

# Web Usage Mining, Pattern Discovery dan Log File

Tri Suratno<sup>a</sup>, Toni Prahasto<sup>b</sup>, Adian Fatchur Rochim<sup>b</sup>,

<sup>a</sup>Jurusan Pertanian, Universitas Jambi, Jambi

<sup>a</sup>Magister Sistem Informasi Program Pascasarjana, Universitas Diponegoro, Semarang

## Abstract

Analysis of data to access the server can provide significant and useful information for performance improvement, restructuring and improving the effectiveness of a web site. Data mining is one of the most effective way to detect a series of patterns of information from large amounts of data. Application of data mining on Internet use called web mining is a set of data mining techniques are used for the web. Web mining technologies and data mining is a combination of web, which is the integration of technology resources extracted from the information world wide web as the implications of the web resources of interest to know the value of the model extraction Potential use of data mining algorithms over a variety of observational data for identify patterns of web resources. Target analysis of web mining is the data from the web, such as data visitors access, web page structure, and format web pages. Target analysis of the study is that web mining web usage mining using association rules on the website [www.faperta.unja.ac.id](http://www.faperta.unja.ac.id) obtained from the log file that is used to discover the navigation patterns, and discover the rules of the association between a combination of items. To determine the pattern of visits in a web, and to identify what pages are frequented by visitors of a website, which can be used to improve website design and recommend the display, as well as the links are often used by visitors, so the quality of services from website [www.faperta.unja.ac.id](http://www.faperta.unja.ac.id) can be provided effectively and efficiently.

**Keywords :** Web Mining; Web usage mining; Data mining; Log file; Website; World wide web

## 1. Pendahuluan

Penggalan data merupakan salah satu cara yang cukup efektif untuk mengetahui adanya serangkaian pola informasi dari sejumlah besar data yang ada. Pola informasi yang didapat akan menjadi sangat berarti apabila bersifat implicit (belum diketahui sebelumnya), dan bermanfaat. *Data mining* memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik, basis datadan juga *information retrieval* (Berry and Linoff, 2004)

Penerapan *data mining* pada penggunaan *internet* yang disebut *web data mining* telah menjadi fokus dari banyak proyek dan penelitian akhir-akhir ini, *webmining* adalah seperangkat teknik-teknik *datamining* yang digunakan untuk *web*. Teknologi *web mining* dan *data mining* adalah kombinasi dari *web*, yang merupakan integrasi teknologi sumber daya yang di ekstrak dari informasi *www* (*world wide web*) sebagai bagian implikasi dari sumberdaya *web* yang menarik untuk mengetahui nilai potesial dari model ekstraksi penggunaan berulang berbagai algoritma *data mining* dari data pengamatan untuk mengidentifikasi pola dari sumber daya *web*, perbedaan data mining dan web mining terletak pada target yang dianalisis.

*Data mining* umumnya menganalisa data yang berasal dari OLTP (*Online Transactional Process*) dan data transaksi lainnya. Sedangkan *web mining* target analisisnya adalah data dari *web*, seperti data akses pengunjung, struktur halaman *web*, format halaman web dan sebagainya. Berdasarkan target analisisnya *webmining* dibagi menjadi tiga yaitu *web content mining*, *web structure mining*, dan *web usage mining*. *Webusagemining* diterapkan dengan banyak masalah dunia nyata untuk menemukan pola navigasi *user* yang menarik untuk perbaikan desain situs *web* dengan membuat topik

tambahan atau rekomendasi pengguna atau mengamati perilaku pelanggan.

Aturan asosiasi menjadi salah satu fungsionalitas yang paling menarik dalam penggalan data. Dikatakan menarik karena sudah banyak dipakai dalam kehidupan manusia sehari-hari, aturan asosiasi akan memberikan gambaran mengenai sejumlah atribut, atau sifat tertentu yang sering muncul bersamaan dalam kumpulan data yang diberikan. Analisis asosiasi atau *association rule mining* adalah teknik *data mining* untuk menemukan aturan asosiasi antara suatu kombinasi item.

Tujuan penelitian ini adalah aplikasi *web usage mining* menggunakan aturan asosiasi digunakan untuk mengetahui pola kunjungan di dalam sebuah *web*, untuk mengidentifikasi halaman apa yang sering dikunjungi oleh pengunjung sebuah *website*. Studi kasus dilakukan pada penerapan *web usage mining* untuk menemukan pola akses kunjungan di *website* Fakultas Pertanian Universitas Jambi, dengan memanfaatkan *log* dari penggunaan *website*, kemudian hasil dari *log* digunakan untuk mengetahui pola dan minat pengguna di *website* Fakultas Pertanian Universitas Jambi.

## 2. Kerangka Teori

### 2.1. Web Mining

*Web* merupakan kumpulan data dan informasi yang sangat berpotensi untuk dilakukan penggalan (*mining*) agar menghasilkan pengetahuan (*knowledge*) yang dapat berguna bagi masyarakat maupun pihak-pihak tertentu. Data dan informasi yang tersimpan di dalam *web* memiliki karakteristik yang berbeda dengan data yang tersimpan dalam penyimpanan konvensional. *Web* memiliki ukuran yang terlalu besar sehingga tidak terlalu efektif jika diterapkan dengan menggunakan data *warehouse* dan *data mining* biasa. Ukuran data dalam *web* mencapai ribuan

terabyte dan akan terus berkembang. Saat ini begitu banyak perusahaan dan organisasi yang mempublikasikan berbagai informasi perusahaan di sebuah *website*. Dengan ukuran yang begitu besar, akan sangat sulit jika harus membangun sebuah data *warehouse* yang akan menyimpan data dan informasi tersebut.

Tingkat kompleksitas dari halaman – halaman *web* jauh lebih tinggi dibanding dokumen dalam format teks biasa. Halaman *web* memiliki struktur yang sangat beragam. Apalagi jika dilihat dari isi atau *content* yang disajikan di halaman *web*, memiliki bahasa, gaya penulisan, struktur penulisan dan tampilan yang beragam.

*Web mining* merupakan ekstraksi pola-pola penting dan bermanfaat namun tersimpan secara implisit pada kumpulan data yang relatif besar pada layanan *world wide web*. *Web mining* terdiri atas tiga bagian yaitu: *web content mining*, *web structure mining*, dan *web usage mining* (Liu, 2007, Tyagi et al., 2010)).

*Web content mining* adalah suatu proses otomatis untuk menemukan informasi yang berguna dari dokumen atau data. Pada prinsipnya teknik ini mengekstraksi kata kunci yang terkandung pada dokumen. Isi data *web* antara lain dapat berupa teks, citra, audio, video, metadata, dan *hyperlink*. Ada dua strategi yang umum digunakan: pertama langsung melakukan *mining* terhadap data, dan kedua melakukan pencarian serta *improve* hasil pencarian seperti layaknya *search engine* (Srivastava et al, 2000). *Web structure mining* adalah teknik yang digunakan untuk menemukan struktur link dari *hyperlink* dan membangun rangkuman *website* dan halaman *web*. Salah satu manfaatnya adalah untuk menentukan *pagerank* pada suatu halaman *web*.

*Web usage mining* adalah teknik untuk mengenali perilaku pelanggan dan struktur *web* melalui informasi yang diperoleh dari *log*, *click stream*, *cookies*, dan *query*. Manfaat *web usage mining* adalah untuk kustomisasi halaman berdasarkan profil pengguna, menentukan ketertarikan pelanggan terhadap produk tertentu, dan menentukan target market yang sesuai (Devi and Sreevani, 2010).

## 2.2. Struktur WebLog

Server *log* adalah teks biasa ASCII *file*, yang independen dari *platform server*. *Log Web* adalah file dimana *Webserver* menulis informasi setiap kali permintaan pengguna sumber daya dari situs tertentu. *weblogfile* memiliki format yang standar, sehingga dapat mempermudah dalam proses *mining*. Format standar dari sebuah *web access log* (Consortium, 1995) adalah [remotehost rfc931 authuser [date] "request" status bytes], yaitu:

1. Remoteaddress Internet Protocol (IP) atau nama domain: Sebuah alamat IP adalah host 32-bit alamat yang sudah ditetapkan oleh Protokol *Internet*; nama *domain* yang digunakan untuk menentukan alamat *Internet* yang unik untuk setiap *host* di *internet*. Satu alamat IP biasanya didefinisikan untuk satu nama *domain*.
2. rfc931 adalah nama *log* dari user.
3. authuser adalah nama user.
4. date adalah tanggal dan waktu request.

5. "request" adalah halaman yang diminta oleh user beserta metodenya .
6. status adalah kode HTTP yang dikirimkan kembali ke user.
7. bytes adalah jumlah byte dokumen yang dikirimkan ke user.

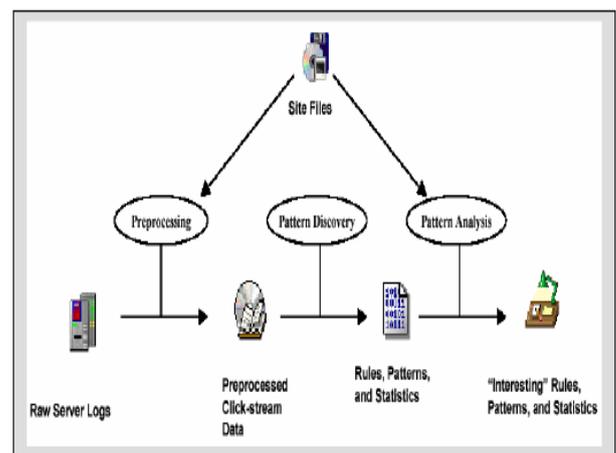
Berikut ini contoh sebuah *web access log*:  
66.249.65.107 - - [08/Oct/2007:04:54:20 -0400] "GET /support.html HTTP/1.1" 200 11179 "-" "Mozilla/5.0 (compatible; Googlebot/2.1;+http://www.google.com/bot.html)".

## 2.3. Web Usage Mining

*Web* sangat berkaitan erat dengan sebuah *web server*, yaitu suatu *software server* yang memiliki tugas utama melayani dan memenuhi permintaan halaman *web* oleh user. Selain itu, *web server* juga akan mencatat setiap aktivitas yang dilakukan oleh user tersebut ke dalam sebuah *file* yang sering disebut *webaccess log*. Hasil catatan aktivitas tersebut yang menjadi sumber data utama dalam *web usage mining*. Dari sebuah *web access log*, dapat diketahui beberapa informasi mengenai pola akses dan Perilaku pengguna dalam mengakses halaman *web*.

## 2.4. Proses Web Usage Mining

Secara garis besar, proses *web usage mining* terbagi menjadi 3 (tiga) tahap , yaitu *preprocessing*, *pattern discovery* dan *pattern analysis* (Srivastava dkk, 2000), digambarkan mengenai proses *webusage mining* seperti terlihat pada gambar 1 berikut ini :



Gambar 1. Proses web usage mining (Srivastava et al.,2000)

### 2.4.1. Tahap Preprocessing

Tahapan ini merupakan proses yang pertama kali dilakukan dari keseluruhan proses *web usage mining*. Tahapan ini penting dilakukan untuk melakukan standarisasi data dan juga menghilangkan bagian – bagian data tertentu yang tidak diperlukan dalam proses *mining*.

Pada tahap *preprocessing*, data *clickstream* dibersihkan dan dibagi menjadi satu set transaksi pengguna mewakili aktivitas dari tiap pengguna yang berbeda selama kunjungan ke situs. Pada tahap penemuan pola, statistik, dan basis data, dilakukan untuk mendapatkan pola-pola tersembunyi yang mencerminkan perilaku khas pengguna. Tahap akhir dari *preprocessing* dapat digunakan sebagai

masukan untuk aplikasi seperti rekomendasi, alat-alat visualisasi, dan analisis web dan alat laporan (Liu 2007).

#### 2.4.2. Pattern Discovery

Tahap yang kedua dari *webusage mining* adalah pencarian pola akses yang dilakukan oleh *user*. Tahap ini merupakan tahap yang sangat penting dan sangat menentukan keluaran dari proses *usage mining*. Pada tahap ini dikenal beberapa algoritma dan teknik, antara lain *statistical analysis, association rules, clustering, classification, sequential pattern, dependency modeling*

#### 2.4.3. Pattern Analysis

*Pattern analysis* merupakan tahap terakhir dalam *webusage mining*. Hasil *pattern analysis* dapat dijadikan suatu rekomendasi misalnya keputusan untuk mengubah tampilan suatu *website*, melakukan optimasi navigasi *website*, meningkatkan kemampuan *website* dengan melakukan *caching* halaman – halaman tertentu yang sering dikunjungi.

#### 2.4.4. Aturan Asosiasi

Analisis asosiasi atau *association rule mining* adalah teknik *data mining* untuk menemukan aturan asosiasi antara suatu kombinasi item. Contoh dari aturan asosiasi dari analisa pembelian di suatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seorang pelanggan membeli roti bersamaan dengan susu. Dengan pengetahuan tersebut pemilik pasar swalayan dapat mengatur penempatan barangnya atau merancang kampanye pemasaran dengan memakai kupon diskon untuk kombinasi barang tertentu, karena analisis asosiasi menjadi terkenal karena aplikasinya untuk menganalisa isi keranjang belanja di pasar swalayan, analisis asosiasi juga sering disebut dengan istilah *market basket analysis* (Kusrini dan Emha 2009).

Analisis asosiasi dikenal juga sebagai salah satu teknik *data mining* yang menjadi dasar dari berbagai teknik *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 persentase kombinasi item tersebut dalam database dan *confidence* (nilai kepastian) yaitu kuatnya hubungan antar item dalam aturan asosiasi (Kusrini dan Emha, 2009), Aturan asosiasi biasanya dinyatakan dalam bentuk : {roti, mentega} -> {susu} (*support* = 40%, *confidence* = 50%), yaitu "50% dari transaksi di database yang memuat item roti dan mentega juga memuat item susu. Sedangkan 40% dari seluruh transaksi yang ada di database memuat ketiga item itu." yang mempunyai makna : "Seorang konsumen yang membeli roti dan mentega punyai kemungkinan 50% untuk juga membeli susu. Aturan ini cukup signifikan karena mewakili 40% dari catatan transaksi selama ini."

Analisis asosiasi didefinisikan suatu proses untuk menemukan semua aturan asosiasi yang memenuhi syarat minimum untuk *support* (*minimum support*) dan syarat minimum untuk *confidence* (*minimum confidence*, (Kusrini dan Emha, 2009).

Metodologi dasar analisis asosiasi terbagi menjadi dua tahap :

##### a. Analisa pola frekuensi tinggi

Tahap ini mencari kombinasi item yang memenuhi syarat minimum dari nilai *support* dalam database. Nilai *support* sebuah item diperoleh dengan persamaan (1) :

$$\text{Support (A)} = \frac{\text{Jumlah transaksi mengandung A}}{\text{Total Transaksi}} \dots\dots\dots (1)$$

Sedangkan nilai *support* dari 2 item diperoleh dari persamaan (2) :

$$\text{Support (A,B)} = P(A \cap B) = \frac{\text{Jumlah transaksi mengandung A dan B}}{\text{Total Transaksi}} \dots\dots\dots (2)$$

##### b. Pembentukan aturan asosiasi

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

Nilai *confidence* dari aturan  $A \rightarrow B$  diperoleh dari Persamaan (3) :

$$\text{Confidence} = P(B|A) = \frac{\text{Jumlah transaksi mengandung A dan B}}{\text{Jumlah Transaksi Mengandung A}} \dots\dots\dots (3)$$

### 2.7 Algoritma Apriori

Algoritma apriori adalah sebuah algoritma klasik untuk mengetahui aturan asosiasi. Algoritma apriori dirancang untuk beroperasi pada database yang berisi transaksi misalnya, koleksi barang yang dibeli oleh pelanggan.

Algoritma apriori digunakan untuk menemukan pola frekuensi tinggi. Pola frekuensi tinggi digunakan untuk menentukan pola-pola item di dalam suatu database yang memiliki frekuensi atau *support* di atas ambang batas tertentu yang dinamakan *minimum support*.

Ide dasar dari algoritma apriori dimulai dengan mengembangkan *frequent itemset* dengan menggunakan satu item dan secara rekursif mengembangkan *frequent itemset* dengan dua item, tiga item dan seterusnya hingga *frequent item set* dengan semua ukuran (Santoso, 2007).

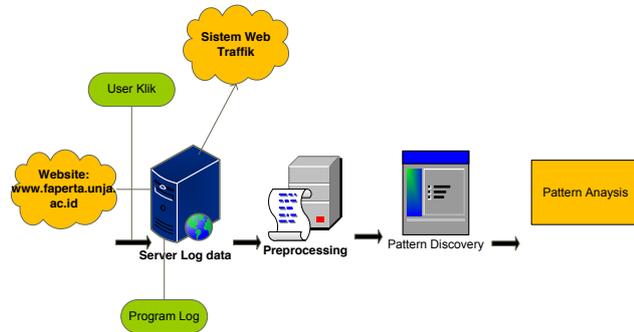
Algoritma Apriori dibagi menjadi beberapa tahap yang disebut iterasi.

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 subsetnya yang berisi k-1 item tidak termasuk dalam pola frekuensi tinggi dengan panjang k-1.
2. Perhitungan *support* dari tiap kandidat k-itemset. *Support* dari tiap kandidat k-itemset didapat dengan menscan database untuk menghitung jumlah transaksi yang memuat semua item di dalam kandidat k-itemset tersebut. Ini juga merupakan ciri dari algoritma apriori dimana diperlukan perhitungan 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*.

Bila tidak didapat pola frekuensi tinggi maka seluruh proses dihentikan. Bila tidak, maka k tambah satu dan kembali ke bagian 1.

### 3. Metodologi

Desain arsitektur yang dibangun dalam penelitian menghasilkan dua hasil yaitu system web trafik dan aturan asosiasi dari pola kunjungan website [www.faperta.unja.ac.id](http://www.faperta.unja.ac.id) yang dapat dilihat pada gambar 2.



Gambar 3.1. Arsitektur system

Alur dari arsitektur system dalam penelitian ini yaitu, user mengklik website [www.faperta.unja.ac.id](http://www.faperta.unja.ac.id), hasil klik pengunjung disimpan di dalam database menggunakan *browser sniffing* (metode tanpa mensubmit form), selanjutnya data yang telah disimpan di database ditampilkan dalam bentuk web trafik secara otomatis, yang dapat mengetahui aktivitas user yang berkunjung di website, setelah data ditampilkan melalui web trafik, langkah selanjutnya yaitu:

#### 1. Tahapan Preprocessing

##### a. Pembersihan data

Data yang digunakan perlu dilakukan pembersihan agar data yang akan diolah benar-benar sesuai dengan yang dibutuhkan. Pembersihan ini dilakukan guna meningkatkan kerja dari proses *mining*. Data *logfile* dikelompokkan berdasarkan variable penelitian dan dikelompokkan berdasarkan tanggal akses, dan apabila user mengklik pada variable yang sama dianggap mewakili satu transaksi. Cara pembersihan dilakukan dengan mengikutsertakan atribut yang tidak digunakan dan menghapus data yang tidak lengkap isinya, yaitu alamat ip yang hanya mempunyai satu transaksi, alamat ip yang mengakses pada kategori yang sama dihitung satu transaksi, serta menghapus alamat url yang tidak terdapat pada lampiran 2, dalam penelitian ini peneliti mengambil data terpenting dari *log file* untuk disesuaikan dengan aplikasi yang dibangun dengan hanya mengambil data berdasarkan halaman yang dikunjungi berdasarkan alamat ip pengunjung.

##### b. Seleksi data

Pada tahapan seleksi data ini, peneliti menyeleksi data, hanya data yang digunakan dalam variabel penelitian untuk diseleksi..

##### c. Tahap Transformasi Data

Tahap selanjutnya dari proses *web mining* yaitu melakukan transformasi data, yang digunakan untuk mengubah *log file* ke transformasi data number, yang dapat dibaca oleh program Matlab, selanjutnya peneliti mengelompokkan konten yang diklik dengan

struktur halaman *link* yang telah ditentukan dalam variable penelitian yaitu:

Main Menu	= 1	Beasiswa	= 6
Program Studi	= 2	weblinks	= 7
Campus News	= 3	komentar	= 8
E-journal	= 4	menu register	= 9
Seputar Pertanian	= 5		

#### 2. Pattern Discovery

Tahap selanjutnya dari proses *web usage mining* adalah pencarian pola akses yang dilakukan oleh pengunjung, untuk menemukan semua hubungan dan korelasi diantara item-item data, dimana kehadiran satu item dalam transaksi menunjukkan kehadiran item yang lainnya, *pattern discovery* pada penelitian ini menggunakan aturan asosiasi menggunakan persamaan *support* dan *confidence* untuk mendapatkan item yang akan dikombinasikan, untuk proses algoritma apriori di mulai dengan langkah, menentukan file yang akan di proses, kemudian menentukan *support dan confidence* untuk membangkitkan aturan yang diinginkan yaitu kaidah aturan asosiasi (“jika –maka”).

#### 3. Pattern analysis

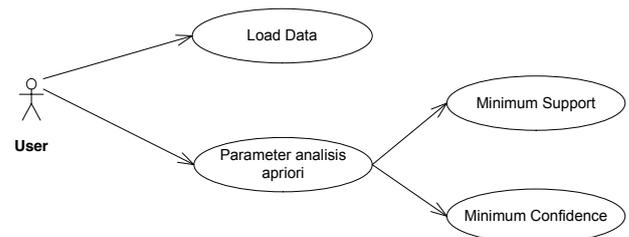
*Pattern analysis* merupakan tahap terakhir dalam *web usage mining*, pada tahap ini, dilakukan proses *visualisasi* hasil analisis pola yang telah dilakukan pada langkah sebelumnya. Penyajian data menjadi hal yang penting dalam langkah ini, dimana penyajian data tentunya tergantung pada kebutuhan *user* dan bisnis. Dari hasil *visualisasi* tersebut, dapat dilakukan suatu keputusan misalnya keputusan untuk mengubah tampilan suatu website, melakukan *optimasi navigasi website*, meningkatkan kemampuan *website* dengan melakukan *caching* halaman – halaman tertentu yang sering dikunjungi.

#### 3.2. Rancangan Perangkat Lunak

*Unified Modeling Language* (UML). Merupakan system arsitektur yang bekerja dalam object Oriented Analysis Design (OOAD) dengan satu bahasa yang konsisten untuk menentukan, visualisasi, mengkonstruksi, dan mendokumentasikan yang terdapat dalam system software. Adapun UML yang dibuat mencakup diagram *use case*, diagram aktifitas, dan diagram sekuen.

##### 1. Diagram Use Case

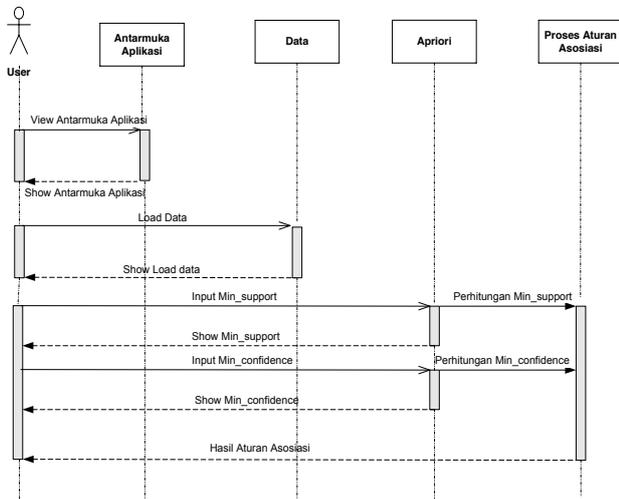
Diagram *use case* menggambarkan fungsionalitas yang diharapkan dari sebuah sistem. Sebuah *use case* mempresentasikan sebuah interaksi antara *actor* dengan sistem, yang digambarkan dalam bentuk diagram *use case* seperti pada gambar 3.



Gambar 3. Diagram Use Case

## 2. Diagram Sequence

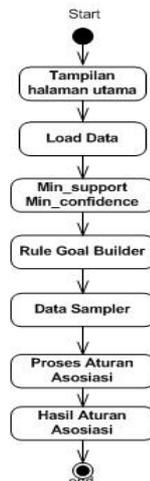
Diagram *sequence* menggambarkan interaksi antar objek di dalam dan di sekitar sistem. Diagram *sequence* digunakan untuk menggambarkan skenario atau rangkaian langkah-langkah yang dilakukan sebagai respons dari sebuah *event* untuk menghasilkan *output* tertentu *sequence* diagram untuk aplikasi yang dikembangkan seperti pada gambar 4.



Gambar 4. Diagram Sequence

## 3. Diagram Aktivitas

Diagram aktivitas menggambarkan berbagai aliran aktifitas dalam sistem yang sedang dirancang, bagaimana masing-masing berawal, keputusan yang mungkin terjadi, dan bagaimana berakhir. Diagram aktifitas menggambarkan proses parallel yang mungkin terjadi pada beberapa eksekusi. Diagram aktivitas dari aplikasi yang dikembangkan dapat dilihat pada gambar 5.



Gambar 5. Diagram Aktivitas

## 4. Hasil dan Pembahasan

Sistem yang di kembangkan dalam penelitian ini dilakukan dengan membuat program metode *browser sniffing* (metode tanpa mensubmit *form*) yang digunakan menangkap alamat *log* yang dihasilkan oleh user secara otomatis, menggunakan *php* dan *mysql* yang disisipkan

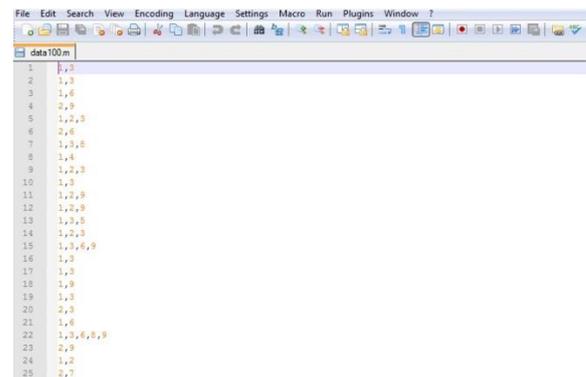
di file website [www.faperta.unja.ac.id](http://www.faperta.unja.ac.id), data pengunjung website yang disimpan dalam database *mysql* selanjutnya di proses secara manual menggunakan program bantu Microsoft Excel untuk dilakukan proses seleksi data, kemudian setelah data diseleksi, kemudian diproses menggunakan program Matlab 2008a.

### 4.1. Aturan Asosiasi

Pada aturan Asosiasi digunakan untuk mencari hubungan kedekatan antara item yang satu dengan yang lain dengan melakukan kombinasi untuk menemukan pola hubungan antar item. Pada aturan asosiasi ini membutuhkan dua nilai masukan yaitu *minimum confidence* dan *minimum support*.

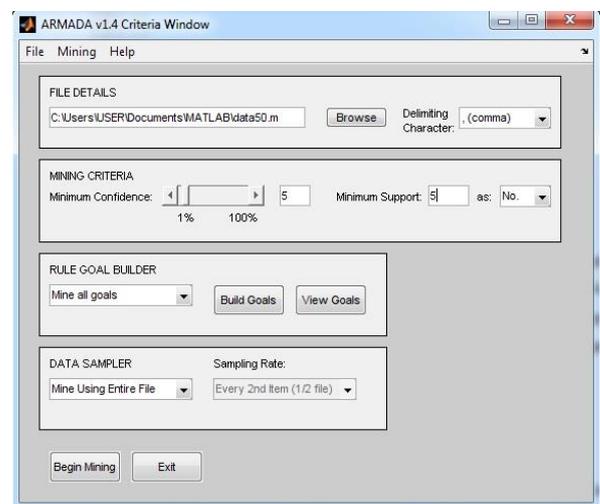
Untuk nilai *minimum confidence* maupun nilai *support* adalah bebas, tidak ada aturan baku untuk memasukan suatu nilai, tetapi harus diperhatikan bahwa semakin besar nilai yang dimasukan, kemungkinan untuk menemukan kombinasi item yang muncul secara bersamaan semakin kecil.

Data log yang didapatkan dari hasil proses browser sniffing, dilakukan pengkodean sesuai dengan variable yang telah ditentukan. Setelah dilakukan pengkodean selanjutnya mentransformasi data log yang didapat sesuai kode yang ada, seperti gambar 5.



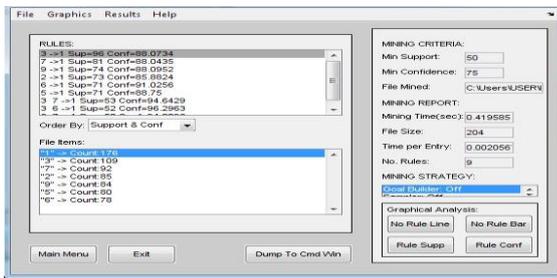
Gambar 5. Tampilan data hasil transformasi

Setelah data di transformasi secara manual dengan melihat pengelompokannya, kemudian data di olah menggunakan aturan asosiasi terlihat pada gambar 6.



Gambar 6. Tampilan aturan asosiasi

Data yang dihasilkan dari 204 data dengan memasukan *minimum support* 50% dan *minimum confidence* 75 % menghasilkan 9 rule, yang terdiri dari kombinasi 2 itemset sebanyak 6 rule, kombinasi 3 itemset sebanyak 3 rule, yang dapat dilihat pada gambar 7.



Gambar 7. Tampilan hasil aturan asosiasi untuk support 50 % dan confidence 75%

Hasil rekomendasi aturan asosiasi minimum support 50% dan minimum confidence 75% untuk penempatan item modul pada website [www.faperta.unja.ac.id](http://www.faperta.unja.ac.id) seperti terlihat pada gambar 8.



Gambar 8. Rekomendasi min\_sup 50% dan min\_conf 75%

### 5. Kesimpulan

Data *log file* yang didapatkan dari klik pengunjung [www.faperta.unja.ac.id](http://www.faperta.unja.ac.id), dapat digunakan untuk mengetahui jumlah pengunjung dari website dengan menampilkan web trafik. *Pattern discovery* pada *log file* [www.faperta.unja.ac.id](http://www.faperta.unja.ac.id), menggunakan aturan asosiasi dapat digunakan sebagai rekomendasi untuk mengatur penempatan item atau modul yang ada pada website [www.faperta.unja.ac.id](http://www.faperta.unja.ac.id). Penempatan item atau modul website [www.faperta.unja.ac.id](http://www.faperta.unja.ac.id), yang direkomendasikan yaitu untuk bagian kiri terdiri dari (1) main menu, (2) campus news, (3) weblinks, (4) beasiswa, (5) menu register, dan untuk bagian kanan terdiri dari (6) program studi, (7) seputar pertanian, (8) komentar anda, (9) e-journal/ebook.

### Daftar Pustaka

Berry, M., Linoff, G., 2004. Data Mining Techniques For Marketing, Sales, Customer Relationship Management Second Edition, Wiley Publishing, Inc.

Consortium, 1995. The Common Log File format [http://www.w3.org/Daemon/User/Config/Logging.html#common\\_logfile\\_format](http://www.w3.org/Daemon/User/Config/Logging.html#common_logfile_format).

Devi, N., Sreevani, 2010. Dynamic modelling approach for web usage mining using open web resources. *International Journal of Engineering Science and Technology*, 2(10): 5605-5610.

Kusrini, L., Emha, T., 2009. Algoritma Data mining, Andi Offset Yogyakarta.

Liu, B. 2007. Web Data Mining Exploring, Hyperlink, Contents, and Usage data, New York, Springer Berlin Heidelberg

Santoso, B., 2007. Data Mining: Teknik Pemanfaatan Data Untuk Keperluan Bisnis, Yogyakarta: Graha Ilmu

Srivastava, J., Cooley, R., Deshpande, M., Tan, P., 2000. Web usage mining: discovery and applications of usage patterns from web data. *ACM SIGKDD*, 1(2): 1-12.

Tyagi, N., Kumar; Solanki; Wadhwa, M., 2010. Analysis of server log by web usage mining for website improvement. *International Journal of Computer Science*, Vol. 7, Issue 4, No 8.