

ANALISIS PERBANDINGAN ALGORITMA APRIORI DAN FP-GROWTH PADA TRANSAKSI KOPERASI

ABSTRAK

Teknik data mining mampu memprediksi tren dan membantu proses pengambilan keputusan bisnis yang penting bagi pemilik usaha. Penggunaan sistem yang telah ada akan menimbulkan permasalahan baru, yaitu peningkatan data transaksi. Algoritma Apriori merupakan algoritma yang paling populer digunakan untuk analisis market basket karena mudah dipahami dan diimplementasikan. Algoritma FP-Growth juga termasuk salah satu teknik association rule untuk analisis market basket. Berdasarkan data transaksi (studi kasus koperasi Uber-Mart Bekasi), peneliti membandingkan Algoritma Apriori dan FP-Growth untuk mendapatkan informasi tentang asosiasi antar produk dari suatu data transaksi dari item-set yang sering muncul secara bersamaan (market basket analysis). Dari hasil penelitian, Algoritma Apriori membutuhkan waktu komputasi yang lama dan membutuhkan alokasi memori yang besar untuk melakukan pencarian item-sets. Hal ini disebabkan pemindaian data yang dilakukan secara berulang-ulang. Algoritma FP-Growth hanya melakukan dua kali pemindaian data dan mampu memampatkan data transaksi yang memiliki item yang sama, sehingga waktu yang dibutuhkan lebih singkat dan alokasi memori lebih kecil.

Kata Kunci: Data Mining, Association Rule, Algoritma Apriori, FP-Growth, Koperasi, Korelasi Produk

Dwi Widiastuti¹
Nelly Sofi²

Jurusan Sistem Informasi,
Ilmu Komputer dan Teknologi Informasi
Universitas Gunadarma
Jurusan Teknik Informatika,
Teknologi Industri Universitas Gunadarma

¹dwidiastuti@staff.gunadarma.ac.id,
²nelly_sofi@staff.gunadarma.ac.id

PENDAHULUAN

Koperasi adalah kegiatan ekonomi rakyat yang berskala kecil dengan bidang usaha yang secara mayoritas merupakan kegiatan usaha kecil dan perlu dilindungi untuk mencegah persaingan usaha yang tidak sehat. Usaha untuk mencapai tingkat margin yang lebih tinggi, merupakan motivasi bagi pemilik untuk meningkatkan mutu pelayanan dan diversifikasi produk yang ditawarkan. Salah satu usaha tersebut adalah kemampuan untuk memahami konsumen.

Oleh sebab itu koperasi akan sangat membutuhkan sistem informasi yang dapat menunjang masalah tersebut, sehingga dapat mempermudah berbagai proses kelancaran bertransaksi. Pada penelitian ini dirancang suatu sistem informasi (studi kasus pada Koperasi Uber-Mart Bekasi) dengan membandingkan Algoritma Apriori dan FP-Growth. Algoritma tersebut akan dapat dituangkan menjadi suatu aplikasi dengan bahasa pemrograman berbasis web, salah satunya PHP.

Algoritma Apriori adalah sebuah algoritma untuk mencari kombinasi *item-set* yang mempunyai suatu nilai keseringan tertentu sesuai kriteria atau *filter* yang diinginkan. *Frequent Pattern Growth (FP-Growth)* adalah salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent item set*) dalam sebuah kumpulan data.

Data mining mengidentifikasi fakta-fakta atau kesimpulan-kesimpulan yang disarankan berdasarkan penyaringan melalui data untuk menjelajahi pola-pola atau anomali-anomali data. Menurut Turban, Rainer, dan Potter (2005), salah satu fungsi data mining adalah *association rule*. *Association rule* mendukung pengambilan keputusan dalam bidang pemasaran, misalnya untuk mengetahui pola pembelian pelanggan, penentuan tata

letak barang dan lain-lain. *Association Rule* diimplementasikan dengan menggunakan Algoritma Apriori dan *FP-Growth*.

Association Rule, Algoritma Apriori, Algoritma FP-Growth

Menurut Han (2006), aturan asosiasi (*association rule*) adalah metode data mining untuk mencari suatu hubungan yang menunjukkan kondisi di dalam satu set data, yang beberapa nilai atribut akan muncul secara bersamaan. Pemilihan produk tidak diperlukan untuk menjalankan analisis keranjang. Semua produk dipertimbangkan, dan perangkat lunak data mining akan menentukan produk yang paling utama. Menentukan produk-produk apa saja yang cenderung sering dibeli oleh para pelanggan dengan mengambil asosiasi seketika maka aturan asosiasi ini dinamakan *market basket analysis* atau analisis afinitas (*affinity analysis*) berkenaan dengan studi tentang 'apa-bersama-apa'. Fasilitas ini meningkatkan pembelian dan membantu pelanggan yang ingin membeli barang agar tidak lupa untuk membeli "pasangannya". Contoh, jika seorang pelanggan membeli roti, maka mereka kemungkinan akan membeli selai [Fanani, 2010].

Dalam menentukan suatu *association rule*, ada 2 nilai *interestingness measure* (ukuran ketertarikan), yaitu *support* dan *confidence* [(Ulmer, David, 2002)]:

a. Support (nilai penunjang/pendukung)

Ukuran yang menunjukkan seberapa besar tingkat dominasi suatu item/item-set dari keseluruhan transaksi (misal dari keseluruhan transaksi yang ada, seberapa besar tingkat dominasi yang menunjukkan bahwa item A dan B dibeli bersamaan), dengan rumus :

$$\text{Support}(A,B) = P(A \rightarrow B) = \frac{JT(A \rightarrow B)}{T} \times 100\%$$

Di mana, $JT(A \rightarrow B)$ adalah jumlah transaksi yang mengandung A dan B, T adalah total transaksi.

b. Confidence (nilai kepastian/keyakinan)

Ukuran yang menunjukkan hubungan antar 2 item secara conditional (misal, seberapa sering item B dibeli jika orang membeli item A), dengan rumus

$$\text{Confidence}(A,B) = P(A \rightarrow B) = \frac{JT(A \rightarrow B)}{JT(A)} \times 100\%$$

$JT(A \rightarrow B)$ adalah jumlah transaksi yang mengandung A dan B, JT adalah jumlah transaksi yang mengandung A.

Kedua ukuran ini nantinya berguna dalam menentukan *interesting association rules*, yaitu untuk dibandingkan dengan batasan (*threshold*) yang ditentukan oleh *user*. Batasan tersebut umumnya terdiri dari *minimum support* dan *minimum confidence*.

Algoritma Apriori adalah salah satu teknik *association rule* yang sangat populer untuk melakukan pencarian pola. Algoritma

ini diajukan oleh R. Agrawal dan R. Srikant tahun 1994. Algoritma ini ditujukan untuk mencari kombinasi *item-set* yang mempunyai suatu nilai keseringan (*frequent itemset*) tertentu sesuai kriteria atau filter yang diinginkan dengan cara memperhatikan *minimum support (minsup)*.

Adapun dua proses utama yang dilakukan algoritma Apriori (Han Jiawei, and M. Kamber. 2006), yaitu *join* (penggabungan) dan *prune* (pemangkasan). Pada proses penggabungan setiap item dikombinasikan dengan item yang lainnya sampai tidak terbentuk kombinasi lagi. Sedangkan pada proses pemangkasan hasil dari item yang telah dikombinasikan tadi dipangkas dengan menggunakan *minimum support* yang telah ditentukan oleh *user*.

```

Input :
D, a database of a transactions;
Min_support, the minimum support
count
threshold
Output : L, frequent itemsets in D
Method :
L1 = find_frequent_1_itemsets(D);
for (k = 2; Lk-1 ≠ ∅ ; k | |) {
    Ck = Apriori_gen(Lk-1);

    for each transaction t ∈ D

    { //scan D for counts
        Ct = subset(Ck, t); //get the
        subsets of t that are candidates

        for each candidate c ∈ Ct

            c.count | |;
        }

        Lk = { c ⊂ Ck | c.counts ≥
min_sup }
    }

    return L = Uk Lk;

```

Gambar 1. Algoritma Apriori

Dua proses utama tersebut merupakan langkah yang akan dilakukan untuk mendapat *frequent item-set*, yang dapat dilihat pada Gambar 1 [Nikky, 2013].

Frequent Pattern Growth (FP-Growth) adalah salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah kumpulan data. Karakteristik algoritma *FP-Growth* adalah struktur data yang digunakan adalah *tree* yang disebut dengan *FP-Tree*. Dengan menggunakan *FP-Tree*, algoritma *FP-Growth* dapat langsung mengekstrak *frequent itemset* dari *FP-Tree*.

Penggalian *item-set* yang *frequent* dengan menggunakan algoritma *FP-Growth* akan dilakukan dengan cara membangkitkan struktur data *tree* atau disebut dengan *FP-Tree*. Metode *FP-Growth* dapat dibagi menjadi 3 tahapan utama yaitu tahap pembangkitan conditional pattern base, pembangkitan conditional *FP-Tree*, dan pencarian frequent item-set (Gosta Grahe, 2005).

Ketiga tahap tersebut merupakan langkah yang akan dilakukan untuk mendapatkan *frequent item-set* yang dapat dilihat pada gambar 2 [Nikky, 2013].

HASIL DAN PEMBAHASAN

Data utama yang digunakan pada penelitian ini adalah data transaksi pada koperasi *Uber-Mart* Bekasi, yang terdiri dari 5.000 transaksi pada tahun 2013. Karena pada penelitian ini dianalisis perbandingan antara *Apriori* dan *FP-Growth* maka data yang digunakan akan dibuat menjadi dua tetapi dengan jumlah data transaksi yang berbeda seperti terlihat pada Tabel 1.

```

Input : FP-Tree Tree
Output : Rt sekumpulan
lengkap pola
frequent
Method : FP-growth (Tree,
null)
Procedure : FP-growth (Tree,
α)
01: if Tree mengandung
single path P;
02: then untuk tiap
kombinasi (dinotasikan β)
dari node-node dalam path do
03: bangkitkan pola β α
dengan support dari node-
node dalam β;
04: else untuk tiap a1 dalam
header dari Tree do
{
05: bangkitkan pola
06: bangun β = a1 α dengan
support = a1.
support {
07: if Tree β = ∅
08: then panggil FP-growth
(Tree, β)
}
}

```

Gambar 2. Algoritma FP-Growth

Tabel 1. Spesifikasi Data Percobaan

No.	Nama Tabel	Jumah Transaksi	Jumlah Record	Jumlah Item
1.	data_1	2500	8478	1794
2.	data_2	5000	16731	1794

Sebelum proses *data mining* dimulai, dilakukan proses *preprocessing* sehingga didapat dataset seperti diperlihatkan dalam Tabel 2.

Tabel 2. Dataset Hasil Preprocessing

No Transaksi	Nama Item
1	Aqua Btl 350ml
1	Baygon LV 750ml
1	Bebelac Madu 400gr
2	Bendera SKM Cklt
2	Beng-beng
2	Blue Band 450gr
....
1501	DAIA Lemon 380gr
1501	Lifeboy merah soap 80gr
1502	Vape Reff 720ml
....

Setelah melalui proses *preprocessing*, dataset tersebut bias di mining dengan menggunakan algoritma *Apriori* dan *FP-Growth*.

Pada algoritma *Apriori*, prosesnya selanjutnya dataset di-*scan*, kemudian setiap item dikombinasikan dengan item yang lainnya. Kemudian kombinasi item tersebut dipangkas untuk menghasilkan *item-sets* yang nantinya akan dikombinasikan lagi sampai tidak dapat terbentuk lagi kombinasi item berdasarkan nilai *minimum_support* yang telah ditentukan oleh *user*.

Penelitian ini menggunakan *minimum support* = 40%, yang hasilnya dapat dilihat pada gambar 3 untuk data_1, dan gambar 4 untuk data_2.

No. Transaksi	Nama Item	Support
1	Aqua Btl Sdg 500ml	500/2500*100%= 20%
1	Beng-Beng 5gr	375/2500*100%= 15%
1	Bendera SKM Cok	250/2500*100%= 10%
2	San Roti Kupas	1250/2500*100%= 50%
2	Bebelac Madu	375/2500*100%= 15%
2	Blue Band	500/2500*100%= 20%
2	Morina Selai kacang	375/2500*100%= 15%
3	Pepsodent Pasta 125gr	375/2500*100%= 15%
....

Nama Item	Support
San Roti Kupas	1250/2500*100%= 50%
Sovia Mnyk Goreng	1175/2500*100%= 47%
Aqua Btl Sdg	1175/2500*100%= 47%
Lifeboy Sbn	1125/2500*100%= 45%
Mie Sedap Kan SP	1125/2500*100%= 45%
Paseo Mni	1050/2500*100%= 42%
Sanwangi The Asli	1000/2500*100%= 40%
Telur Ayam	1000/2500*100%= 40%
....

Itemset	Support
Sari Roti, Sanwangi The Asli	45%
Sari Roti, Keju Kraft	45%
Sari Roti, Aqua Btl Sdg	43%
Sovia Mnyk Goreng, Telur Ayam	43%
Mie Sedap Kan SP, Telur Ayam	42%
Paseo Mini, Pampers Reg M	41%
Bebelac Madu, My Baby Pwdr	40%
....

Gambar 3. Algoritma Apriori dengan data 2500 (a). Kandidat 1-Itemset (C₁), (b). Large 1-itemset (L₁), (c). Large 2-itemset

No Transaksi	Nama Item	Support
1	Fugu seaweed 22gr	40
1	Sprite 1500 ml	40
1	Frutang hot 330ml	20
2	Boom jeruk nipis	41
....

Nama Item	Support
Fugu seaweed 22gr, Sprite 1500 ml	40
Fugu seaweed 22gr, Frutang hot 330 ml	40
Fugu seaweed 22gr, Boom jeruk nipis	40
Frutang hot fill 330 ml, Boom jeruk nipis	41
....

Nama Item	Support
Mizone lemon, Sgm 1 600 Gr	43
Happy Juice Orange, Sgm 1 600 Gr	43
....

Gambar 4. Algoritma Apriori dengan data 5000 (a). Kandidat 1-Itemset (C₁), (b). Large 1-itemset (L₁), (c). Large 2-itemset

Pada algoritma *FP-Growth*, proses yang pertama untuk mencari *frequent itemsets* sama dengan algoritma *Apriori*. Kumpulan *frequent item-sets* dicari dengan melakukan *scanning* database kemudian hasilnya diakumulasikan dengan tiap item lainnya dan dikelompokkan untuk mendapatkan *frequent itemsets* yang memenuhi nilai *minimum support* (Arini, 2007).

Setelah didapatkan *frequent item-sets*, *frequent item-sets* yang memiliki nilai kurang dari *minimum support* akan dipangkas/dibuang, sehingga didapatkan *frequent item-sets* yang memenuhi *minimum support* dan disusun dengan urutan menurun. Hasil dari *list* tersebut dinotasikan dengan L.

Setelah itu, dilakukan *scanning* database untuk kedua kalinya. Pada proses ini *FP-Tree* mulai dibentuk. Pertama, membuat akar dari pohon dan diberi nama *null*. Lalu, L dipindai untuk membuat batang dari setiap transaksi yang ada di L. Pembacaan transaksi pertama yang ada di

L akan membuat simpul, sehingga akan terbentuk lintasan transaksi. *Support count* dari setiap transaksi bernilai satu. Kemudian, dilanjutkan proses pembacaan transaksi kedua.

Dari pembangunan *FP-Tree*, dapat diterapkan algoritma *FP-Growth*. Untuk menemukan *frequent item-sets*, maka perlu ditentukan *tree* dengan lintasan yang berakhir dengan *support count* terkecil, nantinya akan didapatkan *frequent item-set*.

Berdasarkan aturan asosiasi yang terbentuk maka item yang memenuhi *minimum support=40%* dan *minimum confidence=75%*, dapat dilihat pada Tabel 3 dan Tabel 4.

Tabel 3.
Aturan Asosiasi (Association Rule)
untuk data 2500

IF Antecedent ELSE Consequence (X -> Y)	minimum support	minimum confidence
Sari Roti → Sariwangi Teh Asli	45%	50%
Sariwangi Teh Asli → Sari Roti	45%	75%
Sari Roti → Keju Kraft	45%	50%
Keju Kraft → Sari Roti	45%	100%
Sovia Mnyk Goreng → Mie Sedap Kari SP	42%	75%
Mie Sedap Kari SP → Sovia Mnyk Goreng	42%	50%
Sovia Mnyk Goreng → Telur Ayam	43%	50%
Telur Ayam → Sovia Mnyk Goreng	40%	75%
.....

Tabel 4.
Aturan Asosiasi (Association Rule)
untuk data 5000

IF Antecedent ELSE Consequence (X -> Y)	minimum support	minimum confidence

Hasil dari penelitian ini adalah perangkat lunak dengan parameter yang sama pada tabel yang berbeda. Data yang diperoleh dibagi menjadi tabel data_1 untuk data 2500 transaksi, dan tabel data_2 untuk data 5000 transaksi. Sedangkan parameter yang digunakan adalah *minimum_support* dan *minimum_confidence*. Hasil uji coba untuk masing-masing tabel dapat dilihat pada tabel 5 dan tabel 6. Spesifikasi tools untuk penghitungan algoritma *Apriori* dan *FP-Growth* adalah Rapidminer 5. Sedangkan spesifikasi perangkat keras untuk menjalankan tools tersebut adalah processor Pentium 64 bit dengan RAM 4 GB.

Tabel 5.
Hasil Uji Coba Menggunakan data_1

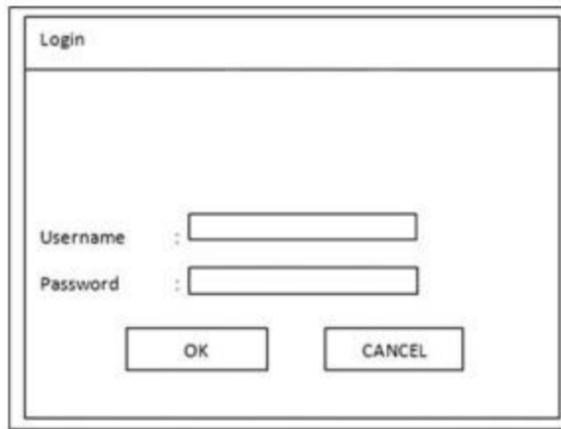
Algoritma	Minimum_Support	Time_Process
Apriori	40	3 jam 12 menit
FP-Growth	40	1 jam 20 menit

Tabel 6.
Hasil Uji Coba Menggunakan data_2

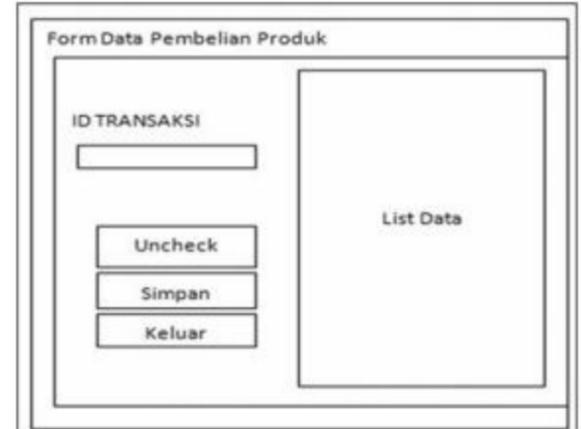
Algoritma	Minimum_Support	Time_Process
Apriori	40	3 jam 12 menit
FP-Growth	40	1 jam 20 menit

Rancangan Antarmuka Aplikasi

Dalam penelitian ini, rancangan antarmuka yang akan dibuat berbasis web. Gambar 5 adalah rancangan antarmuka login, dan gambar 6 merupakan rancangan antarmuka pencarian *item-set*.



Gambar 5. Form Login

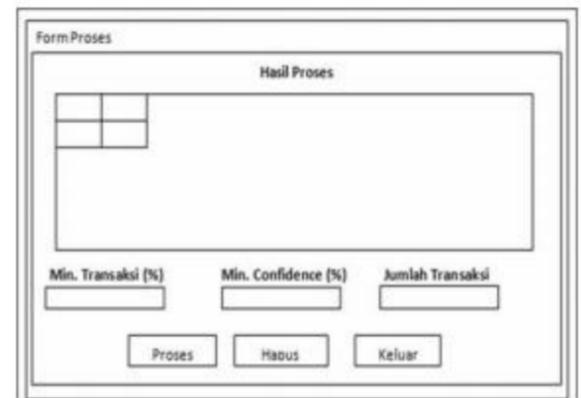


Gambar 8. Form Data Pembelian Produk

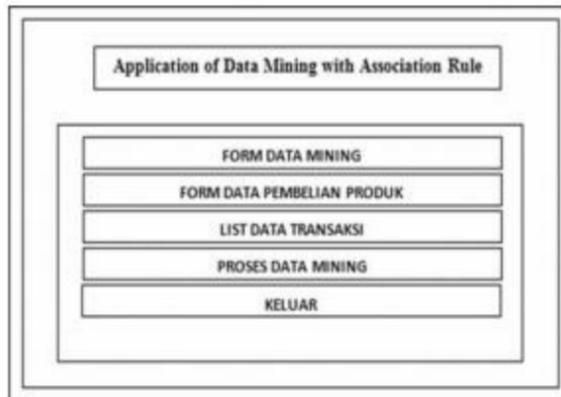


Gambar 9. Form List Data Transaksi

Berikut adalah perancangan tampilan proses yang terdapat pada menu proses data mining, dapat dilihat pada gambar 10.

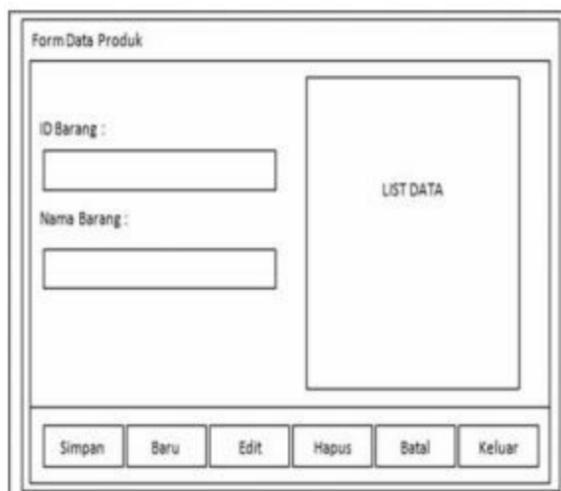


Gambar 10. Form Proses Data Mining



Gambar 6. Form Menu Utama

Rancangan form pada gambar 7, berguna untuk memasukkan data barang ke dalam sistem, seperti pada gambar di bawah ini. Pada gambar 8, berguna untuk memasukkan data penjualan barang ke dalam sistem. Pada gambar 9, dirancang untuk menampilkan daftar data transaksi yang telah diinput oleh user.



Gambar 7. Form Data Produk

KESIMPULAN

Walaupun algoritma *Apriori* mudah dipahami dan diimplementasikan dibanding dengan algoritma yang lainnya yang memang diterapkan untuk proses *association rule*, akan tetapi algoritma *Apriori* juga memiliki kekurangan yaitu, untuk melakukan pencarian *frequent itemset* 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.

Algoritma *FP-Growth* merupakan pengembangan dari algoritma *Apriori* sehingga kekurangan dari algoritma *Apriori* diperbaiki oleh algoritma *FP-Growth*. Pada algoritma *Apriori* diperlukan *generate candidate* untuk mendapatkan *frequent itemsets*. Akan tetapi, di algoritma *FP-Growth generate candidate* tidak dilakukan karena *FP-Growth* menggunakan konsep

pembangunan *tree* dalam pencarian *frequent itemsets*. Hal tersebutlah yang menyebabkan algoritma *FP-Growth* lebih cepat dari algoritma *Apriori*.

DAFTAR PUSTAKA

- Arini, *et al*, 2007. Analisis Kinerja Algoritma *FP-Growth* Pada Penggalian Pola Asosiasi. <http://www.si.its.ac.id/Penelitian/JURNAL/Arin.pdf>. Diakses pada tanggal 17 Februari 2013.
- Erwin. 2009. Analisis Market basket Dengan Algoritma *Apriori* dan *FP-Growth*. <http://www.eprints.unsri.ac.id/83/1/6-Erwin.pdf>. Diakses pada tanggal 17 Februari 2013.
- Fanani, Zainul. 2010. Rancang Bangun Sistem Informasi Inventori Menggunakan Metode *Association Rules* Di CV Damar Langit. <http://www.lib.uin-malang.ac.id/files/thesis/fullchapter/04520044.pdf>. Diakses 17 Februari 2014.
- Gosta, Grahe. 2005. Fast Algorithms for Frequent Itemset Mining Using *FP-Tress*. http://ftp.gunadarma.ac.id/research/IEEE/Knowledge_Data_Engineering/Oct_05/k1347.pdf. Diakses pada tanggal 17 Februari 2013.
- Han Jiawei, and M. Kamber. 2006. *Data mining: Concept and Techniques*. Morgan Kaufmann, USA
- Kokoh, Philips. 2006. Association Rule Mining. <http://philips.wordpress.com/2006/05/10/association-rule-mining/>. Diakses pada tanggal 17 Februari 2013.
- Palace, Bill. 1996. Data Mining: What is Data Mining?. <http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm>. Diakses pada tanggal 17 Februari 2013.
- Nikky. 2013. Mengenal Perbedaan Algoritma *Apriori* dan *FP-Growth* Dalam Menentukan Frequent Itemset. <http://www.saranainformasi.com/2013/10/08/mengenal-perbedaan-algoritma-apriori-dan-fp-growth-dalam-menentukan-frequent-itemset/>. Diakses pada tanggal 17 Februari 2014.
- Ulmer, David. 2002. Mining an Online Auctions DataWarehouse. The Mid Atlantic Student Workshop on Programming Languages and Systems. 19 April 2002. Pace University. <<http://csis.pace.edu/csis/masplas/p8.pdf>>.
- Willyanto, Leo. 2005. Pembuatan Perangkat Lunak Data Mining Untuk Penggalian Kaidah Asosiasi Menggunakan Metode *Apriori*. <http://www.petra.ac.id/~puslit/journals/request.php>. Diakses 17 Februari 2013.

