

## ANALISIS DAN IMPLEMENTASI ALGORITMA FP-GROWTH PADA APLIKASI SMART UNTUK MENENTUKAN MARKET BASKET ANALYSIS PADA USAHA RETAIL (STUDI KASUS : PT.X)

### ANALYSIS AND IMPLEMENTATION OF FP-GROWTH ALGORITHM IN SMART APPLICATION TO DETERMINE MARKET BASKET ANALYSIS ON RETAIL BUSINESS (CASE STUDY : PT.X)

Dyah Pramesthi Larasati<sup>1</sup>, Muhammad Nasrun, S.Si., MT.<sup>2</sup>, Umar Ali Ahmad, ST., MT.<sup>3</sup>

<sup>1,2,3</sup>Prodi S1 Sistem Komputer, Fakultas Teknik, Universitas Telkom Bandung

<sup>1</sup>[dyah.p.larasati@gmail.com](mailto:dyah.p.larasati@gmail.com), <sup>2</sup>[nasrun@telkomuniversity.ac.id](mailto:nasrun@telkomuniversity.ac.id), <sup>3</sup>[umar@telkomuniversity.ac.id](mailto:umar@telkomuniversity.ac.id)

#### Abstrak

Bisnis *retail* membutuhkan suatu aplikasi yang dapat membantu untuk menganalisis produk mana saja yang sering dibeli oleh pelanggan dan kebiasaan pelanggan membeli produk yang satu dengan yang lainnya. Jika terdapat produk yang disediakan tidak sesuai dengan kebutuhan pelanggan, maka dapat menyebabkan kerugian pada bisnis tersebut karena produk tidak berhasil terjual. Oleh karena itu, aplikasi SMART – Sistem Informasi Manajemen *Retail* dibangun untuk memenuhi kebutuhan tersebut.

Aplikasi SMART dalam *Market Basket Analysis* dibangun dengan menggunakan algoritma *FP-Growth*. Algoritma *FP-Growth* dapat menghasilkan *frequent itemset* yang nantinya akan digunakan pada proses *association rules*. Proses *association rules* akan menghasilkan suatu *rule* yang dipengaruhi oleh nilai *minimum support* dan *minimum confidence*. Penelitian ini membahas masalah tentang bagaimana sistem kerja *Market Basket Analysis* dengan menggunakan algoritma *FP-Growth*, perancangan sistem dengan menggunakan algoritma *FP-Growth*, dan menetapkan *frequent itemset* dan *rule*.

Hasil pengujian yang diperoleh adalah rata-rata pada nilai *minsup*=0.006 dan *minconf*=0.6 tidak menghasilkan *rules*, dan rata-rata waktu proses adalah 957 ms..

**Kata Kunci :** FP-Growth, FP-Tree, Association Rule, data mining, retail

#### Abstract

*Retail business requires an application that can help to analyze products that frequently purchased by customers and buying habits of customers for one product with other product. If there is a product that is not supplied in accordance with customer needs, it can cause harm to the business because the product is unsuccessfully sold. Therefore, the application of SMART - Retail Management Information System is built to meet these needs.*

*SMART Application in Market Basket Analysis is built by using the FP-Growth algorithm. FP-Growth algorithm can generate frequent itemset that will be used in the process of association rules. The process of association rules will generate a rule that is affected by the value of minimum support and minimum confidence. In this study addresses the issue of how the Market Basket Analysis' systems work using the FP-Growth algorithm, system design using the FP-Growth algorithm, and assign frequent itemset and rules.*

*The results of the analysis are there is not rules generated if the value of mnsupp=0.006 and minconf=0.6, and the average of running time is 957 ms.*

**Key words :** FP-Growth, FP-Tree, Association Rule, data mining, retail

#### 1. Pendahuluan

Masyarakat yang berkecimpung pada bidang bisnis sering membutuhkan suatu teknologi baru sebagai alat bantu mereka dalam menyelesaikan suatu pekerjaan. Misalnya suatu teknologi yang membantu mereka untuk memberikan analisis yang berhubungan dengan bisnis mereka. Salah satu bisnis yang membutuhkan teknologi seperti ini adalah bisnis *retail*.

Bisnis *retail* membutuhkan teknologi tersebut misalnya untuk memberikan analisis apakah produk yang dijualnya sudah memenuhi kebutuhan pelanggan atau belum, produk apa saja yang biasa dibeli pelanggan, dan kebiasaan pelanggan membeli produk yang satu dengan yang lainnya. Jika terdapat produk yang tidak sesuai dengan kebutuhan pelanggan namun tetap dipertahankan bisa menyebabkan kerugian pada bisnis tersebut karena produk tidak laku dan menjadi kadaluwarsa. Sehingga perusahaan hanya menyediakan produk yang biasa dibeli

oleh pelanggan untuk menghindari kerugian tersebut dan mendapatkan keuntungan. Oleh karena itu, untuk memenuhi kebutuhan masyarakat tersebut penulis membangun aplikasi SMART – Sistem Informasi Manajemen Retail dengan menggunakan algoritma *FP-Growth* untuk menentukan produk yang harus disediakan oleh perusahaan berdasarkan *Market Basket Analysis*.

## 2. Dasar Teori

### 2.1 Algoritma FP-Growth

Algoritma *FP-Growth* merupakan pengembangan dari algoritma Apriori. Algoritma *Frequent Pattern Growth* adalah salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah kumpulan data. [1]

Pada algoritma *FP-Growth* menggunakan konsep pembangunan *tree*, yang biasa disebut *FP-Tree*, dalam pencarian *frequent itemsets* bukan menggunakan *generate candidate* seperti yang dilakukan pada algoritma *Apriori*. Dengan menggunakan konsep tersebut, algoritma *FP-Growth* menjadi lebih cepat daripada algoritma *Apriori*. [2]

Metode *FP-Growth* dibagi menjadi tiga tahapan utama, yaitu [3]:

1. tahap pembangkitan *conditional pattern base*,
2. tahap pembangkitan *conditional FP-Tree*, dan
3. tahap pencarian *frequent itemset*.

### 2.2 Association Rule

*Association rule* merupakan suatu proses pada data mining untuk menentukan semua aturan asosiatif yang memenuhi syarat minimum untuk *support (minsup)* dan *confidence (minconf)* pada sebuah database. Kedua syarat tersebut akan digunakan untuk *interesting association rules* dengan dibandingkan dengan batasan yang telah ditentukan, yaitu *minsup* dan *minconf*. [4]

*Association Rule Mining* adalah suatu prosedur untuk mencari hubungan antar *item* dalam suatu *dataset*. Dimulai dengan mencari *frequent itemset*, yaitu kombinasi yang paling sering terjadi dalam suatu *itemset* dan harus memenuhi *minsup*. [4]

Dalam tahap ini akan dilakukan pencarian kombinasi *item* yang memenuhi syarat minimum dari nilai *support* dalam database. Untuk mendapatkan nilai *support* dari suatu *item* A dapat diperoleh dengan rumus berikut:

$$\text{Support}(A) = \frac{\text{count}(A)}{\text{count}(\text{Database})} \quad (1) [4]$$

Kemudian, untuk mendapatkan nilai *support* dari dua *item* dapat diperoleh dengan rumus berikut:

$$\text{Support}(A,B) = P(A \cap B) = \frac{\text{count}(A \cap B)}{\text{count}(\text{Database})} \quad (2) [5]$$

Setelah semua *frequent item* dan *large itemset* didapatkan, dapat dicari syarat minimum *confidence (minconf)* dengan menggunakan rumus berikut:

$$\text{Confidence}(A \rightarrow B) = P(A|B) = \frac{\text{count}(A \cap B)}{\text{count}(B)} \quad (3) [5]$$

## 3. Perancangan Sistem dan Implementasi

### 3.1 Dataset

Tahap ini adalah tahap dimana memilih *dataset* yang digunakan untuk proses *datamining*.

Tabel 1 Contoh Tabel Penjualan pada Database

Tanggal	No. Transaksi	Nama Barang
10/2/2013	21020130001	Gula
10/2/2013	21020130001	Kopi

10/2/2013	21020130001	Teh
10/2/2013	21020130002	Gula
10/2/2013	21020130002	Kopi
10/2/2013	21020130002	Teh
10/2/2013	21020130003	Gula
10/2/2013	21020130003	Susu
10/2/2013	21020130003	Roti
10/2/2013	21020130004	Roti
10/2/2013	21020130004	Gula
10/2/2013	21020130004	Air
10/2/2013	21020130005	Gula
10/2/2013	21020130005	Susu
10/2/2013	21020130005	Kopi

**3.2 FP-Tree**

Tahap ini adalah tahap dimana dataset yang telah dibatasi dengan menggunakan *support count* yang telah ditentukan, kemudian dibangun menjadi sebuah *Tree*.

Berikut ini adalah tabel dengan semua barang yang dalam satu transaksi sudah disatukan.

Tabel 2 Data Transaksi Awal

No. Transaksi	Nama Barang
21020130001	Gula, Kopi, Teh
21020130002	Gula, Kopi, Teh
21020130003	Gula, Susu, Roti
21020130004	Roti, Gula, Air
21020130005	Gula, Susu, Kopi

Kemudian menentukan frekuensi setiap *item* dari keseluruhan transaksi.

Tabel 3 Nama dan Frekuensi Item dari Data Transaksi Awal

Nama Barang	Jumlah
Gula	5
Kopi	3
Teh	2
Susu	2
Roti	2
Air	1

Setelah frekuensi setiap *item* diperoleh, kemudian dibatasi dengan *support count*. Jika frekuensi *item* tidak kurang dari *support count*, maka *item* tersebut akan dihapus dan tidak dipakai dalam proses *data mining*. Misalkan ditentukan *support count*  $\xi = 2$ , maka hasilnya adalah:

Tabel 4 Nama dan Frekuensi Item Setelah Proses Filter

Nama Barang	Jumlah
Gula	5
Kopi	3
Roti	2

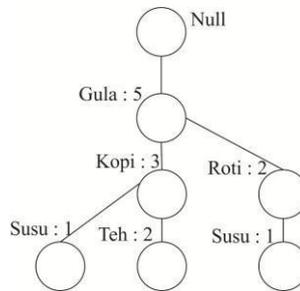
Susu	2
Teh	2

Item Air hilang karena frekuensinya tidak lebih dari sama dengan 2. Tahap selanjutnya adalah pembangunan *Tree* berdasarkan per transaksi dengan item yang telah dibatasi.

Tabel 5 Data Transaksi Setelah Proses *Filter*

No. Transaksi	Nama Barang
21020130001	Gula, Kopi, Teh
21020130002	Gula, Kopi, Teh
21020130003	Gula, Roti, Susu
21020130004	Gula, Roti
21020130005	Gula, Kopi, Susu

Pembangunan *Tree* dari seluruh transaksi tersebut adalah:



Gambar 1 *Tree* dari 5 Transaksi

### 3.3 FP-Growth

Setelah FP-Tree terbentuk, maka langkah selanjutnya adalah tahap pembangkitan *conditional pattern base*, tahap pembangkitan *conditional FP-Tree*, dan tahap pencarian *frequent itemset*. Pada tahap ini dapat dilakukan dengan melihat kembali FP-Tree yang sudah dibuat sebelumnya.

Tabel 6 *Conditional Pattern Base*

Item	<i>Conditional Pattern Base</i>
Teh	{{Gula, Kopi : 2}}
Susu	{{Gula, Kopi : 1}, {Gula, Roti : 1}}
Roti	{{Gula : 2}}
Kopi	{{Gula : 3}}

Tabel 7 *Conditional FP-Tree*

Item	<i>Conditional FP-Tree</i>
Teh	{{Gula : 2}, {Kopi : 2}, {Gula, Kopi : 2}}
Susu	{{Gula : 2}}
Roti	{{Gula : 2}}
Kopi	{{Gula : 3}}

Tabel 8 *Frequent Itemset*

Item	<i>Frequent Itemset</i>
Teh	{{Gula, Teh : 2}, {Kopi, Teh : 2}, {Gula, Kopi, Teh : 2}}
Susu	{{Gula, Susu : 2}}
Roti	{{Gula, Roti : 2}}
Kopi	{{Gula, Kopi : 3}}

### 3.4 Association Rule

Pada tahap ini digunakan untuk menentukan nilai *support* dan *confidence* pada setiap itemset dengan rumus yang sudah dijelaskan sebelumnya pada dasar teori. Pada kasus di atas, misalkan diberikan nilai *minimum support* = 0,4 dan *minimum confidence*= 1, maka hasilnya adalah:

Tabel 9 Hasil Association Rules

<i>Rule</i>	<i>Support</i>	<i>Confidence</i>
Roti => Gula	0,4	1
Susu => Gula	0,4	1
Teh => Gula	0,4	1
Teh => Kopi	0,4	1
Teh => Gula, Kopi	0,4	1
Kopi => Gula	0,6	1

## 4. Pengujian Sistem dan Analisis

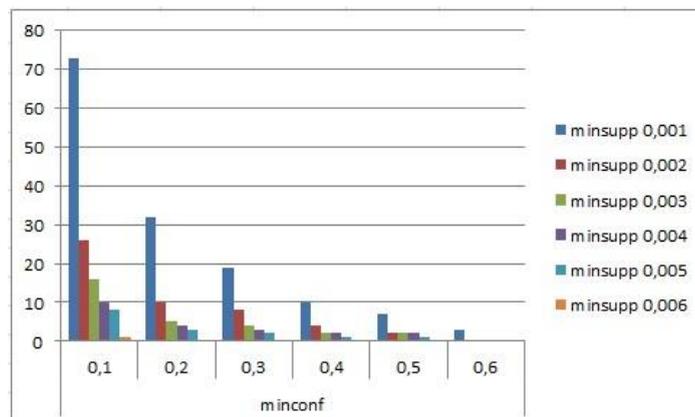
### 4.1 Pengujian Perbandingan Jumlah Rule

Pengujian untuk membandingkan hasil jumlah *rule* yang diperoleh dengan memasukkan *minsupp* dan *minconf* yang berbeda-beda dengan jumlah dataset sama.

Data transaksi yang digunakan adalah data bulan Januari 2014 dengan dataset sejumlah 4541 transaksi. Hasil yang diperoleh adalah sebagai berikut:

Tabel 9 Hasil Pengujian Perbandingan Jumlah Rule

<i>minsupp</i>	<i>minconf</i>					
	0,1	0,2	0,3	0,4	0,5	0,6
0,001	73	32	19	10	7	3
0,002	26	10	8	4	2	0
0,003	16	5	4	2	2	0
0,004	10	4	3	2	2	0
0,005	8	3	2	1	1	0
0,006	1	0	0	0	0	0



Gambar 2 Grafik Hasil Pengujian Perbandingan Jumlah Rule

Dilihat dari hasil pengujian diatas, terlihat bahwa untuk *minsupp*=0,006 dan *minconf*=0,6 terlihat tidak efektif untuk hasil analisis.

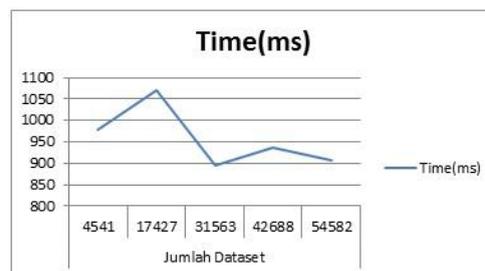
## 4.2 Pengujian Perbandingan Waktu Analisis

Pengujian untuk membandingkan lama waktu analisis dengan jumlah dataset yang berbeda-beda, namun dengan *minsupp* dan *minconf* yang sama.

Pengujian ini menggunakan *minsupp* = 0,001 dan *minconf* = 0,1. Adapun hasil pengujiannya adalah:

Tabel 10 Hasil Pengujian Perbandingan Waktu Analisis

	Jumlah Dataset				
	4541	17427	31563	42688	54582
Time (ms)	977	1070	894	936	908



Gambar 3 Grafik Hasil Pengujian Perbandingan Waktu Analisis

Dari hasil pengujian di atas dapat disimpulkan bahwa banyaknya dataset tidak mempengaruhi lama program eksekusi. Hasilnya menandakan bahwa tidak stabil.

## 4.3 Pengujian Hasil Rules per Bulan

Pengujian untuk melihat bagaimana hasil *rules* yang dihasilkan per bulan dengan nilai *minsupp*=0.002 dan *minconf*=0.5.

### 1. Bulan Januari 2014

No.	Rules	Support	Confidence
1.	PULPY ORANGE 350 ML ==> SOES / MARTABAK/ GEHU PEDAS/ CENTE/ PUDING MANIS/ PISANG	0.0079	0.5989
2.	TEH KOTAK JASMINE SUGAR 300ML ==> SOES / MARTABAK/ GEHU PEDAS/ CENTE/ PUDING MANIS/ PISANG	0.0057	0.64

### 2. Bulan Februari 2014

No.	Rules	Support	Confidence
1.	RISOLES AYAM/ RISOL ==> SOES / MARTABAK/ GEHU PEDAS/ CENTE/ PUDING MANIS/ PISANG	0.007	0.5357
2.	PASTEL / BAKWAN/ ROTI GORENG/ DADAR G ==> SOES / MARTABAK/ GEHU PEDAS/ CENTE/ PUDING MANIS/ PISANG	0.006	0.5347

### 3. Bulan Maret 2014

No.	Rules	Support	Confidence
1.	RISOLES AYAM/ RISOL ==> SOES / MARTABAK/ GEHU PEDAS/ CENTE/ PUDING MANIS/ PISANG	0.0079	0.5989
2.	PASTEL / BAKWAN/ ROTI GORENG/ DADAR G ==> SOES / MARTABAK/ GEHU PEDAS/ CENTE/ PUDING MANIS/ PISANG	0.0057	0.64

## 4. Bulan April 2014

No.	Rules	Support	Confidence
1.	RISOLES AYAM/ RISOL ==> KERIPIK SINGKONG ALIFIA	0.0079	0.5989
2.	FANTA STRAW 425 ML ==> KERIPIK SINGKONG ALIFIA	0.0057	0.64

## 5. Bulan Mei 2014

No.	Rules	Support	Confidence
1.	FEAST COKLAT ==> GARMELIA SANDWICH COKLAT	0.0047	0.6341

Berdasarkan hasil analisis di atas, pengujian ini menunjukkan bahwa dengan  $minsupp=0.002$  dan  $minconf=0.5$  jumlah *rule* yang dihasilkan pada setiap bulannya hampir sama dan terdapat produk yang muncul di beberapa bulan yang berbeda.

## 5. Kesimpulan dan Saran

## 5.1 Kesimpulan

Dari hasil pengujian dan analisa yang telah dilakukan, maka dapat disimpulkan bahwa:

1. pengimplementasian perancangan *Market Basket Analysis* dengan menggunakan *FP-Growth* di aplikasi SMART sudah berhasil,
2. pengimplementasian algoritma *FP-Growth* di aplikasi SMART sudah tervalidasi untuk melakukan *Market Basket Analysis*,
3. *rules* yang dihasilkan di setiap bulan dengan  $minsupp=0.002$  dan  $minconf=0.5$ , jumlahnya hampir sama dan ada beberapa produk yang muncul di beberapa bulan yang berbeda,
4. semakin besar  $minsupp$  dan  $minconf$  yang dimasukkan, maka menghasilkan *rule* yang semakin sedikit. Rata-rata dengan nilai  $minsupp=0.006$  dan  $minconf = 0.6$  tidak menghasilkan *rules*,
5. jumlah *dataset* tidak mempengaruhi waktu proses dalam analisis ini. Rata-rata waktu proses pada analisis ini adalah 957 ms.

## 5.2 Saran

Saran yang dapat diberikan oleh penulis terkait dengan pengembangan selanjutnya adalah:

1. sebaiknya menggunakan data transaksi perusahaan yang mempunyai data per transaksinya terdapat variasi produk yang banyak, supaya menghasilkan hasil analisis yang lebih tidak monoton,
2. sebaiknya menggunakan data transaksi dengan jumlah pembelian setiap produk hampir sama, supaya bisa menentukan nilai *minimum support* dan *minimum confidence* yang lebih besar.

## Daftar Pustaka :

- [1] Samuel, David. 2008. *Penerapan Struktur FP-Tree dan Algoritma FP-Growth dalam Optimasi Penentuan Frequent Itemset*. Institut Teknologi Bandung..
- [2] Erwin. 2009. *Analisis Market Basket dengan Algoritma Apriori dan FP-Growth*. Jurnal Generic 26-30.
- [3] Han, Jiawei, dan Kamber, Micheline. 2006. *Data Mining : Concepts and Technique*. Morgan Kauffman : San Francisco.
- [4] Fatihatul F., Setiawan A., Rosadi R.. *Asosiasi Data Mining Menggunakan Algoritma FP-Growth untuk Market Basket Analysis*. Tersedia : [http://www.academia.edu/4758451/ASOSIASI\\_DATA\\_MINING\\_MENGGUNAKAN\\_ALGORITMA\\_FP-GROWTH\\_UNTUK\\_MARKET\\_BASKET\\_ANALYSIS](http://www.academia.edu/4758451/ASOSIASI_DATA_MINING_MENGGUNAKAN_ALGORITMA_FP-GROWTH_UNTUK_MARKET_BASKET_ANALYSIS). [Diakses tanggal 9 Oktober 2013]
- [5] Larose, Daniel T. 2005. *Discovering Knowledge in Data : an Introduction to Data Mining*. Canada : John Wiley & Sons, Inc.
- [6] Hermawati, Fajar Astuti. 2013. *Data Mining*. CV Andi Offset : Yogyakarta.
- [7] Raharjo B., Heryanto I., Haryono A.. 2010. *Mudah Belajar Java – edisi Revisi*. Bandung : Informatika Bandung.
- [8] Ramadhan, Arief dan Saputra, Hendra. 2005. Tersedia: [http://elib.unikom.ac.id/files/disk1/444/jbptunikompp-gdl-nesyasitia-22163-2-unikom\\_n-i.pdf](http://elib.unikom.ac.id/files/disk1/444/jbptunikompp-gdl-nesyasitia-22163-2-unikom_n-i.pdf). [Diakses tanggal 9 Oktober 2013].
- [9] Soliha, Euis. 2008. *Analisis Industri Ritel di Indonesia*. Jurnal Bisnis dan Ekonomi (JBE) **15:2** 128-142.