

PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA FP-TREE DAN FP-GROWTH PADA DATA TRANSAKSI PENJUALAN OBAT

Yuyun Dwi Lestari

Program Studi Teknik Informatika, Sekolah Tinggi Teknik Harapan

Jl. H. M. Jhoni No. 70 C Medan

yuyun.dl@gmail.com

Abstrak

Penelitian ini membahas tentang bagaimana sistem kerja transaksi penjualan obat menggunakan Algoritma FP-Growth. Data transaksi penjualan obat ini digunakan untuk menemukan produk yang dibeli secara bersamaan. Penggunaan Algoritma FP-Growth untuk menemukan kombinasi pola barang. Penggunaan FP-Tree yang digunakan bersamaan dengan algoritma FP-Growth untuk menentukan frequent itemset dari sebuah database. Metode Association Rule digunakan dalam pencarian pola keterikatan produk untuk strategi penjualan dalam kebijakan pengambilan keputusan. Sehingga dapat diketahui obat yang sering dibeli oleh konsumen, berdasarkan rule-rule yang dihasilkan dari data-data yang terdapat di dalam database. Pengujian dilakukan dengan menggunakan Aplikasi Rapidminer 5. Hasil yang didapat dari pengujian tersebut adalah semakin kecil nilai support maka jumlah itemset yang dihasilkan akan semakin banyak dan jumlah rules yang dibentuk semakin banyak pula.

Kata Kunci : Obat, FP-Growth, FP-Tree, rule, support.

Abstract

This study discusses how the systems work using the drug sales transactions FP-Growth Algorithm . Drug sales transaction data is used to find a product that is purchased simultaneously . The use of FP -Growth algorithm to find the combination pattern of goods . The use of FP-Tree is used in conjunction with FP-Growth algorithm to determine frequent itemset from a database . Association Rule methods used in the search pattern for the product attachment sales strategy in the policy decision-making. So that it can be seen that the drug is often purchased by consumers , based on rules which generated from the data contained in the database. Testing is done by using the application RapidMiner 5. The results obtained from these tests is the smaller the value of support , the number of itemsets generated will be more and more and the number of rules which formed more and more also .

Keyword : Drug, FP-Growth, FP-Tree, rule, support

1. Pendahuluan

Obat merupakan kebutuhan terhadap seseorang yang menderita suatu penyakit. Obat dapat dibeli di apotek, rumah sakit atau tempat penyedia obat lainnya. Setiap hari akan terjadi transaksi penjualan dan pembelian obat di tempat tersebut. Oleh karena itu setiap apotek atau rumah sakit harus memiliki sistem pengolahan data agar data transaksi tersebut dapat digunakan dalam membuat laporan. Laporan dari data transaksi itu akan menghasilkan informasi yang berguna seperti obat apa saja yang paling sering dibeli atau obat yang paling banyak terjual.

Di apotek banyak terjadi transaksi penjualan dan pembelian obat. Data transaksi akan terus bertambah setiap harinya dan di apotek tersebut data transaksi penjualan hanya disimpan sebagai arsip atau pembukuan dan tidak diketahui apa manfaat dari data-data tersebut untuk selanjutnya. Sistem pengolahan data yang ada di apotek belum berjalan dengan baik karena data-data penjualan yang ada dalam pembukuan hanya dibiarkan menumpuk tanpa diketahui untuk apa selanjutnya data-data tersebut.

Oleh karena itu apotek tersebut memerlukan sistem untuk mengolah data yang dapat menghasilkan data penjualan obat yang paling sering dibeli atau paling banyak terjual sehingga dari hasil tersebut dapat menjadi acuan untuk menambah stok obat yang habis dan mengurangi beberapa obat yang jarang dibeli oleh konsumen.

Tujuan penelitian ini adalah :

1. Menggali informasi dari data penjualan obat sehingga manajer mendapatkan informasi lebih yang dapat digunakan untuk pengambilan keputusan atau kebijakan strategis perusahaan.
2. Memahami konsep *FP-Growth* untuk transaksi penjualan obat.
3. Menampilkan hubungan asosiasi *item* obat yang sering dibeli oleh konsumen untuk mengetahui *item-item* yang saling berkaitan.
4. Mengimplementasikan analisa data menggunakan algoritma *Fp-Growth* untuk menghasilkan penerapan dari *Data Mining* dalam

mengoptimasi penjualan obat menggunakan *software* Rapidminer 5.

Data Mining adalah proses mencari pola atau informasi menarik dalam data terpilih menggunakan teknik atau metode tertentu. Salah satu metode yang seringkali digunakan dalam *Data Mining* adalah metode *Association Rule*.

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. [5]

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 [4]:

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

Association rule merupakan suatu proses pada data mining untuk menentukan semua aturan asosiatif yang memenuhi syarat minimum untuk *support* (*minsup*) dan *confidance* (*minconf*) pada sebuah database. Kedua syarat tersebut akan digunakan untuk *interesting association rules* dengan dibandingkan dengan batasan yang telah ditentukan, yaitu *minsup* dan *minconf*. [3]

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*. [3]

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: [1]

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

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

$$\text{Support (A, B)} = \frac{\text{Jumlah Transaksi yang mengandung A dan B}}{\text{Total Transaksi}}$$

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

$$\text{Confidence (A} \rightarrow \text{B)} = \frac{\text{Jumlah Transaksi yang mengandung A dan B}}{\text{Jumlah Transaksi yang mengandung A}}$$

2. Metode Penelitian

Metode dalam penelitian ini adalah sebagai berikut:

1. Identifikasi Masalah
Permasalahan yang ditemukan penulis, dideskripsikan dengan jelas sehingga akan terlihat permasalahan yang akan dibahas.
2. Analisis Masalah
Langkah analisis masalah adalah langkah untuk memahami yang telah ditentukan ruang lingkup atau batasannya. Dengan menganalisis masalah yang telah ditentukan tersebut, maka diharapkan masalah tersebut dapat dipahami dengan baik.
3. Menentukan Tujuan
Tujuan dari penelitian ini untuk menentukan hasil analisa data penjualan menggunakan algoritma *fp-tree* dan *fpgrowth*.
4. Mempelajari Literatur
Untuk mencapai tujuan yang akan ditentukan, maka perlu dipelajari beberapa literatur-literatur yang sesuai dengan penelitian yang akan dilakukan. Kemudian literatur-literatur yang dipelajari tersebut diseleksi untuk dapat ditentukan literatur mana yang akan digunakan dalam penelitian. Sumber literatur didapat dari buku-buku dan jurnal-jurnal yang ada hubungannya dengan penelitian ini maupun referensi yang lain.
5. Mengumpulkan Data
Pengumpulan data dan informasi pada tahap ini dilakukan untuk mengetahui, mendapatkan data dan informasi yang nantinya akan mendukung penelitian ini, dalam pengumpulan data, terdapat beberapa metode yang digunakan yaitu
 - a. Penelitian Lapangan (*field research*)
Penelitian dilakukan pada apotek yaitu melakukan wawancara dengan pihak apotek untuk meneliti data-data transaksi penjualan obat yang terkait dengan permasalahan ini.
 - b. Penelitian Perpustakaan (*library research*)
Penelitian ini dilakukan untuk mengumpulkan informasi tentang literatur dan pedoman dalam penentuan hasil analisa data transaksi penjualan obat.
 - c. Penelitian Laboratorium (*laboratory research*).
Penelitian laboratorium ini dimaksudkan untuk pengujian data penentuan hasil analisa data penjualan yang telah didapatkan dengan

menggunakan algoritma *fp-tree* dan *fp-growth*.

6. Analisa Data dengan Algoritma *fp-tree* dan *fp-growth*

Pada tahap ini akan dilakukan dua analisa yaitu analisa terhadap algoritma *fp-tree* dan *fp-growth* pada data transaksi penjualan untuk menentukan alternatif yang terbaik dari sekumpulan alternatif yang ada melalui suatu proses seperti yang terjadi pada proses yang terstruktur dan analisa bagaimana algoritma *fp-tree* dan *fp-growth* untuk mendapatkan hasil sebagai tujuan yang akan dicapai oleh peneliti yang kemudian dapat dijadikan pihak apotek sebagai pengetahuan dalam meningkatkan penjualan obat kepada konsumen.

7. Merancang Skema

Pada bagian ini akan menampilkan bagaimana proses analisa *data mining* dengan algoritma *fp-tree* dan *fp-growth* dirancang berdasarkan data yang telah terkumpul. Dan bagaimana mengembangkan proses analisa *data mining* dengan *association rule* dan algoritma *fp-tree* dan *fp-growth* untuk mendapatkan hubungan keterikatan antar tiap barang yang ada. Perancangan ini dilakukan pada data transaksi penjualan untuk mengetahui obat yang paling banyak dibeli oleh konsumen.

8. Implementasi

Pada penelitian ini penulis mengimplementasikan berdasarkan hasil analisa data menggunakan algoritma *fp-tree* dan *fp-growth* dengan menggunakan alat bantu komputer dengan *operating system windows* dan menggunakan *software* Rapidminer 5.

3. Hasil dan Pembahasan

3.1 Pengolahan Data

Pengolahan data dilakukan menggunakan aplikasi database yaitu Microsoft Excel 2007. Data real transaksi penjualan dapat dilihat pada gambar berikut.

ID Transaksi	ItemSet
1	Ditiazem HCL, Ketoconazole, Fungoral, Alopurinol, Piroksikam, Asam Mefenamat
2	Ditiazem HCL, Ketoconazole, Batugin, Piroksikam, Klotaren, Asiklovir
3	Ketoconazole, Fungoral, Alopurinol, Piroksikam, Asam Mefenamat, Rifampicin
4	Ditiazem HCL, Ketoconazole, Fungoral, Alopurinol, Batugin, Klotaren
5	Ditiazem HCL, Ketoconazole, Batugin, Rifampicin, Asiklovir
6	Ditiazem HCL, Fungoral, Alopurinol, Batugin, Piroksikam, Asam Mefenamat

Gambar 1 Data Real Transaksi Penjualan Obat

Data real transaksi penjualan kemudian disusun ke dalam bentuk tabular data, maka data real transaksi penjualan obat dikonversikan ke dalam bentuk 1 dan 0 atau bentuk biner. Dimana 1 adalah jika obat dibeli dan 0 jika obat tidak dibeli. Hasil proses konversi

data transaksi penjualan dalam bentuk tabular data dapat dilihat seperti pada gambar berikut.

	A	B	C	D	E	F	G	H	I	J	K	L
ID	Asam Mefenamat	Ditiazem HCL	Ketozonazole	Fungoral	Alopurinol	Batugin	Asam Mefenamat	Ketozonazole	Rifampicin	Asiklovir	Piroksikam	
1	1	1	1	1	1	1	1	1	1	1	1	
2	1	1	1	1	1	1	1	1	1	1	1	
3	1	1	1	1	1	1	1	1	1	1	1	
4	1	1	1	1	1	1	1	1	1	1	1	
5	1	1	1	1	1	1	1	1	1	1	1	
6	1	1	1	1	1	1	1	1	1	1	1	

Gambar 2 Tabular Data Transaksi Penjualan Obat

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 1 Data Transaksi Awal

No	ItemSet
1	Ditiazem HCL, Ketoconazole, Piroksikam, Fungoral, Alopurinol, Asam Mefenamat
2	Ketoconazole, Ditiazem HCL, Piroksikam, Klotaren, Batugin, Asiklovir
3	Asam Mefenamat, Ketoconazole, Piroksikam, Rifampicin, Alopurinol, Fungoral
4	Ditiazem HCL, Ketoconazole, Alopurinol, Batugin, Klotaren, Fungoral
5	Ketoconazole, Ditiazem HCL, Amoksilim, Rifampicin, Batugin, Asiklovir
6	Ditiazem HCL, Fugoral, Batugin, Piroksikam, Asam Mefenamat, Alopurinol

Kemudian menentukan frekuensi setiap *item* dari keseluruhan transaksi.

Tabel 2 Nama dan Frekuensi Item dari Data Transaksi Awal

No	Nama Obat	Jumlah
1	Amoksilin	1
2	Ditiazem HCL	5
3	Klotaren	2
4	Fungoral	4
5	Alopurinol	4
6	Batugin	4
7	Asam Mefenamat	3
8	Ketoconazole	5
9	Rifampicin	2
10	Asiklovir	2
11	Piroksikam	4

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* $\zeta = 2$, maka hasilnya adalah:

Tabel 3 Nama dan Frekuensi Item Setelah Proses Filter

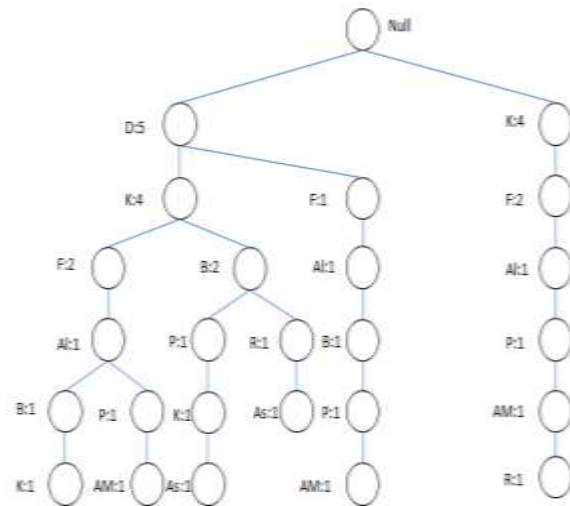
No	Nama Obat	Jumlah
1	Ditiazem HCL	5
2	Ketoconazole	5
3	Fungoral	4
4	Alopurinol	4
5	Batugin	4
6	Piroksikam	4
7	Asam Mefenamat	3
8	Klotaren	2
9	Rifampicin	2
10	Asiklovir	2

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 4 Data Transaksi Setelah Proses Filter

No	ItemSet
1	Ditiazem HCL, Ketoconazole, Fungoral, Alopurinol, Piroksikam, Asam Mefenamat
2	Ditiazem HCL, Ketoconazole, Batugin, Piroksikam, Klotaren, Asiklovir
3	Ketoconazole, Fungoral, Alopurinol, Piroksikam, Asam Mefenamat, Rifampicin
4	Ditiazem HCL, Ketoconazole, Fungoral, Alopurinol, Batugin, Klotaren
5	Ditiazem HCL, Ketoconazole, Batugin, Rifampicin, Asiklovir
6	Ditiazem HCL, Fungoral, Alopurinol, Batugin, Piroksikakm, Asam Mefenamat

Pembangunan *Tree* dari seluruh transaksi tersebut adalah



Gambar 3 Tree dari semua 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. [6]

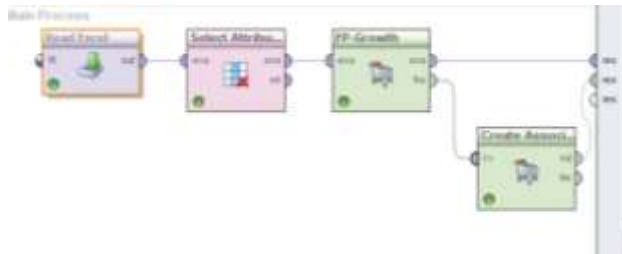
- Tahap Pembangkitan *Conditional Pattern Base*
Conditional Pattern Base merupakan subdata yang berisi *prefix path* (lintasan awal) dan *suffix pattern* (pola akhiran). Pembangkitan *conditional pattern base* didapatkan melalui FP-Tree yang telah dibangun sebelumnya.
- Tahap Pembangkitan *Conditional 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 atau sama dengan minimum *support count* akan dibangun dengan *conditional FP-Tree*.
- 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* (proses memanggil dirinya sendiri)

Untuk pengujian dari data transaksi penjualan yang telah dihasilkan berupa pola hubungan kombinasi antar *items* dan *association rules* sesuai dengan Algoritma FP-Growth, maka digunakan aplikasi Rapidminer untuk menguji analisa pembelian obat yang dilakukan oleh konsumen.

Adapun langkah-langkah pembentukan model mining menggunakan algoritma *FP-Growth* pada *Software Rapidminer* antara lain :

1. Memilih file format xls yang akan dimining.
2. algoritma yang digunakan *algoritma FP-Growth*.
3. Mengatur minimal *support*, *confidence* dan *rules* yang dihasilkan.

Evaluasi data mining akan dilakukan dengan menggunakan dataset pengujian yang berisikan 6 transaksi penjualan obat yang akan dilakukan perhitungan menggunakan algoritma *FP-growth*. Langkah kedua dataset akan dijalankan melalui model mining yang dibuat pada RapidMiner 5. Berikut adalah poses koneksi data menggunakan algoritma *FP-Growth*.



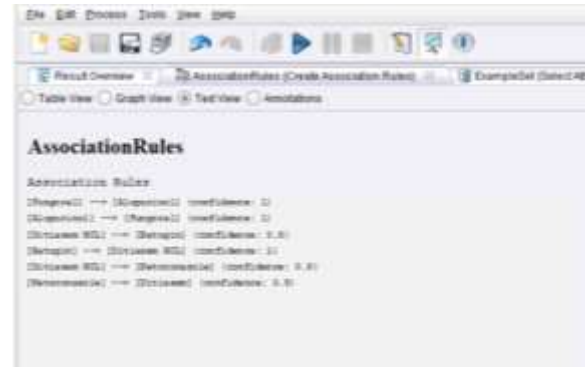
Gambar 4 Proses koneksi data menggunakan *FP-Growth*

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* = 0,6 dan *minimum confidence* = 1, maka hasilnya adalah:

No.	Premises	Conclusion	Support	Confidence	Lift
1	Fungoral	Alopurinol	0.667	1	1.5
2	Alopurinol	Fungoral	0.667	1	1.5
3	Ditiazem HCL	Batugin	0.667	0.8	1.2
4	Batugin	Ditiazem HCL	0.667	1	1.2
5	Ditiazem HCL	Ketoconazole	0.667	0.8	0.96
6	Ketoconazole	Ditiazem HCL	0.667	0.8	0.96

Gambar 5 Hasil *Frequent Item Set*



Gambar 6 Rules yang Dihasilkan

4. Kesimpulan

Berdasarkan analisa dan hasil pengujian maka dapat diperoleh kesimpulan sebagai berikut:

1. *Data Mining* dengan Algoritma *FP-Growth* dapat diimplementasikan dengan menggunakan *database* penjualan obat karena dapat menemukan pola kombinasi *itemsets*. Sehingga informasi tersebut dapat membantu mengembangkan strategi penjualan terhadap konsumen.
2. Metode *Association Rule* dengan menggunakan Algoritma *FP-Growth* dengan parameter *support* dan *confidence* dapat memperoleh korelasi barang pembelian untuk lebih meningkatkan penjualan
3. Setelah dilakukan pengimplementasian Algoritma *FP-Growth* pada Rapidminer, maka hasil pengolahan data penjualan obat paling banyak terjual pada apotek adalah Fungoral, Alopurinol, Ditiazem HCL, Batugin dan Ketoconazole.

References

- [1] Dyah Pramesthi Larasati, Muhammad Nasrun, Umar Ali Ahmad. 2014. Analisis Dan Implementasi Algoritma *FP-Growth* pada Aplikasi Smart Untuk Menentukan Market Basket Analysis Pada Usaha Retail (Studi Kasus : PT. X).
- [2] Erwin. 2009. Analisis Market Basket dengan Algoritma Apriori dan *FP-Growth*. Jurnal Generic 26-30.
- [3] Fatihatul F., Setiawan A., Rosadi R.. *Asosiasi Data Mining Menggunakan Algoritma FP-Growth untuk Market Basket Analysis*. Tersedia : http://www.academia.edu/4758451/ASOSIASI_DAT_A_MINING_MENGGUNAKAN_ALGORITMA_FP

GROWTH_UNTUK_MARKET_BASKET_ANALY
SIS. [Diakses tanggal 02 Agustus 2015]

[4] Han Jiawei, and M. Kamber. 2006. Data Mining: Concepts and Techniques, Morgan Kaufmann, USA.

[5] Samuel, David. 2008. *Penerapan Struktur FP-Tree dan Algoritma FP-Growth dalam Optimasi Penentuan Frequent Itemset*. Institut Teknologi Bandung

[6] Anfatul Maulida, Tia. 2014. Analisa Data Mining Menggunakan Algoritma *Frequent Pattern Growth* pada Data Transaksi Penjualan Restaurant Joglo Kampoeng Doeloe Semarang.

John