

# PENERAPAN METODE *DATA MINING MARKET BASKET ANALYSIS* TERHADAP DATA PENJUALAN PRODUK BUKU DENGAN MENGGUNAKAN ALGORITMA APRIORI DAN *FREQUENT PATTERN GROWTH (FP-GROWTH)* : STUDI KASUS PERCETAKAN PT. GRAMEDIA

<sup>1</sup>Goldie Gunadi, <sup>2</sup>Dana Indra Sensuse

<sup>1</sup>Magister Ilmu Komputer Universitas Budi Luhur  
send2goldie@gmail.com

<sup>2</sup>Ilmu Komputer Universitas Indonesia  
dana@cs.ui.ac.id

## ABSTRAK

*Data mining merupakan proses analisa data untuk menemukan suatu pola dari kumpulan data tersebut. Data mining mampu menganalisa data yang besar menjadi informasi berupa pola yang mempunyai arti bagi pendukung keputusan. Salah satu teknik data mining yang dapat digunakan adalah association data mining atau yang biasa disebut dengan istilah market basket analysis. Market basket didefinisikan sebagai suatu itemset yang dibeli secara bersamaan oleh pelanggan dalam suatu transaksi. Market basket analysis adalah suatu alat yang ampuh untuk pelaksanaan strategi cross-selling. Metode ini dimulai dengan mencari sejumlah frequent itemset dan dilanjutkan dengan pembentukan aturan-aturan asosiasi (association rules). Algoritma Apriori dan frequent pattern growth (FP-growth) adalah dua algoritma yang sangat populer untuk menemukan sejumlah frequent itemset dari data-data transaksi yang tersimpan dalam basis data. Dalam penelitian ini algoritma Apriori dan frequent pattern growth (FP-growth) digunakan untuk membantu menemukan sejumlah aturan asosiasi dari basis data transaksi penjualan produk buku di Percetakan PT. Gramedia, sehingga untuk selanjutnya dapat digunakan sebagai pertimbangan dalam membuat strategi pemasaran dan penjualan yang efektif.*

**Kata kunci :** *data mining, association rules, market basket analysis, Apriori, frequent pattern growth, FP-growth*

## 1. Pendahuluan

Perkembangan teknologi informasi telah memberikan kontribusi pada cepatnya pertumbuhan jumlah data yang dikumpulkan dan disimpan dalam basis data berukuran besar (gunung data). Dibutuhkan sebuah metode atau teknik yang dapat merubah gunung data tersebut menjadi sebuah informasi berharga atau pengetahuan (*knowledge*) yang bermanfaat untuk mendukung pengambilan keputusan

bisnis. Suatu teknologi yang dapat digunakan untuk mewujudkannya adalah *data mining*. *Data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Salah satu metode yang seringkali digunakan dalam teknologi *data mining* adalah metode asosiasi atau *association rule mining*. Di dalam bidang usaha *retail* metode *association rule mining* ini lebih dikenal dengan istilah analisa

keranjang belanja (*market basket analysis*). *Market basket analysis* adalah suatu metode analisa atas perilaku konsumen secara spesifik dari suatu golongan / kelompok tertentu. Sumber data dari *market basket analysis* antara lain dapat bersumber dari transaksi kartu kredit, kartu lotere, kupon diskon, panggilan keluhan pelanggan. *Market basket analysis* umumnya dimanfaatkan sebagai titik awal pencarian pengetahuan dari suatu transaksi data ketika kita tidak mengetahui pola spesifik apa yang kita cari. Kebutuhan *market basket analysis* berawal dari keakuratan dan manfaat yang dihasilkannya dalam wujud aturan asosiasi (*association rules*). Yang dimaksud dengan *association rules* adalah pola-pola keterkaitan data dalam basis data.

Percetakan PT. Gramedia adalah sebuah perusahaan yang bergerak di bidang industri percetakan dan telah beroperasi sejak tahun 1972. Hingga kini Percetakan PT. Gramedia telah melayani sekian banyak transaksi pesanan cetakan. Setiap data transaksi tersebut disimpan di dalam suatu sistem basis data melalui aplikasi sistem informasi manajemen.

Dalam penelitian ini penulis akan mencoba untuk melakukan suatu eksperimen terhadap data transaksi penjualan produk buku di Percetakan PT. Gramedia dengan menggunakan sebuah metode *market basket analysis* untuk melihat hubungan asosiasi (korelasi) antara sejumlah atribut penjualan. Algoritma yang akan digunakan adalah algoritma Apriori dan algoritma *frequent pattern growth* (FP-growth). Kedua algoritma ini digunakan untuk membentuk *frequent itemsets* yang nantinya akan dijadikan sebagai acuan untuk merumuskan aturan-aturan asosiasi yang dihasilkan oleh model *market basket analysis* dengan menggunakan alat bantu perangkat lunak aplikasi Weka versi 3.6. Aturan-aturan asosiasi yang dihasilkan oleh masing-masing algoritma Apriori dan FP-growth akan dievaluasi dan dianalisa untuk menemukan algoritma yang paling baik dan sesuai (*best-fit algorithm*) untuk permasalahan yang dikaji dalam penelitian ini.

Komponen-komponen penjualan utama yang akan digunakan sebagai acuan didalam penentuan hubungan asosiasi dengan menggunakan *market basket analysis* adalah :

- a. Jenis cover buku
- b. Teknik penjilidan
- c. Penggunaan plastik buku
- d. Penggunaan jenis kertas khusus untuk cover buku
- e. Penggunaan jenis kertas khusus untuk isi buku
- f. Pekerjaan-pekerjaan dalam pembuatan cover seperti : UV, laminating, embossed, dan hot print.

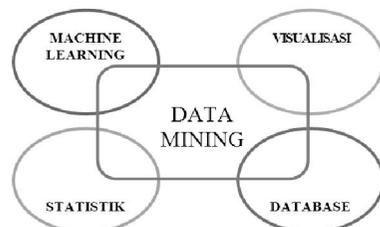
Dengan dilakukannya penelitian ini, diharapkan dapat memberikan hasil berupa informasi yang bermanfaat bagi pihak-pihak terkait dalam melakukan proses pengambilan keputusan manajerial, terutama yang berkaitan dengan perumusan atau pembuatan strategi pemasaran dan penjualan, khususnya untuk produk-produk buku di Percetakan PT. Gramedia.

## 2. Landasan Pemikiran

### 2.1. Tinjauan Pustaka

#### 2.1.1. Data Mining

Menurut Turban dalam bukunya yang berjudul "*Decision Support Systems and Intelligent Systems*", *data mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam basis data. *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 basis data besar [1].



Gambar 1. *Data mining* merupakan irisan dari berbagai disiplin [2].

### 2.1.2. Pengelompokan *Data Mining*

Menurut Larose dalam bukunya yang berjudul "*Discovering Knowledge in Data: An Introduction to Data Mining*", *data mining* dibagi menjadi beberapa kelompok berdasarkan tugas/pekerjaan yang dapat dilakukan, yaitu [1]:

- a. Deskripsi  
Terkadang peneliti dan analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Deskripsi dari pola kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.
- b. Estimasi  
Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik dari pada ke arah kategori. Model dibangun menggunakan baris data (*record*) lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.
- c. Prediksi  
Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.
- d. Klasifikasi  
Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.
- e. Pengklasteran (*Clustering*)  
Pengklasteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas obyek-obyek yang memiliki kemiripan. Klaster adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki

ketidakmiripan *record* dalam klaster yang lain.

Berbeda dengan klasifikasi, pada pengklasteran tidak ada variabel target. Pengklasteran tidak melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target, akan tetapi, algoritma pengklasteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan *record* dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan *record* dalam kelompok lain akan bernilai minimal.

### f. Asosiasi

Tugas asosiasi dalam *data mining* adalah untuk menemukan atribut yang muncul dalam satu waktu. Salah satu implementasi dari asosiasi adalah *market basket analysis* atau analisis keranjang belanja, sebagaimana yang akan dibahas dalam tesis ini.

### 2.1.3. *Knowledge Discovery in Databases*

KDD terdiri dari tiga proses utama yaitu:

- a. *Preprocessing*  
*Preprocessing* dilakukan terhadap data sebelum algoritma *data mining* diaplikasikan. Proses ini meliputi *data cleaning*, integrasi, seleksi dan transformasi.
- b. *Data mining*  
Proses utama dalam KDD adalah proses *data mining*, dalam proses ini algoritma-algoritma *data mining* diaplikasikan untuk mendapatkan pengetahuan dari sumber data.
- c. *Post processing*  
Hasil yang diperoleh dari proses *data mining* selanjutnya akan dievaluasi pada *post processing*.

### 2.1.4. *Market Basket Analysis*

*Market basket analysis* adalah suatu metodologi untuk melakukan analisis *buying habit* konsumen dengan menemukan asosiasi antar beberapa *item* yang berbeda, yang diletakkan konsumen dalam *shopping basket* (keranjang belanja) yang dibeli pada

suatu transaksi tertentu. Tujuan dari *market basket analysis* adalah untuk mengetahui produk-produk mana yang mungkin akan dibeli secara bersamaan [3].

**2.1.5. Analisis Asosiasi**

Analisis asosiasi atau *association rule mining* adalah teknik *data mining* untuk menemukan aturan asosiasi antara suatu kombinasi *item* [1]. *Interestingness measure* yang dapat digunakan dalam *data mining* adalah :

- a. *Support*, adalah suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu *item* atau *itemset* dari keseluruhan transaksi.
- b. *Confidence*, adalah suatu ukuran yang menunjukkan hubungan antar dua *item* secara *conditional* (berdasarkan suatu kondisi tertentu).

Metodologi dasar analisis asosiasi terbagi menjadi 2 tahap, yakni melakukan analisa pola frekuensi tinggi (*frequent pattern*) dan berikutnya adalah proses pembentukan aturan asosiasi.

**2.1.6. Algoritma Apriori**

Algoritma Apriori adalah suatu algoritma dasar yang diusulkan oleh Agrawal & Srikant pada tahun 1994 untuk penentuan *frequent itemsets* untuk aturan asosiasi *boolean* [3].

Bentuk algoritma dari metode Apriori dapat dituliskan sebagai berikut [3]:

```

L1 = {frequent itemset with one element}
for (k=2; Lk-1≠∅; k++)
{
Ck = apriori-gen(Lk-1); //pembuatan
kandidat
//baru
for all transactions t
{
C't = subset(Ck, t); //kandidat yang
//tampil pada t
for all candidates c ∈ C't do
c.count ++;
}
Lk = { c ∈ C't | c.count ≥ minsup }
}
return ∪kLk;
    
```

Dimana :

- L* : himpunan *frequent itemset*
- minsup* : *minimum support*
- C* : himpunan *kandidat itemset*
- c* : *kandidat itemset*
- t* : *transaksi*

**2.1.7. Analisis Pola Frekuensi Tinggi dengan Algoritma Apriori**

Mencari kombinasi *item* yang memenuhi syarat minimum dari nilai *support* dalam basis data.

Nilai *support* sebuah *item* diperoleh dengan menggunakan rumus berikut:

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

Nilai *support* dari 2 *item* diperoleh dengan menggunakan rumus :

$$Support(A, B) = P(A \cap B)$$

$$Support(A, B) = \frac{\sum Transaksi\ mengandung\ A\ dan\ B}{\sum Transaksi}$$

*Frequent itemset* menunjukkan *itemset* yang memiliki frekuensi kemunculan lebih dari nilai minimum yang ditentukan ( $\Phi$ ). Misalkan  $\Phi = 2$ , maka semua *itemsets* yang frekuensi kemunculannya lebih dari atau sama dengan 2 kali disebut *frequent*. Himpunan dari *frequent k-itemset* dilambangkan dengan  $F_k$ .

**2.1.8. Algoritma FP-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. FP-growth menggunakan pendekatan yang berbeda dari paradigma yang digunakan pada algoritma Apriori.

**2.1.9. Analisis Pola Frekuensi Tinggi dengan Algoritma FP-growth**

Pada penentuan *frequent itemset* terdapat 2 tahap proses yang dilakukan yaitu: pembuatan FP-tree dan penerapan algoritma FP-growth untuk menemukan *frequent itemset*. Struktur data yang digunakan untuk mencari *frequent itemset*

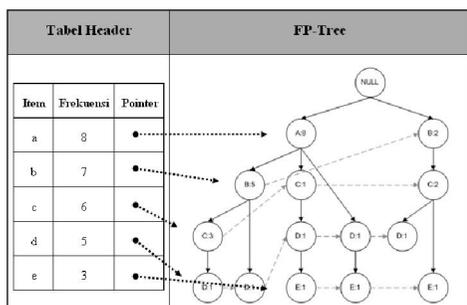
dengan algoritma FP-growth adalah perluasan dari penggunaan sebuah pohon *prefix*, yang biasa disebut adalah FP-tree. Dengan menggunakan FP-tree, algoritma FP-growth dapat langsung mengekstrak *frequent itemset* dari FP-tree yang telah terbentuk dengan menggunakan prinsip *divide and conquer*.

**2.1.9.1. Pembuatan FP-Tree**

FP-tree merupakan struktur penyimpanan data yang dimampatkan. FP-tree dibangun dengan memetakan setiap data transaksi ke dalam setiap lintasan tertentu dalam FP-tree. Karena dalam setiap transaksi yang dipetakan, mungkin ada transaksi yang memiliki *item* yang sama, maka lintasannya memungkinkan untuk saling menimpa. Semakin banyak data transaksi yang memiliki *item* yang sama, maka proses pemampatan dengan struktur data FP-tree semakin efektif.

Adapun FP-tree adalah sebuah pohon dengan definisi sebagai berikut:

- a. FP-tree dibentuk oleh sebuah akar yang diberi label *null*, sekumpulan *sub-tree* yang beranggotakan *item-item* tertentu, dan sebuah tabel *frequent header*.
- b. Setiap simpul dalam FP-tree mengandung tiga informasi penting, yaitu label *item*, menginformasikan jenis *item* yang direpresentasikan simpul tersebut, *support count*, merepresentasikan jumlah lintasan transaksi yang melalui simpul tersebut, dan *pointer* penghubung yang menghubungkan simpul-simpul dengan label *item* sama antar-lintasan, ditandai dengan garis panah putus-putus.



Gambar 2. Contoh FP-tree dan Tabel Header

**2.1.9.2. Penerapan Algoritma FP-Growth**

Setelah tahap pembangunan FP-tree dari sekumpulan data transaksi, akan diterapkan algoritma FP-growth untuk mencari *frequent itemset* yang signifikan. Algoritma FP-growth dibagi menjadi tiga langkah utama, yaitu :

- a. Tahap Pembangkitan *Conditional Pattern Base*

*Conditional Pattern Base* merupakan *sub-database* yang berisi *prefix path* (lintasan prefix) dan *suffix pattern* (pola akhiran). Pembangkitan *conditional pattern base* didapatkan melalui FP-tree yang telah dibangun sebelumnya.

- b. Tahap Pembangkitan *Conditional FP-tree*

Pada tahap ini, *support count* dari setiap *item* pada setiap *conditional pattern base* dijumlahkan, lalu setiap *item* yang memiliki jumlah *support count* lebih besar sama dengan *minimum support count*  $\xi$  akan dibangkitkan dengan *conditional FP-tree*.

- c. Tahap Pencarian *frequent itemset*

Apabila *Conditional FP-tree* merupakan lintasan tunggal (*single path*), maka didapatkan *frequent itemset* dengan melakukan kombinasi *item* untuk setiap *conditional FP-tree*. Jika bukan lintasan tunggal, maka dilakukan pembangkitan FP-growth secara rekursif.

**2.1.10. Pembentukan Aturan Asosiasi**

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

Nilai *confidence* dari aturan  $A \Rightarrow B$  diperoleh dengan rumus berikut.

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

Untuk menentukan aturan asosiasi yang akan dipilih maka harus diurutkan berdasarkan  $Support \times Confidence$ . Aturan diambil sebanyak  $n$  aturan yang memiliki hasil terbesar [1].

### 2.2. Tinjauan Studi

Dalam [4] diukur nilai *confidence* dari setiap aturan asosiasi yang dihasilkan. Aturan asosiasi yang diambil adalah aturan-aturan asosiasi yang memiliki nilai *confidence* lebih besar atau sama dengan nilai *minimum confidence* yang telah ditentukan.

Dalam tulisannya, [5] menggunakan hasil dari nilai *Support* dikalikan dengan *Confidence* terhadap masing-masing aturan asosiasi yang dihasilkan untuk mengukur kekuatan dari aturan asosiasi yang bersangkutan.

Dalam penelitiannya yang berjudul "A Scalable Algorithm for the Market Basket Analysis", Luis Cavique melakukan perbandingan performa terhadap akurasi dari dua algoritma asosiasi yaitu algoritma similis dan FP-growth, dengan menghitung tingkat akurasinya.

Fungsi yang digunakan untuk menghitung persentase akurasi dari algoritma similis terhadap FP-growth adalah :

$$\frac{\sum \text{Similis Support}}{\sum \text{FP - growth Support}} \times 100$$

### 2.3. Tinjauan Obyek Penelitian

#### 2.3.1. Percetakan PT. Gramedia

Percetakan PT. Gramedia berdiri sejak tahun 1972, bergerak di bidang layanan jasa cetak koran, tabloid, buku, majalah dan material promosi. Disamping melayani pelanggan dalam kelompok Kompas Gramedia, juga telah dipercaya oleh pelanggan dari luar kelompok Kompas Gramedia untuk mencetak produk-produknya.

#### 2.3.2. Visi dan Misi Perusahaan

Visi yang ingin dicapai oleh Percetakan PT. Gramedia adalah: "Menjadi percetakan kelas dunia terbaik di Asia Tenggara."

Misi dari Percetakan PT. Gramedia adalah :

- a. Mendukung pencapaian sasaran bisnis pelanggan dengan memahami bisnis, menjamin keamanan produk, memberi kualitas yang dipilih, memberikan

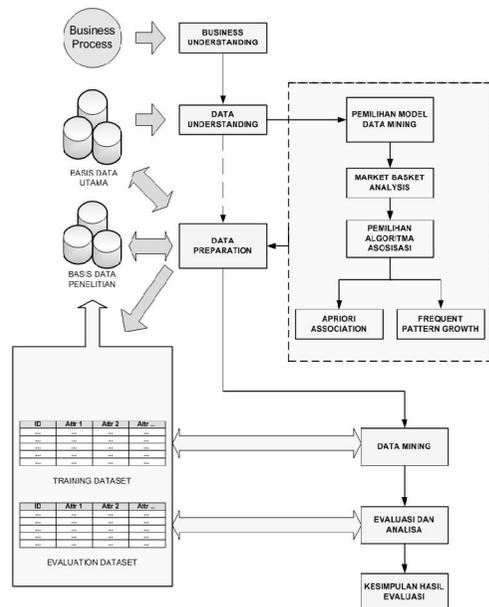
inovasi produk dan layanan yang memuaskan pelanggan.

- b. Menyediakan layanan desain, percetakan, pengemasan dan layanan distribusi yang paling disukai pelanggan untuk mencerahkan dan mencerdaskan masyarakat.

### 2.4. Kerangka Pemikiran

Kerangka pemikiran penelitian *market basket analysis*, meliputi :

- a. *Business Understanding*
- b. *Data Understanding*
- c. Pemilihan Model *Data Mining*
- d. Pemilihan Algoritma Asosiasi
- e. *Data Preparation*
- f. *Data Mining*
- g. Evaluasi dan Analisa



Gambar 3. Kerangka pemikiran penelitian *market basket analysis*

#### 2.4.1. Evaluasi

Ada 2 faktor yang akan dievaluasi dalam penelitian ini, yaitu ukuran generalitas (*generality*) dan ukuran reliabilitas (*reliability*) dari *association rules* yang dihasilkan.

Ukuran generalitas digunakan untuk mengetahui seberapa tingkat kemunculan setiap *item* yang dirumuskan dalam aturan

asosiasi terhadap keseluruhan transaksi. Ukuran-ukuran yang digunakan adalah :

a. *Support*

*Support* adalah ukuran dari seberapa sering koleksi *item* dalam asosiasi terjadi bersama sebagai persentase dari seluruh transaksi.

Rumus :

$$s(A \Rightarrow B) = P(A \cap B)$$

dimana :

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

b. *Coverage*

*Coverage* adalah ukuran seberapa sering koleksi setiap *item* yang dirumuskan pada sisi sebelah kiri (faktor penyebab) terjadi sebagai persentase dari seluruh transaksi.

Rumus :

$$P(A) = \frac{\sum \text{Transaksi mengandung A}}{\sum \text{Transaksi}}$$

Ukuran reliabilitas digunakan untuk mengetahui tingkat kehandalan dari aturan asosiasi yang dihasilkan dalam *association rules mining*. Ukuran-ukuran yang digunakan adalah :

a. *Confidence*

*Confidence* adalah ukuran ketidakpastian atau kepercayaan kelayakan yang terkait dengan setiap pola yang ditemukan.

Rumus :

$$\alpha(A \Rightarrow B) = P(B|A)$$

dimana :

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

b. *Added Value*

*Added Value* digunakan untuk mengukur berapa besar selisih antara nilai *confidence* dengan peluang terjadinya kombinasi setiap *item* pada sisi akibat pada aturan asosiasi terhadap seluruh transaksi.

$$\text{Added Value } (A \Rightarrow B) = P(B|A) - P(B)$$

c. *Correlation*

*Correlation* digunakan untuk menentukan apakah antara *itemset* yang satu

dengan yang lainnya memiliki suatu ketergantungan atau tidak.

*Correlation* ( $A \Rightarrow B$ ) bernilai 0 jika A dan B adalah bersifat independen (tidak memiliki ketergantungan), dan 1 (nilai positif, >0) atau -1 (nilai negatif, <0) jika mereka tergantung. 1 menunjukkan daya tarik, dan -1 menunjukkan tolakan.

Rumus :

$$\text{Correlation } (A \Rightarrow B) = \frac{P(A \cap B) - [P(A)P(B)]}{\sqrt{P(A)P(B)}}$$

dimana :

$$P(A) = \sum \text{Transaksi mengandung A}$$

$$P(B) = \sum \text{Transaksi mengandung B}$$

2.4.2. **Analisa**

Analisa dilakukan terhadap sejumlah aturan asosiasi dengan nilai *confidence* terbesar yang dihasilkan oleh masing-masing algoritma Apriori dan FP-growth. Analisa dilakukan dengan menghitung tingkat kekuatan dan akurasi dari masing-masing aturan yang diimplementasikan terhadap *evaluation dataset* data transaksi penjualan terkini.

Untuk menghitung tingkat kekuatan dilakukan dengan rumus :

$$\frac{\sum_{i=1}^n (S_i \times C_i)}{n}$$

dimana:

n = jumlah aturan asosiasi

S<sub>i</sub> = nilai support untuk aturan asosiasi ke-i

C<sub>i</sub> = nilai *confidence* untuk aturan asosiasi ke-i

Untuk menghitung besar persentase akurasi algoritma A terhadap algoritma B dilakukan dengan menggunakan rumus :

$$\frac{\sum \text{Support Algoritma A}}{\sum \text{Support Algoritma B}} \times 100$$

3. **Rancangan Penelitian**

3.1. **Data Penelitian**

Data yang akan digunakan di dalam penelitian ini adalah data transaksi pesanan cetakan di Percetakan PT. Gramedia yang terkumpul mulai dari tanggal 1 Januari 2005 hingga 1 Agustus 2010.

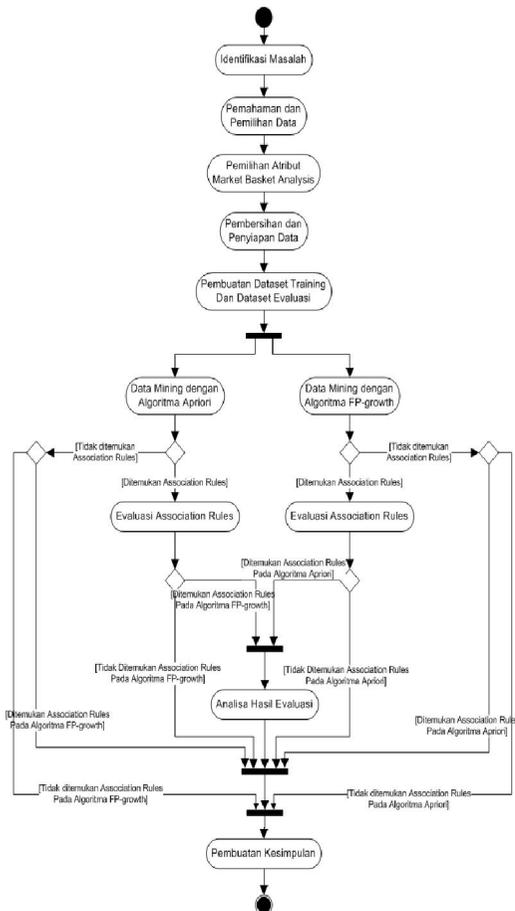
Tabel-tabel yang digunakan sebagai obyek penelitian ini adalah :

Tabel 1. Tabel obyek penelitian

No	Nama Tabel	Tuples
1	Tbl_Master_Order	147498
2	Tbl Master Pelanggan	2038
3	Tbl Rincian Order	258278
4	Tbl Order Sortir Pak	143620
5	Tbl Jenis Permintaan	6
6	Tbl Jenis Produk	18
7	Tbl Nama Kertas	182
8	Tbl Teknik Jilid	14
9	Tbl Jenis Cover	7
10	Tbl Jenis Laminating	6
11	Tbl Jenis UV	4
12	Tbl Master Site	4
13	Tbl Nama Komponen	5

3.2. Kerangka Kerja

Kerangka kerja penelitian dapat dilihat dalam bentuk diagram aktivitas berikut:



Gambar 4. Diagram aktivitas penelitian

3.3. Perangkat Lunak

Aplikasi-aplikasi yang digunakan untuk membantu proses penelitian *market basket analysis* terdiri dari :

a. SQL Server 2005 Express Edition

Dalam penelitian ini SQL Server 2005 Express Edition digunakan untuk menyimpan basis data yang akan digunakan dalam penelitian sebagai hasil dari proses pemilihan data penelitian. Setelah itu dilakukan proses pembersihan dan penyiapan data serta pembuatan *dataset* penelitian dengan menggunakan perangkat SQL Server Management Studio Express.

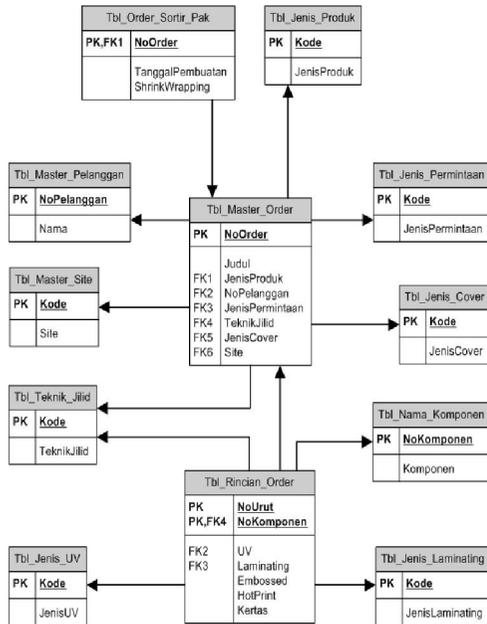
b. Weka 3.6

Aplikasi Weka digunakan dalam penelitian untuk mengimplementasikan model *data mining* terhadap basis data penelitian. Weka juga digunakan untuk menampilkan presentasi knowledge yang dapat disimpulkan dari hasil penerapan model *data mining* terhadap sejumlah atribut penjualan produk buku di Percetakan PT. Gramedia.

4. Analisis Dan Interpretasi

4.1. Entity Relationship Diagram (ERD)

Hubungan antar tabel yang digunakan dapat digambarkan dengan menggunakan sebuah diagram relasi antar entitas (*Entity Relationship Diagram*).



Gambar 5. Entity Relationship Diagram

4.2. Pemilihan Data

Pemilihan data mengacu pada permasalahan yang telah diidentifikasi sebelumnya.

Hasil dari proses pemilihan data terhadap masing-masing tabel adalah :

Tabel 1. Tabel hasil pemilihan data

Nama Tabel	Tuples
Tbl Master Order Ex	36558
Tbl Rincian Order Ex	69673
Tbl Order Sortir Pak Ex	35875
Tbl Master Pelanggan Ex	690
Tbl Jenis Permintaan Ex	2
Tbl Jenis Produk Ex	4
Tbl Nama Kertas Ex	54
Tbl Teknik Jilid Ex	9
Tbl Jenis Cover Ex	5
Tbl Jenis Laminating Ex	4
Tbl Jenis UV Ex	3
Tbl Master Site Ex	4
Tbl Nama Komponen Ex	3

4.3. Pemilihan Atribut

Dari setiap *item* data akan dihitung besar persentase jumlah frekuensi kemunculan *item* tersebut terhadap jumlah transaksi keseluruhan. Penentuan nilai batas minimum (*Threshold*) sebesar 0,2 %, sehingga setiap *item* data yang memiliki

persentase kemunculan melebihi atau sama dengan 0,2 % akan terpilih menjadi atribut.

Data-data persentase dari masing-masing komponen dapat dilihat pada sejumlah tabel berikut :

Tabel 2. Data frekuensi kandidat atribut *cover* buku

Item	Frek	Persen
Soft Cover	34678	96,54
Semi Hard Cover	57	0,16
Hard Cover	1185	3,3
<b>Total</b>	<b>35920</b>	<b>100</b>

Tabel 3. Data frekuensi kandidat atribut teknik penjilidan

Item	Frek	Persen
Benang	3	0,008
Kawat Spiral	230	0,63
Punch	8	0,02
Kawat	5127	14,05
Lem	29075	79,67
Lem + Benang	1962	5,38
Lipat	10	0,03
Sisip	1	0,002
Potong Jadi	77	0,21
<b>Total</b>	<b>36493</b>	<b>100</b>

Tabel 4. Data frekuensi kandidat atribut teknik UV

Item	Frek	Persen
UV Doft	16798	47,76
UV Spot	4765	13,55
Lainnya	13608	38,69
<b>Total</b>	<b>35171</b>	<b>100</b>

Tabel 5. Data frekuensi kandidat atribut teknik *laminating*

Item	Frek	Persen
<i>Laminating Gloss</i>	2865	8,15
<i>Laminating Doft</i>	7344	20,88
Lainnya	24962	70,97
<b>Total</b>	<b>35171</b>	<b>100</b>

Tabel 6. Data frekuensi kandidat atribut teknik *embossed* dan *hot print*

Item	Frek	Persen
<i>Embossed</i>	2877	8,18
<i>Hot Print</i>	295	0,8

Tabel 7. Data frekuensi kandidat atribut penggunaan plastik *shrink*

Item	Frek	Persen
<i>Shrink</i>	26992	75,24

Tabel 8. Data frekuensi kandidat atribut jenis kertas cover buku

Item	Frek	Persen
Art Paper	2601	7,4
Karton Board	1109	3,15
Art Karton	30489	86,69
Newsprint	9	0,02
Lainnya	963	2,74
<b>Total</b>	<b>35171</b>	<b>100</b>

Tabel 9. Data frekuensi kandidat atribut jenis kertas halaman isi buku

Item	Frek	Persen
Art Paper	4228	12,38
Newsprint	13307	39
HVS/ Woodfree	1713	5
HVS Sheet	10048	29,42
Koran Sheet	4346	12,72
Lainnya	510	0,01
<b>Total</b>	<b>34152</b>	<b>100</b>

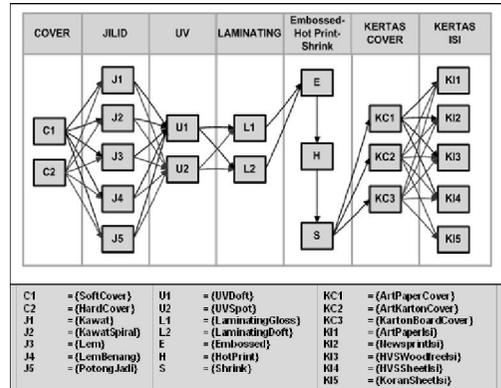
Dari hasil proses pemilihan atribut, dapat diperoleh sejumlah atribut yang memenuhi syarat *threshold* minimum dan dapat digunakan dalam *market basket analysis*. Seluruh atribut penelitian yang merupakan hasil seleksi dari setiap komponen penjualan produk buku dapat dilihat pada tabel berikut :

Tabel 10. Data frekuensi atribut terpilih

Komponen	Atribut Terpilih	Jumlah
Jenis cover	{Soft Cover}, {Hard Cover}	2
Teknik jilid	{Kawat}, {Lem}, {Lem + Benang}, {Kawat Spiral}, {Potong Jadi}	5
Shrink	{Shrink}	1
Jenis kertas cover	{Art Paper}, {Karton Board}, {Art Karton}	3
Jenis kertas isi	{Art Paper}, {Newsprint}, {HVS/Woodfree}, {HVS Sheet}, {Koran Sheet}	5
UV	{UV Doft}, {UV Spot}	2
Laminating	{Laminating Gloss}, {Laminating Doft}	2
Embossed	{Embossed}	1
Hot Print	{Hot Print}	1

**4.3.1. Kombinasi Atribut**

Kombinasi atribut yang dapat digunakan untuk pencarian aturan asosiasi tampak seperti pada gambar dibawah ini :



Gambar 6. Kombinasi atribut penelitian

**4.3.2. Penentuan Atribut Penelitian**

Dari sekian banyak kombinasi atribut yang ada, kombinasi atribut yang nantinya akan digunakan dalam penelitian ini adalah kombinasi dari atribut-atribut yang memiliki persentase kemunculan paling besar dari masing-masing komponen.

Tabel 11. Atribut yang digunakan dari masing-masing komponen

Komponen	Atribut Terpilih	Atribut Penelitian
Jenis cover	{SoftCover = 6,54%}, {HardCover = 3,3%}	{SoftCover}
Teknik jilid	{Kawat = 14,05%}, {Lem = 79,67%}, {LemBenang = 5,38%}, {KawatSpiral = 0,63%}, {PotongJadi = 0,21%}	{Lem}
UV	{UVDoft = 47,76%}, {UVSpot = 13,55%}	{UVDoft}
Laminating	{LaminatingGloss = 8,15%}, {LaminatingDoft = 20,88%}	{LaminatingDoft}
Embossed	{Embossed = 8,18%}	{Embossed}
Hot Print	{HotPrint = 0,8 %}	{HotPrint}
Shrink	{Shrink = 75,24%}	{Shrink}
Jenis kertas cover	{ArtPaperCover = 7,4 %}, {KartonBoardCover = 3,15%}, {ArtKartonCover = 86,69%}	{ArtKartonCover}
Jenis kertas isi	{ArtPaperIsi = 12,38%}, {NewsprintIsi = 39%}, {HVSWoodfreeIsi = 5%}, {HVSWoodfreeIsi = 29,42%}, {KoranSheetIsi = 12,72%}	{NewsprintIsi}

Berdasarkan nilai persentase terbesar dari masing-masing atribut, dengan

demikian atribut-atribut yang akan digunakan dalam penelitian adalah : jenis *cover hard cover*, teknik penjiplidan dengan menggunakan lem, proses : UV *doft*, laminating *doft*, *embossed* dan *hot print*, penggunaan plastik *shrink*, jenis kertas Art Karton untuk *cover* buku serta jenis kertas Newsprint untuk halaman isi buku.

**4.4. Pembersihan dan Penyiapan Data**

Perintah SQL yang digunakan untuk melakukan pembersihan data adalah perintah DML SQL.

Berikut adalah jumlah *tuples* terakhir dari masing-masing tabel transaksi setelah dilakukannya pembersihan data :

- a. Tbl\_Master\_Order : 32821 *Tuples*
- b. Tbl\_Rincian\_Order : 69996 *Tuples*
- c. Tbl\_Order\_Sortir : 32821 *Tuples*

**4.5. Pembuatan Dataset**

*Dataset* dibuat berdasarkan atribut-atribut yang telah dipilih pada data transaksi penjualan produk buku.

Berikut adalah struktur *dataset* yang dibuat :

Tabel 12. Struktur *dataset* penelitian

Nama Kolom	Tipe Data	Atribut Kolom	Keterangan
ID	INTE-GER	Primary Key	ID transaksi penjualan
SoftCover	CHAR(1)	NotNull	Atribut jenis <i>cover</i>
HardCover	CHAR(1)	NotNull	Atribut jenis <i>cover</i>
Kawat	CHAR(1)	NotNull	Atribut teknik penjiplidan
KawatSpiral	CHAR(1)	NotNull	Atribut teknik penjiplidan
Lem	CHAR(1)	NotNull	Atribut teknik penjiplidan
LemBerang	CHAR(1)	NotNull	Atribut teknik penjiplidan
PotongJadi	CHAR(1)	NotNull	Atribut teknik penjiplidan
UVDoft	CHAR(1)	NotNull	Atribut jenis UV
UVSpot	CHAR(1)	NotNull	Atribut jenis UV
LaminatingGloss	CHAR(1)	NotNull	Atribut jenis <i>laminating</i>
LaminatingDoft	CHAR(1)	NotNull	Atribut jenis <i>laminating</i>
Embossed	CHAR(1)	NotNull	Atribut <i>embossed</i>
HotPrint	CHAR(1)	NotNull	Atribut <i>hotprint</i>
Shrink	CHAR(1)	NotNull	Atribut <i>shrink</i>
ArtPaperCover	CHAR(1)	NotNull	Atribut kertas <i>cover</i>
ArtKartonCover	CHAR(1)	NotNull	Atribut kertas <i>cover</i>
KartonBoardCover	CHAR(1)	NotNull	Atribut kertas <i>cover</i>
ArtPaperIsi	CHAR(1)	NotNull	Atribut kertas halaman isi
NewsprintIsi	CHAR(1)	NotNull	Atribut kertas halaman isi
HVSWoodfreeIsi	CHAR(1)	NotNull	Atribut kertas halaman isi
HVSSheetIsi	CHAR(1)	NotNull	Atribut kertas halaman isi
KoranSheetIsi	CHAR(1)	NotNull	Atribut kertas halaman isi

Dataset dengan struktur tersebut diatas, dibuat sebanyak 3 buah dalam bentuk tabel dengan nama Tbl\_Dataset\_Training, Tbl\_Dataset\_Evaluation1, dan Tbl\_Dataset\_Evaluation2. Masing-masing *dataset* tersebut kemudian diisi dengan data-data sebagai berikut :

- a. Tabel Tbl\_Dataset\_Training diisi dengan seluruh data transaksi penjualan yang terdapat pada basis data penelitian.
- b. Tabel Tbl\_Dataset\_Evaluation1 diisi dengan 5000 data transaksi penjualan terakhir yang terdapat pada basis data penelitian.
- c. Tabel Tbl\_Dataset\_Evaluation2 diisi dengan 10000 data transaksi penjualan yang diambil secara acak dari basis data penelitian. Data acak yang diambil adalah data-data transaksi penjualan mulai tanggal 1 Januari 2008.

**4.6. Data Mining**

Proses data mining dilakukan terhadap tabel Tbl\_Dataset\_Training dengan menggunakan aplikasi Weka versi 3.6. Sepuluh aturan asosiasi dengan nilai *confidence* terbesar diambil untuk dievaluasi dan dianalisa dalam penelitian ini.

**4.6.1. Aturan Asosiasi Algoritma Apriori**

Aturan-aturan asosiasi yang diperoleh dari penerapan *market basket analysis* dengan algoritma apriori adalah seperti pada tabel berikut :

Tabel 13. Aturan asosiasi hasil algoritma Apriori

Urt	A	B	F <sub>A</sub>	F <sub>AB</sub>	Cnf
1	{Lem=y}, {Shrink=y}, {ArtKartonCover=y}	{SoftCover=y}	23849	23846	1
2	{Lem=y}, {ArtKartonCover=y}	{SoftCover=y}	26042	26035	1
3	{Lem=y}, {Shrink=y}	{SoftCover=y}	24078	24068	1
4	{Lem=y}	{SoftCover=y}	26705	26684	1
5	{Shrink=y}, {ArtKartonCover=y}	{SoftCover=y}	24864	24814	1

Tabel 13. Aturan asosiasi hasil algoritma Apriori (Lanjutan)

Urt	A	B	F <sub>A</sub>	F <sub>AB</sub>	Cnf
6	{ArtKartonCover=y}	{SoftCover=y}	29947	29878	1
7	{SoftCover=y}, {Lem=y}, {Shrink=y}	{ArtKartonCover=y}	24068	23846	0,99
8	{Lem=y}, {Shrink=y}	{ArtKartonCover=y}	24078	23849	0,97
9	{Lem=y}, {Shrink=y}	{SoftCover=y}, {ArtKartonCover=y}	24078	23846	0,94
10	{SoftCover=y}, {Shrink=y}	{ArtKartonCover=y}	25086	24814	0,94

Aturan-aturan asosiasi yang dihasilkan oleh algoritma Apriori ditentukan oleh sejumlah kombinasi *items* sebagai berikut :

Tabel 14. Kombinasi *items* pada aturan asosiasi Apriori

Ukuran <i>Itemset</i>	<i>Items</i>
2	{Lem,SoftCover}, {ArtKartonCover,SoftCover}
3	{Lem, ArtKartonCover, SoftCover}, {Lem, Shrink, SoftCover}, {Shrink, ArtKartonCover, SoftCover}, {Lem, Shrink, ArtKartonCover}
4	{Lem,Shrink,ArtKartonCover,SoftCover}

Dari tabel di atas dapat dilihat bahwa algoritma apriori menemukan adanya hubungan antara teknik penjilidan lem, penggunaan plastik shrink, jenis kertas Art Karton untuk *cover* buku, dan jenis *cover* soft cover, yang digambarkan oleh aturan-aturan asosiasi yang dihasilkan.

#### 4.6.2. Aturan Asosiasi Algoritma FP-growth

Aturan-aturan asosiasi yang diperoleh dari penerapan *market basket analysis* dengan algoritma FP-growth adalah seperti pada tabel berikut :

Tabel 15. Aturan asosiasi hasil algoritma FP-growth

Urt	A	B	F <sub>A</sub>	F <sub>AB</sub>	Cnf
1	{Shrink=y}, {UVDof=y}, {Newsprints=y}	{Lem=y}	8199	8189	1
2	{Shrink=y}, {Newsprints=y}	{Lem=y}	12306	12228	0,99

Tabel 15. Aturan asosiasi hasil algoritma FP-growth (Lanjutan)

3	{Lem=y}, {UVDof=y}, {Newsprints=y}	{Shrink=y}	8336	8189	0,98
4	{Lem=y}, {Newsprints=y}	{Shrink=y}	12465	12228	0,98
5	{Shrink=y}, {UVDof=y}	{Lem=y}	13529	13247	0,98
6	{UVDof=y}, {Newsprints=y}	{Lem=y}	8540	8336	0,98
7	{UVDof=y}, {Newsprints=y}	{Shrink=y}	8540	8199	0,96
8	{Newsprints=y}	{Lem=y}	12998	12465	0,96
9	{UVDof=y}, {Newsprints=y}	{Lem=y}, {Shrink=y}	8540	8189	0,96
10	{Newsprints=y}	{Shrink=y}	12998	12306	0,95

Aturan-aturan asosiasi yang dihasilkan oleh algoritma FP-growth ditentukan oleh sejumlah kombinasi *items* sebagai berikut :

Tabel 16. Kombinasi *items* pada aturan asosiasi FP-growth

Ukuran <i>itemset</i>	<i>Items</i>
2	{Newsprints,Lem},{Newsprints,Shrink}
3	{Shrink,Newsprints,Lem}, {Shrink,UVDof,Lem}, {UVDof, Newsprints, Lem}, {UVDof, Newsprints, Shrink}
4	{Shrink,UVDof,Newsprints,Lem}

Dari tabel di atas dapat dilihat bahwa algoritma FP-growth menemukan adanya hubungan antara penggunaan plastik *shrink*, proses UV Dof pada *cover*, jenis kertas Newsprint untuk isi halaman buku, dan teknik jilid lem, yang digambarkan oleh aturan-aturan asosiasi yang dihasilkan.

#### 4.7. Evaluasi Aturan Asosiasi

Evaluasi aturan asosiasi dilakukan dengan menggunakan 2 buah *evaluation dataset*, yang telah disiapkan sebelumnya. Evaluasi dilakukan dengan mengukur generalitas dan reliabilitas dari masing-masing aturan yang dihasilkan.

##### 4.7.1. Evaluasi Aturan Asosiasi Apriori

Hasil evaluasi terhadap *evaluation dataset* yang tersimpan pada tabel Tbl\_Dataset\_Evaluation1 dapat dilihat pada tabel berikut :

Tabel 17. Evaluasi aturan asosiasi Apriori terhadap Tbl\_Dataset\_Evaluation1

Aturan Urut	Generalitas		Reliabilitas		
	Support	Coverage	Confidence	Added Value	Correlation
1	0,6806	0,6806	1	0,0434	0,4157
2	0,7658	0,7664	0,9992	0,0426	0,4444
3	0,6896	0,6896	1	0,0434	0,4189
4	0,7868	0,7876	0,9989	0,0423	0,451
5	0,711	0,7124	0,998	0,0414	0,426
6	0,9112	0,9138	0,9971	0,0405	0,4871
7	0,6806	0,6896	0,9869	0,0731	0,4244
8	0,6806	0,6896	0,9869	0,0731	0,4244
9	0,6806	0,6896	0,9869	0,0757	0,4251
10	0,711	0,7212	0,9858	0,072	0,4348
Rata-rata	0,72978	0,73404	0,99397	0,05475	0,43518

Hasil evaluasi terhadap *evaluation dataset* yang disimpan pada tabel Tbl\_Dataset\_Evaluation2 dapat dilihat pada tabel berikut :

Tabel 18. Evaluasi aturan asosiasi Apriori terhadap Tbl\_Dataset\_Evaluation2

Aturan Urut	Generalitas		Reliabilitas		
	Support	Coverage	Confidence	Added Value	Correlation
1	0,7107	0,711	0,9995	0,0464	0,427
2	0,7851	0,7857	0,9992	0,0461	0,4515
3	0,7189	0,7195	0,9991	0,046	0,4298
4	0,8106	0,8121	0,9981	0,045	0,4592
5	0,7361	0,7382	0,9971	0,044	0,4352
6	0,9014	0,9048	0,9962	0,0431	0,4851
7	0,7107	0,7189	0,9885	0,0837	0,4377
8	0,711	0,7195	0,9881	0,0833	0,4377
9	0,7107	0,7195	0,9877	0,0863	0,4384
10	0,7361	0,7459	0,9868	0,082	0,4459
Rata-rata	0,75313	0,75751	0,99403	0,06059	0,44475

#### 4.7.2. Evaluasi Aturan Asosiasi FP-growth

Hasil evaluasi terhadap *evaluation dataset* yang disimpan pada tabel Tbl\_Dataset\_Evaluation1 dapat dilihat pada tabel berikut :

Tabel 19. Evaluasi aturan asosiasi FP-growth terhadap Tbl\_Dataset\_Evaluation1

Aturan Urut	Generalitas		Reliabilitas		
	Support	Coverage	Confidence	Added Value	Correlation
1	0,2326	0,2338	0,9948	0,2072	0,2277
2	0,338	0,3418	0,9888	0,2012	0,2992
3	0,2326	0,24	0,9691	0,2163	0,2342

Tabel 19. Evaluasi aturan asosiasi FP-growth terhadap Tbl\_Dataset\_Evaluation1 (Lanjutan)

Aturan Urut	Generalitas		Reliabilitas		
	Support	Coverage	Confidence	Added Value	Correlation
4	0,338	0,346	0,9768	0,224	0,3076
5	0,2992	0,304	0,9842	0,1966	0,274
6	0,24	0,2506	0,9577	0,1701	0,2311
7	0,2338	0,2506	0,9329	0,1801	0,233
8	0,346	0,3594	0,9627	0,1751	0,3016
9	0,2326	0,2506	0,9281	0,2385	0,2473
10	0,3418	0,3594	0,951	0,1982	0,3073
Rata-rata	0,28346	0,29362	0,96461	0,20073	0,2663

Hasil evaluasi terhadap *evaluation dataset* yang disimpan pada tabel Tbl\_Dataset\_Evaluation2 dapat dilihat pada tabel berikut :

Tabel 20 Evaluasi aturan asosiasi FP-growth terhadap Tbl\_Dataset\_Evaluation2

Aturan Urut	Generalitas		Reliabilitas		
	Support	Coverage	Confidence	Added Value	Correlation
1	0,2462	0,2468	0,9975	0,1854	0,2325
2	0,3606	0,3635	0,992	0,1799	0,3067
3	0,2462	0,2486	0,9903	0,2094	0,2391
4	0,3606	0,3634	0,9922	0,2113	0,3151
5	0,3524	0,3599	0,9791	0,167	0,3006
6	0,2486	0,2562	0,9703	0,1582	0,2327
7	0,2468	0,2562	0,9633	0,1824	0,2379
8	0,3634	0,3735	0,9729	0,1608	0,3065
9	0,2462	0,2562	0,9609	0,2414	0,2523
10	0,3635	0,3735	0,9732	0,1923	0,3148
Rata-rata	0,30345	0,30978	0,97917	0,18881	0,27382

#### 4.8. Analisa Hasil Evaluasi

Analisa hasil evaluasi dilakukan dengan menghitung tingkat kekuatan dari aturan-aturan asosiasi yang dihasilkan oleh masing-masing algoritma. Tingkat kekuatan dari aturan asosiasi ditentukan oleh nilai *support* yang mewakili aspek generalitas dan nilai *confidence* yang mewakili aspek reliabilitas.

Berikut adalah pengukuran tingkat kekuatan terhadap aturan-aturan asosiasi yang dihasilkan oleh algoritma Apriori :

Tabel 21. Tingkat kekuatan aturan asosiasi algoritma Apriori

Aturan Asosiasi	Support	Confidence	Supp×Conf
{Lem, Shrink, ArtKartonCover}⇒ {SoftCover}	0,6806	1	0,6806
{Lem, ArtKartonCover}⇒ {SoftCover}	0,7658	0,9992	0,765187
{Lem, Shrink}⇒{SoftCover}	0,6896	1	0,6896
{Lem}⇒{SoftCover}	0,7868	0,9989	0,785935
{Shrink, ArtKartonCover}⇒ {SoftCover}	0,711	0,998	0,709578
{ArtKartonCover}⇒{SoftCover}	0,9112	0,9971	0,908558
{SoftCover, Lem, Shrink}⇒ {ArtKartonCover}	0,6806	0,9869	0,671684
{Lem, Shrink}⇒ {ArtKartonCover}	0,6806	0,9869	0,671684
{Lem, Shrink}⇒ {SoftCover, ArtKartonCover}	0,6806	0,9869	0,671684
{SoftCover, Shrink}⇒ {ArtKartonCover}	0,711	0,9858	0,700904
Tingkatkekuatanaturanasosiasi			0,725541

Sedangkan hasil pengukuran tingkat kekuatan terhadap aturan-aturan asosiasi yang dihasilkan oleh algoritma FP-growth adalah :

Tabel 22. Tingkat kekuatan aturan asosiasi algoritma FP-growth

Aturan Asosiasi	Support	Confidence	Supp×Conf
{Shrink, UVDof, Newsprints}⇒ {Lem}	0,2326	0,9948	0,23139
{Shrink, Newsprints}⇒{Lem}	0,338	0,9888	0,334214
{Lem, UVDof, Newsprints}⇒ {Shrink}	0,2326	0,9691	0,225413
{Lem, Newsprints}⇒{Shrink}	0,338	0,9768	0,330158
{Shrink, UVDof}⇒{Lem}	0,2992	0,9842	0,294473
{UVDof, Newsprints}⇒{Lem}	0,24	0,9577	0,229848
{UVDof, Newsprints}⇒{Shrink}	0,2338	0,9329	0,218112
{Newsprints}⇒{Lem}	0,346	0,9627	0,333094
{UVDof, Newsprints}⇒{Lem, Shrink}	0,2326	0,9281	0,215876
{Newsprints}⇒{Shrink}	0,3418	0,951	0,325052
Tingkatkekuatanaturanasosiasi			0,273763

Dari hasil analisa diketahui bahwa tingkat kekuatan aturan-aturan asosiasi yang dihasilkan oleh algoritma Apriori lebih besar dibandingkan dengan yang dihasilkan oleh algoritma FP-growth, penyebabnya adalah rendahnya nilai *support* yang menggambarkan terjadinya

kombinasi *items* dalam aturan terhadap keseluruhan data transaksi terkini penjualan produk buku.

Tingkat akurasi dari algoritma Apriori terhadap FP-growth adalah sebesar:

$$\frac{\sum Support Apriori}{\sum Support FP - growth} \times 100 = \frac{7,2978}{2,8346} \times 100 = 257,4543 \%$$

### 5. Kesimpulan Dan Saran

Beberapa kesimpulan yang dapat diambil dari hasil penelitian ini adalah :

- Dari hasil observasi terhadap data-data sejumlah komponen transaksi penjualan produk buku yang terjadi percetakan PT. Gramedia hingga tanggal 1 Agustus 2010, telah ditemukan sebanyak 22 atribut-atribut yang dapat dimanfaatkan untuk melakukan penelitian *market basket analysis*.
- Penelitian *market basket analysis* dengan menggunakan algoritma Apriori dan FP-growth terhadap atribut-atribut penjualan produk buku telah menghasilkan sejumlah aturan asosiasi yang berbeda antara algoritma yang satu dengan yang lainnya.
- Analisa yang dilakukan terhadap tingkat kekuatan aturan-aturan asosiasi menunjukkan bahwa aturan-aturan asosiasi yang dihasilkan oleh algoritma Apriori memiliki tingkat kekuatan yang lebih tinggi dibandingkan dengan yang dihasilkan oleh algoritma FP-growth.

Beberapa saran yang dapat diberikan antara lain :

- Perbaiki prosedur kerja yang berhubungan dengan proses pemasukan data-data transaksi ke dalam sistem basis data perlu mendapat perhatian yang khusus sehingga ketersediaan data pada saat diperlukan dapat lebih terjamin.
- Dapat dilakukan penelitian lebih lanjut dengan memanfaatkan algoritma-algoritma *association data mining* lainnya, seperti: *Generalized Rule Induction, Algoritma Hash Based, Predictive Apriori*, dan *Tertius* untuk

- penelitian aturan-aturan asosiasi berdasarkan atribut-atribut penjualan produk buku di Percetakan PT. Gramedia.
- c. Perlu dilakukan penelitian untuk mengkaji kelemahan pada algoritma FP-growth yang mengakibatkan lemahnya aturan-aturan asosiasi yang dihasilkan melalui penerapan *market basket analysis* terhadap atribut penjualan produk buku di Percetakan PT. Gramedia.
- d. Pemanfaatan teknologi *data warehouse* pada basis data SQL Server sehingga proses penyiapan data yang merupakan syarat awal untuk melakukan *data mining* dapat dilakukan dengan lebih cepat dan efisien. Disamping itu, kualitas dan integritas data yang tersimpan dalam *data warehouse* lebih terjamin.
- www.crisp-dm.org/Process/index.htm*, (Diakses 20 Agustus 2010).
- [7] Fayyad, U., Shapiro, G. Piatetsky, dan Smyth, N Padhraic, "From Data Mining to Knowledge Discovery in Databases", *AI Magazine*, Hal. 37-54, 1996.

#### Daftar Pustaka

- [1] Kusriani dan Emha Taufiq Luthfi, "ALGORITMA DATA MINING", PENERBIT ANDI, 2009.
- [2] Budi Santosa, "Data Mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis", GRAHA ILMU, 2007
- [3] Jiawei Han dan Micheline Kamber, "Data Mining: Concepts and Techniques : Chapter 6. Mining Association Rules in Large Databases", Simon Fraser University, 2000.
- [4] Gregorius Satia Budhi, Andreas Handoyo dan Christine Oktavina Wirawan, 2009, "Algoritma Generalized Sequential Pattern Untuk Menggali Data Sekuensial Sirkulasi Buku Pada Perpustakaan Uk Petra", Seminar Nasional Aplikasi Teknologi Informasi, 2009.
- [5] Kusriani, "Penerapan Algoritma Apriori Pada Data Mining Untuk Mengelompokkan Barang Berdasarkan Kecenderungan Kemunculan Bersama Dalam Satu Transaksi", E-learning Center STMIK AMIKOM Yogyakarta, 2007
- [6] CRISP-DM, "Cross Industry Standard Process for Data Mining", <http://>