa R.
4. Bisa deprogram. Pengguna dapat memprogram metode baru atau mengembangkan modifikasi dari fungsi – fungsi analisis statistika yang telah ada dalam system R.
5. Bahasa berbasis analisis matriks. Bahasa R sangat baik untuk melakukan pemrograman dengan bais matriks
6. Fasilitas grafik yang relatif baik.

Kelemahan bahasa R :

1. *Point* dan *Click* GUI. Interaksi utama dengan R bersifat *Command Line Interface* (CLI). Walaupun saat ini sudah tersedia menu point and *Click* GUI (*Grafik User Interface*) sederhana untuk keperluan statistika tertentu.
2. Ketidakterseidanya beberapa fungsi statistic tertentu. Walaupun analisis statistika dalam R cukup lengkap, tidak semua metode statistika diimplementasi-kan kedalam R(Rosadi,2011).

3. HASIL DAN PEMBAHASAN

3.1 Model Pengambilan Data

Tahapan yang dilakukan dalam pengambilan data penelitian ini adalah sebagai berikut :

1. Metode Pengamatan Langsung
Melakukan pengamatan langsung ke pusat pangkalan data yaitu pada PTIPD untuk memperoleh data yang dibutuhkan.
2. Metode Dengan Wawancara
Mengadakan wawancara dengan Bapak Agung selaku ketua dari pusat pangkalan data PTIPD terkait izin dan data-data yang dibutuhkan pada penelitian ini.

3.2 Model Pembentukan Data

Dari data mentah yang diperoleh kemudian dilakukan tahapan untuk pembentukan data untuk bahan analisis. Data mentah di olah dalam beberapa tahapan, tahapan-tahapan dalam pembentukan data sebagai berikut :

1. Pembersihan data
Data yang diperoleh dari PTIPD tidak sepenuhnya sempurna. Terdapat beberapa data kosong, oleh karenanya untuk pemenuhan kebutuhan penelitian ini. Maka dari itu data-data yang kosong akan dihilangkan atau dilakukan pembersihan data. Pembersihan data ini mempengaruhi performansi dari teknik data mining, karena data yang ditangani berkurang jumlah dan kompleksitasnya.
2. Integrasi data
Pada tahap ini melakukan penggabungan atau pembuatan satu database baru dari data mentah yang sudah di peroleh dari PTIPD di olah kembali atau mengintegrasikan data-data yang terkait dan diperlukan dalam penelitian.
3. Seleksi Data
Pada tahap seleksi data, tidak semua data mentah yang di berikan dari PTIPD di pakai karena ada beberapa atribut yang tidak memungkinkan di pakai, seperti NIM, nama, tanggal masuk dan keluar, pekerjaan ibu. Data-data ini tidak dipakai karena tidak masuk dalam kebutuhan data yang akan diolah atribut tersebut hanya membantu dalam pengolahan data base baru kaitanya dengan kebutuhan data untuk penelitian.
4. Transformasi data
Bentuk data dari PTIPD berupa data mentah dengan format excel.xls. Penulis dalam melakukan penelitian menggunakan aplikasi tools R dan dengan menggunakan excel sebagai penguat hasil analisis. Disini format yang di pakai berbentuk excel jadi data-data yang sudah di olah berupa data yang dibutuhkan telah disiapkan yaitu dalam bentuk format excel.xls.
5. Proses Mining
Pada penerapan metode melakukan pencarian data-data yang menarik seperti pengelompokan dalam tiga kategori semester lulus dan selanjutnya implemetasi pada algoritma apriori yang dapat dicapai dalam menentukan hasil analisis..
6. Evaluasi pola
Mengevaluasi hasil analisis yaitu pola seperti dari asal sekolah, jenis pekerjaan orangtua, kelas, dan nilai ipk mahasiswa. Sehingga dapat memeproleh informasi yang berharga ketika hasil analisis sudah tercapai.
7. Presentasi pengetahuan
Dari hasil analisis yang telah di terapkan dengan metode data mining algoritma apriori ke dalam aplikasi Bahasa R dan Microsoft excel didapatkan sebuah informasi dan pengetahuan tentang hasil dari analisis penelitian.

3.3 Proses Algoritma *Apriori*

Dalam menemukan aturan suatu kombinasi item dari data mahasiswa, dilakukan pencarian pola frekuensi tinggi dari database tiga kategori semester lulus, tahapan-tahapan yang di lakukan sebagai berikut :

1. Melakukan scan terhadap database semester 6-7, semester 8, dan semester 9-14 untuk mendapatkan kandidat pertama dari 1 itemset dengan menghitung nilai pada masing-masing 1 itemset. Kemudian nilai 1 itemset dibandingkan dengan batas minimal nilai yang telah di tentukan.

2. 1 itemset yang nilainya di atas atau sama dengan batas minimal nilai akan di ikutkan dalam iterasi berikutnya, sedangkan nilai 1 itemset yang nilainya dibawah batas minimal nilai tidak di ikutkan dalam proses berikutnya.
3. Nilai dari 1 itemset yang diatas atau sama dengan batas minimal nilai akan di pakai untuk pembentukan kandidat kedua dari kombinasi itemset. Yaitu kombinasi dari asal sekolah, pekerjaan orang tua, kelas, kategori ipk.
4. Setelah kandidat kedua yang berisi kategori asal sekolah, pekerjaan orangtua, kelas, dan kategori ipk, di hitung kembali jumlah kemunculannya pada kombinasi tersebut.

Selanjutnya kombinasi 2 itemset yang nilainya dibawah batas minimal nilai tidak akan di ikutkan pada iterasi berikutnya. Proses pembentukkan kombinasi itemset ini akan trus dilakukan selama nilai kombinasi itemset sama dengan atau di atas batas minimal nilai dan sampai tidak ada lagi kombinasi dari asal sekolah , pekerjaan orantua, kelas dan ketegori ipk. Proses penggabungan kombinasi itemset berhenti ketika tidak ada lagi kombinasi yang dibentuk.

5. Dari hasil yang terbentuk di tetapkan pola kombinasi itemset paling tinggi dari asal sekolah pekerjaan orangtua, kelas , dan kategori ipk. Kemudian dibentuk aturan asosiasi untuk menghitung nilai *support* dan *confidence*.

Untuk lebih jelasnya seperti yang disajikan pada gambar 1 adalah skema proses pengolahan data menggunakan Algoritma Apriori.



Gambar 1 Skema Proses Pengolahan Data Dengan Algoritma Apriori

3.4 Pembersihan Data

Setelah data mentah di peroleh kemudian dilakukan pembersihan data. Pada tahap ini dilakukan proses menghilangkan data yang kosong. Dari proses pembersihan data tersebut diperoleh hasil pembersihan data seperti yang disajikan pada tabel 1 :

Tabel 1 Perbandingan Sebelum Dan Sesudah Pembersihan Data

Sebelum Pembersihan Data	Sesudah Pembersihan Data
444 Record	414 Record

3.5 Integrasi Data

Untuk mengolah data-data yang sesuai dengan kebutuhan analisis dari data mentah tentunya tidak semua data menjadi bahan analisis. penulis mengolah kembali menjadi kategori semester lulus yaitu dengan cara menghitung lama tahun dijadikan tiap-tiap semester. Kategori Penjelasanya seperti yang di sajikan pada tabel 2 sebagai berikut :

Tabel 2 Kategori Per-Semester

Tgl Masuk	Tgl Lulus	Tahun	Semester	Keterangan
01/09/2010	01/03/2011	0.50-0.99	1.00	Satu Semester
01/09/2010	01/09/2011	1.00-1.49	2.00	Dua Semester
01/09/2010	01/03/2012	1.50-1.99	3.00	Tiga Semester
01/09/2010	01/09/2012	2.00-2.49	4.00	Empat Semester
01/09/2010	01/03/2013	2.50-2.99	5.00	Lima Semester
01/09/2010	01/09/2013	3.00-3.49	6.00	Enam Semester
01/09/2010	01/03/2014	3.50-3.99	7.00	Tujuh Semester
01/09/2010	01/09/2014	4.00-4.49	8.00	Delapan Semester
01/09/2010	01/03/2015	4.50-4.99	9.00	Sembilan Semester
01/09/2010	01/09/2015	5.00-5.49	10.00	Sepuluh Semester
01/09/2010	01/03/2016	5.50-6.00	11.00	Sebelas Semester
01/09/2010	01/09/2016	6.01-6.49	12.00	Duabelas Semester
01/09/2010	01/03/2017	6.50-7.00	13.00	Tigabelas Semester
01/09/2010	01/09/2017	7.01-7.49	14.00	Empat Belas Semester

Tahapan selanjutnya yaitu pengkategorisasian IPK, Pada data mentah nilai IPK setiap mahasiswa memiliki pola yang berbeda-beda, oleh karena itu untuk membuat data olahan agar dapat disetarakan menjadi bentuk pola analisis penulis mengkategorikan nilai IPK seperti yang disajikan pada tabel 3 :

Tabel 3 Kategori IPK

Kategori	Nilai
IPK1	2.00 - 2.25
IPK2	2.26 - 2.50
IPK3	2.51 - 2.75
IPK4	2.76 - 3.00
IPK5	3.01 - 3.25
IPK6	3.26 - 3.50
IPK7	3.51 - 3.75
IPK8	3.76 - 4.00

3.6 Seleksi Data

Pada tahap ini menentukan data-data yang dipakai untuk analisis. Sehingga data-data yang tidak dipakai akan dibuang dan hanya data yang sesuai dengan bahan analisis saja yang digunakan.

Kemudian pada tahap seleksi data ini, data-data yang digunakan yaitu meliputi : (Asal_Sekolah), (Pekerjaan_Orangtua), (Kelas), dan (Kategori_IPK). Sedangkan kategori semester pada tahap ini di kelompokkan sebagai bahan analisis dengan tiga kategori rujukan pencarian pola, yaitu kategori :

1. Lama studi 6-7 semester
2. Lama studi 8 semester
3. Dan lama studi 9-14 semester

3.7 Analisis data dengan metode *Apriori*

Pada tahap ini adalah proses menerapkan metode algoritma *apriori* untuk menemukan pengetahuan terkait pola-pola apa saja yang terbentuk dari data yang sudah di kelompokkan dari tahapan-tahapan sebelumnya. Tentukan batas minimal *support* dan *confidence* untuk membatasi nilai minimal yang akan dicapai terkait aturan-aturan tinggi di atas minimal dari *support* dan *confidence*.

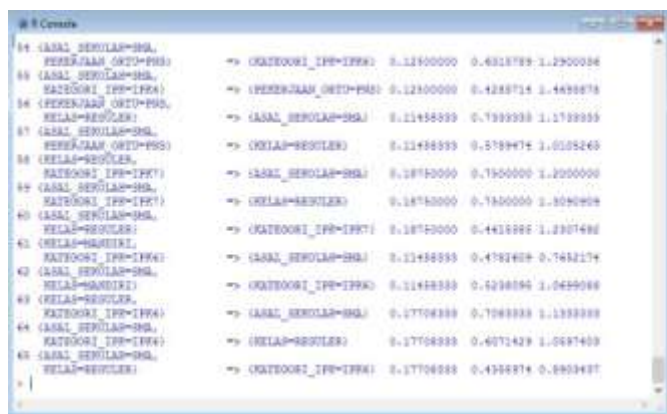
Support adalah suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu item atau *itemset* dari keseluruhan transaksi. Sedangkan *Confidence*, adalah suatu ukuran yang menunjukkan hubungan antar item.

Penulis menentukan batas minimum *support* sebesar 7 %. Dan minimum *confidence* sebesar 40 % dalam penelitian ini. Batas minimum *support* dan *confidence* ditentukan untuk mencari pola kombinasi *itemset* tinggi dalam database transaksi.

Data-data yang sudah dikelompokkan sebelumnya yaitu meliputi tiga kategori akan dianalisis untuk mengetahui dari ketiga jenis pola yang terbentuk terkait penerapannya dengan algoritma.

Lama Studi Semester 6 Dan 7

Tahap untuk mencari pola kombinasi asosiasi *rule* dengan memasukan *script* aplikasi, dan menentukan nilai *support* 7% dan *confidence* 40% sebagai berikut : `> rules <- apriori(semester6_7, parameter= list(supp=0.07, conf=0.40))`. Setelah dilakukan *script* di peroleh hasil seperti yang disajikan pada gambar 2 sebagai berikut :

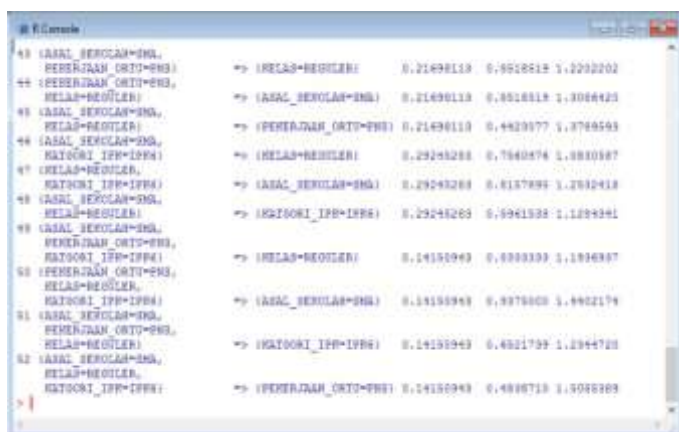


Itemset	Support	Confidence	Lift
(KATASARI_IPH=IPH)	0.1250000	0.631875	1.2300000
(PERHAJAN_OTU=PH)	0.1250000	0.428715	1.4438875
(KATASARI_IPH=IPH)	0.1250000	0.799933	1.1799333
(KATASARI_IPH=IPH)	0.1250000	0.5799475	1.0105240
(KATASARI_IPH=IPH)	0.1250000	0.7900000	1.2000000
(KATASARI_IPH=IPH)	0.1250000	0.7900000	1.2000000
(KATASARI_IPH=IPH)	0.1250000	0.4418888	1.2307488
(KATASARI_IPH=IPH)	0.1250000	0.4792408	0.7432174
(KATASARI_IPH=IPH)	0.1250000	0.6298888	1.0459088
(KATASARI_IPH=IPH)	0.1250000	0.7700000	1.1333333
(KATASARI_IPH=IPH)	0.1250000	0.6071429	1.0974000
(KATASARI_IPH=IPH)	0.1250000	0.4304874	0.8903487

Gambar 2 Hasil Asosiasi Rule Kombinasi

Lama Studi Semester 8

Tahap untuk mencari pola kombinasi asosiasi *rule* dengan memasukan *script* aplikasi, dan menentukan nilai *support* 7% dan *confidence* 40% sebagai berikut : `> rules <- apriori(semester8, parameter= list(supp=0.07, conf=0.40))`. Setelah dilakukan *script* di peroleh hasil seperti yang disajikan pada gambar 3 sebagai berikut :

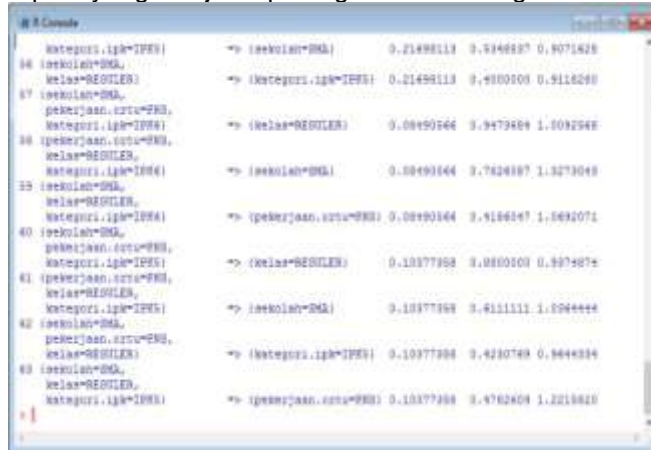


Itemset	Support	Confidence	Lift
(KATASARI_IPH=IPH)	0.2143811	0.5518812	1.2202002
(PERHAJAN_OTU=PH)	0.2143811	0.4518812	1.3004420
(KATASARI_IPH=IPH)	0.2143811	0.4423077	1.3764543
(KATASARI_IPH=IPH)	0.2143811	0.7840474	1.3810187
(KATASARI_IPH=IPH)	0.2143811	0.6137899	1.2812418
(KATASARI_IPH=IPH)	0.2143811	0.5941328	1.1294341
(KATASARI_IPH=IPH)	0.2143811	0.6303333	1.1814387
(KATASARI_IPH=IPH)	0.2143811	0.5878888	1.4402174
(KATASARI_IPH=IPH)	0.2143811	0.4821799	1.2344728
(KATASARI_IPH=IPH)	0.2143811	0.4898715	1.5053388

Gambar 3 Hasil Asosiasi Rule Kombinasi

Lama Studi Lulus Semester 9-14

Tahap untuk mencari pola kombinasi asosiasi *rule* dengan memasukan *script* aplikasi, dan menentukan nilai *support* 7% dan *confidence* 40% sebagai berikut : `> rules <- apriori(semester9_14, parameter=list(supp=0.07, conf=0.40))`. Setelah dilakukan *script* di peroleh hasil seperti yang disajikan pada gambar 4 sebagai berikut :



Rule Number	Antecedent	Consequent	Support	Confidence
34	sekolah=SMA, kelas=REGULER	ipk=IPK1	0.21488113	0.5346827
37	sekolah=SMA, pekerjaan_orangtua=PNS, ipk=IPK1	kelas=REGULER	0.21488113	0.4900100
38	ipk=IPK1, kelas=REGULER	sekolah=SMA	0.08490344	0.9473484
39	sekolah=SMA, kelas=REGULER	ipk=IPK1	0.08490344	0.7424387
40	sekolah=SMA, pekerjaan_orangtua=PNS, ipk=IPK1	kelas=REGULER	0.08490344	0.9186247
41	ipk=IPK1, kelas=REGULER	sekolah=SMA	0.10377368	0.8800100
42	sekolah=SMA, kelas=REGULER	ipk=IPK1	0.10377368	0.4111111
43	sekolah=SMA, pekerjaan_orangtua=PNS, ipk=IPK1	kelas=REGULER	0.10377368	0.4230748

Gambar 4 Hasil Asosiasi *Rule* Kombinasi

4. KESIMPULAN

Berdasarkan pengolahan data kelulusan yang ada, dapat disimpulkan bahwa :

1. Dapat menerapkan data *mining* dengan *algoritma apriori* pada data alumni TIF terkait mencari pola-pola yang mempengaruhi lama studi mahasiswa berdasarkan tiga kategori kelulusan yang dapat di ukur dari semester lulus.
2. Pola-pola yang terbentuk dari tiga kategori lulus semester yaitu :
 - a. Lulus semester 6 dan 7, pola itemset yng terbentuk adalah mahasiswa dari asal sekolah sma, dari kelas regular, cenderung memiliki nilai ipk 3,51-3,75 dengan tingkat dominasi (*support*) sebesar 18,75%. Sedangkan untuk nilai kepercayaan (*confidence*) yang paling berpengaruh terhadap lulus studi di semester 6 dan 7 adalah faktor dari asal sekolah sma, dan dari kelas regular yang sama-sama memiliki nilai kepercayaan (*confidence*) sebesar 75%. 46% mahasiswa dari asal sekolah sma dan dari kelas regular memiliki ipk 3,51-3,75.
 - b. Lulus pada semester 8, pola itemset yang terbentuk adalah mahasiswa dari asal sekolah sma, jenis pekerjaan orangtuanya PNS, dari kelas regular, dan memiliki ipk 3,26-3,50 dengan tingkat dominasi (*support*) sebanyak 14,15%. Sedangkan yang paling berpengaruh terhadap lulus di semester 8 ini adalah berasal dari sekolah sma dengan nilai kepercayaan(*confidence*) sebesar 93,75% , pengaruh kedua yaitu dari kelas regular dengan nilai kepercayaan (*confidence*) sebesar 83%, dan mahasiswa yang pekerjaan orangtuanya dari PNS dengan nilai kepercayaan (*confidence*) sebesar 48%, serta mahasiswa cenderung memiliki ipk 3,26-3,50 dengan nilai kepercayaan (*confidence*) sebesar 78% dari ketiga kombinasi tersebut.
 - c. Lulus pada semester 9-14, pola itemset yang terbentuk pada semester ini yaitu mahasiswa dari asal sekolah sma, pekerjaan orangtuanya PNS, dari kelas regular, dan cenderung memiliki ipk 3,01-3,25 dengan tingkat dominasi (*support*) sebanyak 10,38%. Sedangkan untuk nilai kepercayaan (*confidence*) yang paling berpengaruh terhadap lulus di semester ini yaitu dari kelas reeguler dengan nilai kepercayaan (*confidence*) 88%, dan pengaruh kedua yaitu dari asal sekolah sma dengan nilai kepercayaan (*confidence*) sebanyak 61,11%, kemudian dari jenis pekerjaan orangtua PNS dengan nilai kepercayaan (*confidence*) 47,83%, serta cenderung dari kombinasi ini memiliki ipk 3,01-3,25 dengan nilai kepercayaan (*confidence*) 42,31%.

DAFTAR PUSTAKA

- Bangkik, 2015, *Pengertian, Fungsi Dan Kegunaan Microsoft Excel*. Sumber : <http://forum.idws.id/threads/pengertian-fungsi-dan-kegunaan-microsoft-excel.550824/> (diakses pada tanggal 12 Juni 2016)
- Dedi Rosadi, 2011, *Analisis ekonometrika dan runtut waktu terapan dengan R*. penerbit ANDI.
- Devi dinda setiawan, 2009, *Penggunaan Metode Apriori Untuk Analisa Keranjang Pasar Pada Data Transaksi Penjualan Minimarket Menggunakan Java & Mysql*.
- Han, J. and Kamber, M, 2006, *Data Mining Concepts and Techniques Second Edition*. Morgan Kauffman, San Francisco
- Hermawati, F. A.(2013), *Data Mining*, Penerbit ANDI : Yogyakarta.
- Kusrini, & Luthfi, E. T, 2009, *Algoritma Data Mining*, Penerbit ANDI : Yogyakarta.
- Larose, Daniel T. 2005, *Discovering Knowledge in Data : An Introduction to Data Mining*, John Willey & Sons, Inc.
- Pramudiono, I. 2007, *Pengantar Data Mining : Menambang Permata Pengetahuan di Gunung Data*, <http://www.ilmukomputer.org/wpcontent/uploads/2006/08/iko-datamining.zip> Diakses pada tanggal 17 Juni 2016 jam 09.00.
- Prasetyo, E. (2012), *Data Mining Konsep dan Aplikasi menggunakan Matlab*, Penerbit ANDI : Yogyakarta.
- Rainardi, Vincent, 2008, *Building a Data Warehouse with Examples in SQL Server*, Springer, New York.
- Santosa, Budi, 2007, *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Graha Ilmu, Yogyakarta.
- Tan, P. et al.2006, *Introduction to Data Mining*, Boston:Pearson Education
- Turban,E., dkk. 2005, *Decicion Support System and Intelligent System*, Yogyakarta : Andi Offset
- Wahyu, Tyas, D.,2008 *Melakukan Penelitian Menggunakan Metode Asosiasi Rule*