

PENGGALIAN KAJIDAH ASOSIASI DALAM MEMILIH PROGRAM KEGIATAN PENDUKUNG MAHASISWA STIE-GK MUARA BULIAN

Azwar Anas

*Jurusan Manajemen STIE – GK Muara Bulian
Jl. Gajah Mada Muara Bulian 36613
Email: aans_07@yahoo.co.id*

Abstract

Elements that can not be avoided from a college is a student. The main objective college student is studying, getting science as evidenced by the formal recognition in the form of an undergraduate degree or diploma. To complete the lecture program, students have the right in accordance with the conditions of their respective colleges. One student rights STIE-GK Muara Bulian is to choose a program supporting activities apart from Tri Dharma College, namely, computer courses, English or SME activities. In this study, the authors used a priori algorithm to dig association rules from the selected additional program options students STIE-GK Muara Bulian. This is done in order to obtain information which programs are likely to appear together based on student choice. In this study, the authors used the data of 50 students who choose additional programs. Software I use for testing is Orange Data Mining. The results of this study are, for the data 2 itemsets tend to choose a combination of computer and English language courses with support levels reached 22% and 34% confidence. While the data with 3 itemsets students tend to choose a combination of the SME program, English and Computer with support levels reached 18% and the confidence level reaches 100%

Keywords: College Student, Data Mining, Algorithm Apriori, Association, Itemsets.

Abstrak

Unsur yang tidak bisa dihindarkan dari sebuah perguruan tinggi adalah mahasiswa. Tujuan utama mahasiswa kuliah adalah menuntut ilmu, mendapatkan ilmu yang dibuktikan dengan pengakuan yang formal dalam bentuk ijazah sarjana atau diploma. Untuk menyelesaikan program perkuliahnya, mahasiswa memiliki hak sesuai dengan kondisi perguruan tinggi masing-masing. Salah satu hak mahasiswa STIE-GK Muara Bulian adalah memilih program kegiatan pendukung selain dari Tri Dharma Perguruan Tinggi yaitu, mengikuti kursus komputer, Bahasa Inggris ataupun kegiatan UKM. Dalam penelitian ini, penulis menggunakan Algoritma Apriori untuk menggali kaidah asosiasi dari pilihan program tambahan yang dipilih mahasiswa STIE-GK Muara Bulian. Hal ini dilakukan agar mendapatkan informasi program mana yang cenderung muncul bersamaan berdasarkan pilihan mahasiswa. Dalam penelitian ini, penulis menggunakan 50 data mahasiswa yang memilih program tambahan. Software yang penulis gunakan untuk melakukan uji coba adalah Orange Data Mining. Adapun hasil dari penelitian ini adalah, untuk data 2 itemsets mahasiswa cenderung memilih kombinasi program Komputer dan Bahasa Inggris dengan tingkat support mencapai 22% dan confidence mencapai 34%. Sedangkan data dengan 3 itemsets mahasiswa cenderung memilih kombinasi program UKM, Bahasa Inggris dan Komputer dengan tingkat support mencapai 18% dan tingkat confidence mencapai 100%.

Kata kunci: Mahasiswa, Data Mining, Algoritma Apriori, Asosiasi, Itemset.

© 2017 Jurnal MEDIASISFO.

1. Pendahuluan

Unsur yang tidak bisa dihindarkan dari sebuah perguruan tinggi adalah mahasiswa. Tujuan utama mahasiswa kuliah adalah menuntut ilmu, mendapatkan ilmu yang dibuktikan dengan pengakuan yang formal dalam bentuk ijazah sarjana atau diploma. Untuk menyelesaikan program perkuliahnya,

mahasiswa memiliki hak sesuai dengan kondisi perguruan tinggi masing-masing. Salah satu hak mahasiswa STIE-GK Muara Bulian adalah memilih program kegiatan pendukung selain dari Tri Dharma Perguruan Tinggi yaitu, mengikuti kursus komputer, Bahasa Inggris ataupun kegiatan UKM. Ketiga kegiatan tersebut merupakan garis besar. Karena pada kegiatan UKM, terdapat pula beberapa kegiatan pula. Sesuai dengan aturan yang ada, mahasiswa boleh memilih satu, dua ataupun ketiga program tersebut. Penulis sengaja melakukan hal agar mendapatkan informasi program mana yang cenderung muncul bersamaan berdasarkan pilihan mahasiswa.

Teknik penambangan data dapat dilakukan dengan beberapa algoritma. Khusus untuk data yang memuat kombinasi beberapa variabel, dapat dianalisis dengan menggunakan Algoritma Apriori. Penerapan Algoritma Apriori telah banyak digunakan sebelumnya untuk mendapatkan informasi berharga dari sejumlah frekuensi data. Berikut beberapa penerapan Algoritma Apriori yang telah dilakukan :

1. Analisa Algoritma Apriori untuk Mendapatkan Pola Peminjaman Buku Perpustakaan SMPN 3 Batanghari^[1].
2. Penerapan Data Mining Algoritma Asosiasi untuk Meningkatkan Penjualan^[2].
3. Analisis *Market Basket* dengan Algoritma Apriori dan FP-Growth^[3].
4. Implementasi *Association Rule* terhadap Penyusunan *Layout* Makanan dan Penentuan Paket Makanan Hemat di RM Roso Echo dengan Algoritma Apriori^[4].
5. Implementasi Data Mining pada Penjualan Tiket Pesawat Menggunakan Algoritma Apriori^[5].
6. Implementasi Data Mining dengan Metode Algoritma Apriori dalam Menentukan Pola Pembelian Obat^[6].

Tugas dari Algoritma Aprioria adalah untuk menghasilkan semua kaidah asosiasi pada suatu tabel transaksional, yang mempunyai nilai *support* lebih dari *minimum support*. Kaidah tersebut juga harus mempunyai *support* yang lebih besar dari *confidence*.

2. Tinjauan Pustaka

2.1. Definisi *Data Mining*

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual^[7]. Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Turban, dkk 2007)^[8].

2.2. Teknik – teknik *Data Mining*

Data mining dibagi menjadi beberapa teknik berdasarkan tugas yang dapat dilakukan, yaitu^[9] :

1. Deskripsi
Terkadang analis/ peneliti ingin mendeskripsikan pola dan trend yang tersimpan dalam data.
2. Estimasi
Estimasi mirip dengan klasifikasi, kecuali variabel tujuan yang lebih kearah numerik daripada kategori. Misalnya, akan dilakukan estimasi tekanan darah *systolic* dari pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level sodium darah.
3. Prediksi
Prediksi memiliki kemiripan dengan estimasi dan klasifikasi. Hanya saja, prediksi hasilnya menunjukkan sesuatu yang belum terjadi (mungkin terjadi di masa depan).
4. Klasifikasi
Dalam klasifikasi variabel, tujuan bersifat kategorik. Misalnya, kita akan mengklasifikasikan pendapatan dalam 3 kelas, yaitu mendapatkan tinggi, pendapatan sedang, dan pendapatan rendah.
5. *Clustering*
Clustering lebih kearah pengelompokan *record*, pengamatan, atau kasus dalam kelas yang memiliki kemiripan. Sebuah *cluster* adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lain dan memiliki ketidakmiripan dengan *record-record* dalam *cluster* yang lain.
6. Asosiasi
Mengidentifikasi hubungan antara berbagai peristiwa yang terjadi pada satu waktu. Pendekatan asosiasi tersebut menekankan sebuah kelas masalah yang dicirikan dengan analisis keranjang pasar.

Algoritma-algoritma *data mining* kebanyakan berasal atau merupakan bentuk pengembangan dari algoritma-algoritma dari bidang ilmu *machine learning*, statistika, intelegensia buatan dan jaringan syaraf tiruan, karena tidak dirancang untuk menangani data dalam ukuran yang sangat besar, sedangkan *data mining* dimaksudkan untuk menangani data dalam ukuran yang demikian, maka salah satu arah penelitian di bidang *data mining* adalah mengembangkan algoritma-algoritma tersebut agar dapat menangani data yang berukuran sangat besar.

Selain itu, integrasi teknik-teknik *data mining* ke dalam DBMS, khususnya *Object-Relation DBMS* (ORDBMS) yang merupakan teknologi DBMS terbaru, juga masih merupakan bidang penelitian yang aktif. Tujuan utama dari integrasi ini adalah untuk “melebur” algoritma *data mining* agar menjadi fungsi internal ORDBMS yang berkualitas, sehingga pengguna dapat menggunakannya sesuai dengan kebutuhan. DBMS merupakan teknologi yang sudah matang, digunakan secara luas, dapat mengelola data dalam ukuran yang sangat besar, memfasilitasi kueri tabel yang mudah dengan *Structured Query Language* (SQL) dan memiliki fitur yang mengakomodasi kebutuhan pengguna ORDBMS, maka pengembangan dan peleburan algoritma *data mining* ke dalam ORDBMS yang baik adalah dengan memanfaatkan semua ini.

2.3. Penerapan Data Mining

Banyak penerapan yang dapat dilakukan oleh *data mining*, apalagi ditunjang kakayaan dan keanekaragaman bidang ilmu (*artificial intelligence*, *database*, statistik, pemodelan matematika, dan pengolahan citra) membuat penerapan *data mining* menjadi makin luas. Bidang penerapan *data mining* dapat dilakukan diantaranya adalah :

1. Analisa Pasar dan Manajemen
 - Untuk analisa pasar, banyak sumber data yang dapat digunakan seperti transaksi kartu kredit, kartu anggota klub tertentu, kupon diskon, keluhan pembeli, ditambah dengan studi tentang gaya hidup publik. Beberapa kondisi yang bisa diselesaikan dengan *data mining* diantaranya :
 - a. Menembak target pasar
 - b. Melihat pola beli pemakai dari waktu ke waktu
 - c. *Cross Market Analysis*
 - d. Profil *customer*
 - e. Identifikasi kebutuhan *customer*
 - f. Menilai loyalitas *customer*
 - g. Informasi *summary*
2. Analisa Perusahaan dan Manajemen Resiko
 - a. Perencanaan keuangan dan evaluasi aset
Data mining dapat membantu melakukan analisis dan prediksi *cash flow* serta melakukan *contingent claim analysis* untuk mengevaluasi aset. Selain itu juga dapat menggunakannya untuk analisis tren.
 - b. Perencanaan sumber daya (*Resource Planning*)
Dengan melihat informasi ringkas (*summary*) serta pola pembelajaran dan pemasukan dari masing-masing *resource*, dan dapat memanfaatkannya untuk melakukan *resource planning*.
 - c. Persaingan (*Competition*)
 - 1) Sekarang ini banyak perusahaan yang berupaya untuk dapat melakukan *competitive intelligence*. *Data mining* dapat membantu memonitor pesaing-pesaing dan melihat *market direction*.
 - 2) Dapat melakukan pengelompokan *customer* dan memberikan variasi harga/ layanan/ bonus untuk masing-masing grup.
 - 3) Menyusun strategi penetapan harga di pasar yang sangat kompetitif. Hal ini diterapkan oleh perusahaan minyak REPSOL di Spanyol dalam menetapkan harga jual gas di pasaran.
3. Telekomunikasi
Sebuah perusahaan telekomunikasi menerapkan *data mining* untuk melihat dari jutaan transaksi yang masuk, transaksi mana sajakah yang masih harus ditangani secara manual (dilayani oleh orang). Tujuannya tidak lain adalah untuk menambah layanan otomatis khusus untuk transaksi-transaksi yang masih dilayani secara manual. Dengan demikian jumlah operator penerima transaksi manual tetap bisa ditekan minimal.
4. Keuangan
Financial Crimes Enforcement Network di Amerika Serikat baru-baru ini menggunakan *data mining* untuk menambang trilyunan dari berbagai subjek seperti properti, rekening bank dan transaksi

keuangan lainnya untuk mendeteksi transaksi-transaksi keuangan yang mencurigakan (seperti *money laundry*), yang susah dilakukan jika menggunakan analisa standar.

5. Asuransi

Australian Health Insurance Commision menggunakan *data mining* untuk mengidentifikasi layanan kesehatan yang sebenarnya tidak perlu tetapi tetap dilakukan oleh peserta asuransi. Hasilnya, berhasil menghemat satu juta dolar per tahunnya. Tentu saja ini tidak hanya bisa diterapkan untuk asuransi kesehatan, tetapi juga untuk berbagai jenis asuransi lainnya.

6. Olah Raga

IBM Advanced Scout menggunakan *data mining* untuk menganalisis statistik permainan NBA (jumlah *shots blocked*, *assist* dan *fouls*) dalam rangka mencapai keunggulan bersaing (*competitive advantage*) untuk tim New York Knicks dan Miami Heat.

7. Astronomi

Jet Propulsion Laboratory (JPL) di Pasadena, Clifornia dan *Palomar Observatory* berhasil menemukan 22 quasar dengan bantuan *data mining*. Hal ini merupakan salah satu kesuksesan penerapan *data mining* di bidang astronomi dan ilmu ruang angkasa.

8. *Internet Web Surf-Aid*

IBM Surf-Aid menggunakan algoritma *data mining* untuk mendata akses halaman web khususnya yang berkaitan dengan pemasaran guna melihat perilaku dan minat *customer* serta melihat keefektifan pemasaran melalui web.

Data mining merupakan salah satu aktivitas di bidang perangkat lunak yang dapat memberikan ROI (*return on investment*) yang tinggi. Namun demikian *data mining* tetaplah hanya alat bantu yang dapat membantu manusia untuk melihat pola, menganalisis *trend* dalam rangka mempercepat pembuatan keputusan.

2.4. Association Rule

Association rule merupakan salah satu metode yang bertujuan mencari pola yang sering muncul di antara banyak transaksi, dimana setiap transaksi terdiri dari beberapa *item*^[10]. Contoh dari aturan asosiatif dari analisa pembelian di suatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seorang pelanggan membeli roti bersama 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 isi keranjang belanja di pasar swalayan, analisis asosiasi juga sering disebut dengan istilah *market basket analysis*.

Association rule merupakan teknik *data mining* untuk mencari pola hubungan dalam data atau basis data. Yang paling populer adalah *market basket analysis* (MBA). Teknik MBA ini digunakan untuk mencari pelanggan yang membeli barang pada took X, akan membeli barang Y dan seterusnya^[10].

Tabel 1. Contoh Transaksi untuk Association Rule Analysis

Trans	Itemset
1	A, B, E
2	B, D
3	B, C
4	A, B, D
5	A, C
6	B, C
7	A, C
8	A, B, C, E
9	A, B, C

Dalam *association analysis* *itemset* merupakan kumpulan nol atau lebih item. Pada contoh tabel di atas diumpamakan *itemset* adalah barang yang dibeli pada took tersebut, trans adalah kode transaksi. Pada tabel tersebut dapat dilihat pada transaksi 1 yang dibeli adalah barang A, B dan E. Transaksi 2 yang dibeli adalah B dan D, demikian seterusnya. *Association analysis* akan melakukan analisis hubungan dengan aturan tertentu sering disebut sebagai *association rule*.

Sebagai contoh tabel tersebut, pembeli yang biasanya membeli barang A dan B akan membeli barang C. aturan ini dapat ditulis $\{A, B\} \rightarrow \{C\}$. Untuk memperoleh kuantifikasinya terminologi *support* dan *confidence* perlu diperhatikan. *Support* dan *confidence* dapat dihitung sebagai berikut :

$$\text{Support, } s(X \rightarrow Y) = \frac{\sigma(XUY)}{N} \quad (1)$$

$$\text{Confidence, } c(X \rightarrow Y) = \frac{\sigma(XUY)}{\sigma_X} \quad (2)$$

Di mana X dan Y sebagai *itemset* dan σ adalah *support* dari *itemset*.

Dengan rumus tersebut maka *support* yang didapat dari hubungan $s(A, B \rightarrow C)$ adalah $\frac{2}{9} = 0, 22$. Sedangkan nilai *confidence* yang diperoleh adalah $c(A, B \rightarrow C)$ adalah $\frac{2}{4} = 0, 5$. Nilai 2 pada *support* diperoleh dari *itemset* A, B, C berjumlah 2 yaitu pada trans 8 dan 9, jumlah keseluruhan trans adalah 9. Sedangkan 2 pada *confidence* sama seperti pada *support*, 4 diperoleh dari jumlah A, B yaitu pada trans 1, 4, 8 dan 9. Dari perolehan tersebut, kita dapat membuat aturan persentasi yang diinginkan untuk menentukan kepastian asosiasinya. Paling baik adalah apabila nilai *support* maupun *confidence* tinggi.

2.5. Algoritma Apriori

Apriori adalah suatu algoritma untuk melakukan pencarian *frequent itemset* untuk mendapatkan *association rules*. Sesuai dengan namanya, algoritma ini menggunakan *prior knowledge* mengenai *frequent itemset properties* yang telah diketahui sebelumnya, untuk memproses informasi selanjutnya. Apriori menggunakan pendekatan secara *iterative* yang disebut juga sebagai *level-wise search* di mana *k-itemset* digunakan untuk mencari $(k+1)$ -*itemset*^[11].

Algoritma Apriori menggunakan *knowledge* mengenai *frequent itemset* yang telah diketahui sebelumnya, untuk memproses informasi selanjutnya. Pola frekuensi tinggi adalah pola-pola item dalam *database* yang memiliki frekuensi atau *support* di atas ambang batas tertentu yang disebut istilah *minimum support*. Pola frekuensi tinggi ini digunakan untuk menyusun aturan asosiatif dan juga beberapa teknik *data mining* lainnya.

Prinsip dari Algoritma Apriori adalah :

1. Kumpulkan jumlah item tunggal, dapatkan item besar
2. Dapatkan *candidate pairs*, hitung \rightarrow *large pairs* dari item-item
3. Dapatkan *candidate triplets*, hitung \rightarrow *large triplets* dari item-item dan seterusnya
4. Sebagai petunjuk : setiap *subset* dari sebuah *frequent itemset* harus menjadi *frequent*

Dua proses utama dalam Algoritma Apriori merupakan langkah yang akan dilakukan untuk mendapat *frequent itemset*. Walaupun Algoritma Apriori mudah untuk dipahami dan diimplementasikan dibandingkan algoritma yang lainnya yang memang diterapkan untuk proses *association rule*, akan tetapi Algoritma Apriori juga memiliki kekurangan yaitu untuk melakukan pencarian *frequent itemset*, Algoritma Apriori harus melakukan *scanning database* berulang kali untuk setiap kombinasi item. Hal tersebut menyebabkan banyaknya waktu yang dibutuhkan untuk melakukan *scanning database*. Selain itu dibutuhkan *generate candidate* yang besar untuk mendapatkan kombinasi item dari *database*.

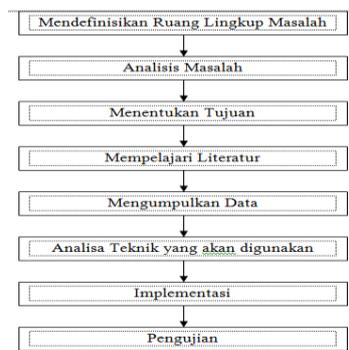
3. METODE PENELITIAN

3.1. Pendahuluan

Dalam melakukan sebuah penelitian, tentu harus berdasarkan metode penelitian yang benar sehingga dapat mempermudah jalannya penelitian tersebut. Metode penelitian merupakan kerangka kerja dalam melakukan penelitian. Dengan mengikuti kerangka kerja tersebut maka penelitian yang dilakukan akan berjalan dengan sistematis dan memberikan hasil yang baik.

3.2. Kerangka Penelitian

Pada bagian ini akan diuraikan kerangka kerja penelitian, kerangka kerja ini merupakan langkah-langkah yang akan dilakukan dalam penyelesaian masalah yang akan dibahas.



Gambar 1. Kerangka Penelitian

Berdasarkan kerangka kerja pada gambar 1, maka masing-masing langkahnya dapat diuraikan seperti berikut ini :

1. Mendefinisikan Ruang Lingkup Masalah
Ruang lingkup masalah yang akan diteliti harus ditentukan terlebih dahulu, karena tanpa mampu menentukan serta mendefinisikan batasan masalah yang akan diteliti, maka tidak akan pernah didapat suatu solusi yang terbaik dari masalah tersebut. Jadi ini adalah langkah awal yang terpenting dalam penulisan penelitian.
2. Analisis Masalah
Langkah analisis masalah adalah untuk dapat memahami masalah yang telah ditentukan ruang lingkup atau batasannya. Dengan menganalisa masalah yang telah ditentukan tersebut, maka diharapkan masalah dapat dipahami dengan baik. Pada bagian ini digambarkan proses untuk mendapatkan kaidah asosiasi kecenderungan pilihan program tambahan mahasiswa STIE-GK Muara Bulian dengan menggunakan Algoritma Apriori. Disamping itu juga diuraikan tentang kebutuhan data untuk menganalisa kaidah asosiasi pilihan program tambahan mahasiswa di STIE-GK Muara Bulian.
3. Penentuan Tujuan
Berdasarkan pemahaman dari masalah, maka ditentukan tujuan yang akan dicapai dari penulisan ini. Pada tujuan ini ditentukan target yang dicapai, terutama yang dapat mengatasi masalah-masalah yang ada.
4. Mempelajari Literatur
Untuk mencapai tujuan, maka dipelajari berbagai literatur yang relevan dengan masalah yang diteliti. Kemudian literatur-literatur yang dipelajari tersebut diseleksi untuk dapat ditentukan literatur-literatur mana yang akan digunakan dalam penelitian. Literatur diambil dari internet, berupa artikel dan jurnal ilmiah tentang *data mining*, *association rule* dan Algoritma Apriori, serta bahan bacaan lain yang mendukung penelitian.
5. Pengumpulan Data dan Informasi
Dalam pengumpulan data dilakukan observasi yaitu pengamatan secara langsung di tempat penelitian sehingga permasalahan yang ada dapat diketahui dengan jelas. Kemudian dilakukan *interview* yang bertujuan untuk mendapatkan informasi atau data yang dibutuhkan. Selain itu juga dilakukan studi kepustakaan yaitu dengan membaca buku-buku yang menunjang dalam melakukan analisa terhadap data dan informasi yang didapat. Adapun data-data yang diperlukan dalam penelitian ini adalah data pilihan mahasiswa STIE-GK Muara Bulian tahun akademik 2016/2017 tentang pilihan program tambahan yang diambil.
6. Analisa Teknik yang digunakan
Bagian ini bertujuan menganalisis dan memahami teknik yang akan digunakan dalam pengolahan data yang telah diperoleh dari tempat penelitian, terutama pada proses, yang mana teknik yang digunakan untuk pengolahan data dengan Algoritma Apriori.
7. Implementasi
Setelah dilakukan analisa teknik yang digunakan, maka dilakukan implementasi terhadap *software* dalam hal ini adalah *software data mining Orange*. Tentu saja penerapan analisa pada *software* tersebut menggunakan Algoritma Apriori. Jika penerapan sistem sudah berjalan dengan lancar dan akurat, maka sistem dapat diimplementasikan sesuai dengan kebutuhan.
8. Pengujian
Pengujian dilakukan untuk mendapatkan hasil yang sesuai dengan melakukan perbandingan antara perhitungan manual dan komputerisasi, berikut ini langkah-langkah pengujian yang akan dilakukan :
 - a. Melakukan pengujian Algoritma Apriori dari data pilihan program tambahan mahasiswa untuk mendapatkan kaidah asosiasi secara manual.

- b. Setelah dilakukan perhitungan manual, selanjutnya dilakukan perhitungan menggunakan *software data mining Orange* dengan memasukkan data-data pilihan program tambahan mahasiswa dan diproses menggunakan Algoritma Apriori.
- c. Tahap terakhir adalah melakukan perbandingan antara perhitungan manual dengan komputerisasi, jika hasil yang dicapai sama atau hampir sama, maka data yang diproses dan teknik yang digunakan telah benar.

4. PEMBAHASAN

4.1. Analisa Asosiasi dengan Algoritma Apriori

Analisis asosiasi atau *association rule mining* adalah teknik *data mining* untuk menentukan aturan asosiatif antara aturan kombinasi item. Dalam penelitian ini, penulis hanya melakukan analisa terhadap 50 data mahasiswa. Dari 50 data yang diamati, variabel yang akan diproses adalah, Kursus Komputer, Kursus Bahasa Inggris dan Unit Kegiatan Mahasiswa (UKM).

4.2. Menentukan Variabel

Data yang akan diolah dalam menentukan penggalian kaidah asosiasi ini adalah data mahasiswa beserta pilihan program tambahannya. Adapun variabel-variabel data mahasiswa tersebut adalah :

1. Kursus Komputer
2. Kursus Bahasa Inggris
3. UKM

4.3. Analisis Frekuensi Tinggi

Tahapan ini mencari kombinasi item yang memenuhi syarat minimum dari nilai *support* dalam *database*. Nilai *support* sebuah item diperoleh dengan rumus berikut :

$$\text{Support}(A) = \frac{\text{JumlahtransaksimengandungA}}{\text{Jumlahtransaksi}}$$

Sedangkan nilai *support* dari 2 item diperoleh dari rumus berikut :

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

$$\text{Support}(A,B) = \frac{\sum \text{JumlahtransaksimengandungAdanB}}{\sum \text{Jumlahtransaksi}}$$

4.4. Melakukan Proses

Berdasarkan variabel-variabel data mahasiswa di atas, maka format data dapat dilihat pada tabel berikut ini:

Tabel 2. Data Program Tambahan Mahasiswa

Mahasiswa	Program Tambahan 1	Program Tambahan 2	Program Tambahan 3
1	Bahasa Inggris	UKM	
2	UKM	Komputer	Bahasa Inggris
3	UKM	Bahasa Inggris	Komputer
4	UKM		
5	Komputer	Bahasa Inggris	
6	Bahasa Inggris	Komputer	
7	UKM	Komputer	
8	UKM	Bahasa Inggris	
9	UKM		
10	Bahasa Inggris	UKM	
11	UKM	Komputer	Bahasa Inggris
12	UKM	Bahasa Inggris	Komputer
13	UKM		
14	Komputer	Bahasa Inggris	
15	UKM	Komputer	Bahasa Inggris
16	UKM	Bahasa Inggris	Komputer
17	Bahasa Inggris		
18	Komputer	Bahasa Inggris	
19	Bahasa Inggris	Komputer	
20	UKM	Komputer	
21	UKM	Bahasa Inggris	

22	UKM		
23	Bahasa Inggris	UKM	
24	UKM	Komputer	Bahasa Inggris
25	UKM	Bahasa Inggris	Komputer
26	Bahasa Inggris		
27	UKM	Bahasa Inggris	
28	Bahasa Inggris		
29	Bahasa Inggris	UKM	
30	UKM	Komputer	Bahasa Inggris
31	UKM	Bahasa Inggris	Komputer
32	UKM		
33	Komputer	Bahasa Inggris	
34	UKM	Komputer	Bahasa Inggris
35	UKM	Bahasa Inggris	Komputer
36	UKM		
37	Komputer	Bahasa Inggris	
38	Bahasa Inggris	Komputer	
39	Bahasa Inggris	UKM	
40	UKM	Komputer	Bahasa Inggris
41	UKM	Bahasa Inggris	Komputer
42	Komputer	Bahasa Inggris	
43	Komputer		
44	Bahasa Inggris	UKM	
45	UKM	Komputer	Bahasa Inggris
46	UKM	Bahasa Inggris	Komputer
47	Komputer	Bahasa Inggris	
48	UKM	Komputer	Bahasa Inggris
49	Komputer	Bahasa Inggris	
50	UKM	Komputer	Bahasa Inggris

Setelah data mahasiswa diperoleh, selanjutnya adalah mengelompokkan mahasiswa berdasarkan data masing-masing. Namun sebelum hal ini dilakukan, agar mempermudah dalam analisa, maka dibuatlah tabel mahasiswa berdasarkan pilihan program masing-masing.

Tabel 3. *Program Tambahan Mahasiswa*

Mahasiswa	Program Tambahan yang Dipilih
1	Bahasa Inggris, UKM
2	UKM, Komputer, Bahasa Inggris
3	UKM, Bahasa Inggris, Komputer
4	UKM
5	Komputer, Bahasa Inggris
6	Bahasa Inggris, Komputer
7	UKM, Komputer
8	UKM, Bahasa Inggris
9	UKM
10	Bahasa Inggris, UKM
11	UKM, Komputer, Bahasa Inggris
12	UKM, Bahasa Inggris, Komputer
13	UKM
14	Komputer, Bahasa Inggris
15	UKM, Komputer, Bahasa Inggris
16	UKM, Bahasa Inggris, Komputer
17	Bahasa Inggris
18	Komputer, Bahasa Inggris
19	Bahasa Inggris, Komputer
20	UKM, Komputer

21	UKM, Bahasa Inggris
22	UKM
23	Bahasa Inggris, UKM
24	UKM, Komputer, Bahasa Inggris
25	UKM, Bahasa Inggris, Komputer
26	Bahasa Inggris
27	UKM, Bahasa Inggris
28	Bahasa Inggris
29	Bahasa Inggris, UKM
30	UKM, Komputer, Bahasa Inggris
31	UKM, Bahasa Inggris, Komputer
32	UKM
33	Komputer, Bahasa Inggris
34	UKM, Komptuer, Bahasa Inggris
35	UKM, Bahasa Inggris, Komputer
36	UKM
37	Komputer, Bahasa Inggris
38	Bahasa Inggris, Komputer
39	Bahasa Inggris, UKM
40	UKM, Komputer, Bahasa Inggris
41	UKM, Bahasa Inggris, Komputer
42	Komputer, Bahasa Inggris
43	Komputer
44	Bahasa Inggris, UKM
45	UKM, Komputer, Bahasa Inggris
46	UKM, Bahasa Inggris Komputer
47	Komputer, Bahasa Inggris
48	UKM, Komputer, Bahasa Inggris
49	Komputer, Bahasa Inggris
50	UKM, Komputer, Bahasa Inggris

Data transaksional di atas lalu direpresentasikan dalam bentuk seperti terlihat pada tabel 4 berikut ini :

Tabel 4. *Representasi Data Program Tambahan Mahasiswa*

Mahasiswa	Program Tambahan yang Dipilih
1	Bahasa Inggris
1	UKM
2	UKM
2	Komputer
2	Bahasa Inggris
3	UKM
3	Bahasa Inggris
3	Komputer
4	UKM
5	Komputer
5	Bahasa Inggris
6	Bahasa Inggris
6	Komputer
7	UKM
7	Komputer
8	UKM
8	Bahasa Inggris
9	UKM
10	Bahasa Inggris
10	UKM
11	UKM

11	Komputer
11	Bahasa Inggris
12	UKM
12	Bahasa Inggris
12	Komputer
13	UKM
14	Komputer
14	Bahasa Inggris
15	UKM
15	Komputer
15	Bahasa Inggris
16	UKM
16	Bahasa Inggris
16	Komputer
17	Bahasa Inggris
18	Komputer
18	Bahasa Inggris
19	Bahasa Inggris
19	Komputer
20	UKM
20	Komputer
21	UKM
21	Bahasa Inggris
22	UKM
23	Bahasa Inggris
24	UKM
24	Komputer
24	Bahasa Inggris
25	UKM
25	Bahasa Inggris
25	Komputer
26	Bahasa Inggris
27	UKM
27	Bahasa Inggris
28	Bahasa Inggris
29	Bahasa Inggris
29	UKM
30	UKM
30	Komputer
30	Bahasa Inggris
31	UKM
31	Bahasa Inggris
31	Komputer
32	UKM
33	Komputer
33	Bahasa Inggris
34	UKM
34	Komputer
34	Bahasa Inggris
35	UKM
35	Bahasa Inggris
35	Komputer
36	UKM
37	Komputer
37	Bahasa Inggris

38	Bahasa Inggris
38	Komputer
39	Bahasa Inggris
39	UKM
40	UKM
40	Komputer
40	Bahasa Inggris
41	UKM
41	Bahasa Inggris
41	Komputer
42	Komputer
42	Bahasa Inggris
43	Komputer
44	Bahasa Inggris
44	UKM
45	UKM
45	Komputer
45	Bahasa Inggris
46	UKM
46	Bahasa Inggris
46	Komputer
47	Komputer
47	Bahasa Inggris
48	UKM
48	Komputer
48	Bahasa Inggris
49	Komputer
49	Bahasa Inggris
50	UKM
50	Komputer
50	Bahasa Inggris

Pada pengujian ini, penulis menetapkan nilai *minimum support* sebesar 5% dan *minimum confidence* sebesar 10%. Alasan penulis menetapkan angka tersebut karena data yang penulis gunakan hanya 50. Sehingga data yang akan diproses oleh *software Orange* hanyalah *rule-rule* yang memenuhi kriteria di atas. Tabel 5 berikut ini menunjukkan calon *2-itemset* dari data pilihan program tambahan yang memenuhi ketentuan *minimum support* dan *minimum confidence* berdasarkan pada tabel 4.

Tabel 5. *Calon 2-itemset*

Kombinasi	Jumlah
Bahasa Inggris, UKM	8
Komputer, Bahasa Inggris	11
UKM, Komputer	2

Dari tabel 5 di atas, dapat disimpulkan jika ditetapkan nilai dari *threshold* di mana $\Phi = 2$, maka :
 $F_2 = \{(Bahasa\ Inggris,\ UKM), (Komputer,\ Bahasa\ Inggris), (UKM,\ Komputer)\}$.

Tabel 6. *Calon 3-itemset*

Kombinasi	Jumlah
UKM, Komputer, Bahasa Inggris	9
UKM, Bahasa Inggris, Komputer	9

Kombinasi dari *itemset* dalam F_2 dapat digabungkan menjadi calon *3-itemset*. *Itemset-itemset* dari F_2 yang dapat digabungkan adalah *itemset-itemset* yang memiliki kesamaan dalam *k-1 item* pertama. Calon *3-itemset* yang dapat dibentuk dari F_2 tampak pada tabel 6. Dengan demikian $F_3 = \{(UKM,\ Komputer,\ Bahasa\ Inggris)\}$.

4.5. Pembentukan Aturan Asosiasi

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiasi yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan asosiasi $A \rightarrow B$. Nilai *confidence* dari aturan $A \rightarrow B$ diperoleh dari rumus berikut.

$$Confidence = P(B | A) =$$

Jika memilih program UKM dan Komputer, maka akan memilih program Bahasa Inggris. Untuk mendapatkan nilai *confidence* dari aturan ini maka masukkan nilai total pilihan yang mengandung program UKM dan Komputer lalu dibagi dengan total pilihan yang mengandung program UKM dan Komputer. Rujukan dari rumus ini adalah pada tabel 7 dengan *3-itemset* dan tabel 6 dengan *2-itemset*.

Sementara untuk mendapatkan nilai *support* dari jika memilih program Bahasa Inggris, maka akan memilih program UKM merupakan hasil dari perhitungan total pilihan yang mengandung program Bahasa Inggris dan UKM lalu dibagi dengan total pilihan.

Aturan asosiasi yang terbentuk berdasarkan *minimum support* 5% dan *minimum confidence* 10% dapat dilihat pada tabel 7 berikut.

Tabel 7. Aturan Asosiasi 2-itemset

Aturan	Support	Confidence
Jika memilih Bahasa Inggris, maka memilih UKM	16%	19%
Jika memilih Komputer, maka memilih Bahasa Inggris	22%	34%
Jika memilih UKM, maka memilih Komputer	4%	6%

Dari tabel 7 di atas dapat disimpulkan nilai *support* dan *confidence* dengan acuan *2-itemset* yang memiliki nilai tertinggi adalah kombinasi antara buku Agama dan buku Penjaskes dengan total *support* \times *confidence* adalah 80%.

Sementara aturan asosiasi dengan *3-itemset* yang memenuhi kriteria *minimum support* dan *minimum confidence* dapat dilihat pada tabel 8 berikut.

Tabel 8. Aturan Asosiasi 3-itemset

Aturan	Support	Confidence
Jika memilih program UKM dan Komputer, maka memilih program Bahasa Inggris	18%	45%
Jika memilih program UKM dan Bahasa Inggris, maka memilih program Komputer	18%	112%

Dari tabel 8 di atas diperoleh 2 aturan asosiasi yang memenuhi syarat *minimum support* dan *minimum confidence*. Dapat pula disimpulkan bahwa semua kombinasi pemilihan program UKM, Bahasa Inggris dan Komputer di STIE-GK Muara Bulian memiliki tingkat kemungkinan pilihan secara bersamaan paling tinggi.

4.6. Hasil Pengujian pada Software

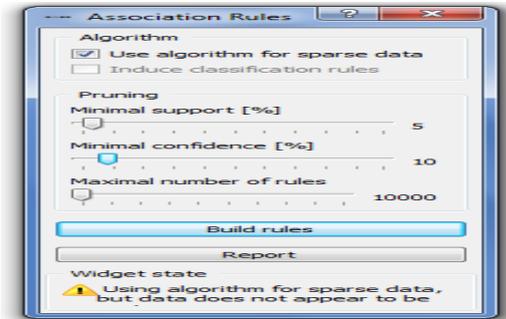
Pengujian terhadap hasil analisa, sangat penting dilakukan untuk menentukan dan memastikan apakah hasil analisa tersebut benar atau tidak.



Gambar 2. Proses Pembacaan Data

Gambar di atas memperlihatkan hubungan antara file yang telah diakses dan *rule* asosiasi yang akan dicari. Data yang telah diakses akan diproses oleh menu *Association Rules* dan akan ditampilkan melalui menu *Association Rules Explorer*.

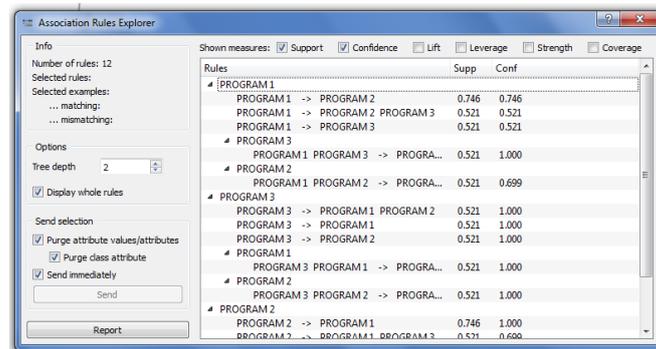
Setelah dihubungkan antara file dan *rule* asosiasi. Lakukan pengaturan nilai *minimum support* dan *minimum confidence*.



Gambar 3. Pengaturan Nilai Minimum Support dan Confidence

Gambar 3 memperlihatkan halaman pengaturan *minimum support* dan *minimum confidence*. Langkah-langkahnya adalah dengan mengklik kanan menu *Association Rules* → *Open* sehingga tampil sebagaimana gambar di atas.

Untuk melihat *rule-rule* yang terbentuk, klik kanan *Association Rules Explorer* dan pilih *Open*, maka *rule-rule* yang memenuhi kriteria *minim support* dan *minimum confidence* yang telah ditetapkan akan ditampilkan sebagaimana gambar 4 berikut.



Gambar 4. Rule Asosiasi yang Terbentuk

6. PENUTUP

6.1. Kesimpulan

Dari uraian pada bab-bab yang sudah dibahas sebelumnya dapat ditarik kesimpulan :

1. Sistem yang dibangun dapat membantu dalam menentukan pola pemilihan program tambahan yang umumnya masih dilakukan secara manual
2. Algoritma Apriori dengan *rule* asosiasinya dapat memberikan informasi efektif untuk menggambarkan proses yang terkait dengan pola pemilihan program tambahan
3. *Rule-rule* yang dihasilkan dapat digunakan sebagai referensi dalam pemilihan program tambahan di STIE-GK Muara Bulian.

6.2. Saran

Pengembangan sistem ini adalah dari sistem manual atau berdasarkan basis pengetahuan ke sistem komputerisasi yang dibangun atas dasar pertimbangan permasalahan-permasalahan yang timbul dari sistem yang lama. Berikut ini adalah saran yang dapat diperlihatkan untuk masa yang akan datang.

1. Pada penelitian ini, penulis mencoba salah satu teknik yang digunakan dalam menentukan pola pemilihan program tambahan yaitu Algoritma Apriori. Untuk mendapatkan hasil pola yang baik dapat digunakan beberapa atau penggabungan beberapa teknik prediksi. Oleh sebab itu perlunya

dilakukan penelitian lanjut untuk membandingkan hasil pola dan menentukan teori mana yang menghasilkan prediksi yang terbaik.

2. Untuk mendapatkan hasil pola pilihan program yang baik diperlukan sumber data yang lengkap. Oleh sebab itu diharapkan pada proses penyimpanan data program kemahasiswaan sudah dilakukan secara elektronik dengan demikian akan memudahkan dalam pencarian data yang dibutuhkan untuk melakukan proses penemuan *knowledge* seperti untuk menentukan pola pilihan program.
3. Walaupun aplikasi ini sangat mudah untuk dijalankan, namun perlu diasosiasikan cara-cara pengoperasiannya.

Daftar Rujukan

- [1] Anas, Azwar. (2016). Analisa Algoritma Apriori untuk Mendapatkan Pola Peminjaman Buku Perpustakaan SMPN 3 Batanghari. *Jurnal Ilmiah Media SISFO*. 10 (2).
- [2] Erwin. (2009). Analisis *MarketBasket* dengan Algoritma Apriori dan *FP-Growth*. *Jurnal Generic*. 4 (2).
- [3] Elsa.W & Kania.E.D. (2014).Implementasi *Association Rule* terhadap Penyusunan *Layout* Makanan dan Penentuan Paket Makanan Hemat di RM Roso Echo dengan Algoritma Apriori. *Jurnal KOMPUTA*. 3 (2).
- [4] Sri R.S. (2014). Implementasi Data Mining pada Penjualan Tiket Pesawat menggunakan Algoritma Apriori. *Jurnal Pelita Informatika Budi Darma*. VII (1).
- [5] Robi.Y & Riri.K. (2015).Implementasi Data Mining dengan Metode Algoritma Apriori dalam Menentukan Pola Pembelian Obat. *Citec Journal*. 2 (2).
- [6] Wirdah. C. (2016).Penggunaan Algoritma Apriori Data Mining untuk Mengetahui Tingkat Kesetiaan Konsumen (*Brand Loyalty*) terhadap Merek Kendaraan Bermotor (Studi Kasus Dealer Honda Rumbai).*Jurnal Teknologi Informasi dan Komunikasi Digital Zone*.7 (1).
- [7] Dewi.K.P. (2013). Implementasi Data Mining pada Penjualan Produk Elelektronik dengan Algoritma Apriori (Studi Kasus: Kredit Plus). *Jurnal Pelita Informatika Budi Darma*. IV (3).
- [8] Kennedy. T, Hoga.S & Bobby.R. (2013).Implementasi Data Mining Algoritma Apriori pada Sistem Persediaan Alat-alat Kesehatan.*Jurnal Informasi dan Teknologi Ilmiah*.1 (1).
- [9] Nugroho W, Rully. A.H & Ahmad.M. (2012).Pengembangan Sistem Rekomendasi Penelusuran Buku dengan Penggalian *Association Rule* menggunakan Algoritma Apriori (Studi Kasus Badan Perpustakaan dan Kearsipan Provinsi Jawa Timur).*Jurnal Teknik ITS*.1 (1).
- [10] Widodo. (2008, Mei).*Prediksi Mata Kuliah Pilihan dengan Aturan Asosiasi*.Makalah disajikan dalam Konferensi dan Temu Nasional Teknologi Informasi dan Komunikasi untuk Indonesia. Jakarta: Universitas Negeri Jakarta.