

Implementasi Data Mining untuk Memprediksi Itemset Promosi Penjualan Pada CV. Sumber Segar Utama

Nova Mayasari

Fakultas Sains dan Teknologi
Universitas Pembangunan Panca Budi
Medan, Sumatera Utara, Indonesia
maya7886@pancabudi.ac.id

Debi Yandra Niska

Fakultas Sains dan Teknologi
Universitas Pembangunan Panca Budi
Medan, Sumatera Utara, Indonesia
debiyandraniska@gmail.com

Erni Ronauli Samosir

Fakultas Sains dan Teknologi
Universitas Pembangunan Panca Budi
Medan, Sumatera Utara, Indonesia
ernironaulisamosir@gmail.com

Abstrak—Data mining yang diterapkan pada CV. Sumber Segar Utama untuk memprediksi itemset promosi. Hal tersebut bertujuan untuk meningkatkan penjualan dan pengembangan strategi pemasaran. Algoritma yang digunakan pada data mining ini adalah Algoritma Frequent Pattern Growth (FP-Growth). Data yang diolah pada Data Mining ini berupa data sampel dari data transaksi penjualan buah-buahan seperti apel, jeruk, mangga, dan lain-lain dalam satu bulan, yaitu pada bulan Maret 2017. Pengolahan data transaksi penjualan tersebut bertujuan untuk melihat buah mana yang lebih diminati oleh pembeli, dan buah mana yang kurang diminati. Dengan mengetahui hal tersebut, pihak manajer pemasaran dapat meningkatkan promosi terhadap buah yang kurang diminati tersebut, sehingga transaksi penjualan pada CV. Sumber Segar Utama bisa meningkat.

Kata kunci : Data Mining, Itemset, FP-Growth.

I. PENDAHULUAN

Data transaksi penjualan pada CV. Sumber Segar Utama setiap hari semakin bertambah. Data tersebut dapat diolah menjadi informasi yang berguna bagi perusahaan. Salah satu tujuannya adalah untuk memprediksi itemset promosi guna peningkatan penjualan dan pengembangan strategi pemasaran.

Itemset promosi merupakan suatu strategi pada ilmu ekonomi pemasaran untuk memasarkan suatu produk atau bahkan jasa untuk meningkatkan kuantitas penjualan suatu perusahaan. Informasi pemasaran yang dihasilkan akan membantu manajer dalam memutuskan suatu kegiatan promosi guna pengenalan suatu item kepada masyarakat luas.

Untuk memperoleh pengetahuan dari informasi pemasaran tersebut diperlukan suatu teknik yang disebut Data Mining. Data mining merupakan proses pencarian pola yang menarik dan tersembunyi (*hidden pattern*) dari suatu kumpulan data yang berukuran besar yang tersimpan dalam basis data, seperti data *were house* dan tempat penyimpanan data lainnya.

Data Mining dapat menyatukan teknik pembelajaran mesin, pengolahan pola, statistik, *database*, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari *database* yang besar. Hubungan yang dicari dalam data mining dapat berupa hubungan antara dua atau lebih dalam satu dimensi. Misalnya dalam dimensi produk dapat melihat keterkaitan pembelian suatu produk dengan

produk yang lain. Selain itu, hubungan juga dapat dilihat antara dua atau lebih atribut dan dua atau lebih objek.

Salah satu faktor penting dalam data mining adalah adanya aturan untuk menemukan suatu pola frekuensi tinggi antar himpunan itemset yang disebut dengan Aturan Asosiasi (*Association Rule*). Algoritma yang digunakan pada data mining ini adalah Algoritma *FP-Growth* (*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 (Ikhwan A., Nofriansyah D., and Sriani 2015). Pada algoritma Apriori frequent itemsets dicari dengan menggunakan Association Rule dan untuk mendapatkan hal tersebut maka perlu diketahui nilai *support* dan *confidence* (Santoso H., Hariyadi I.P. and Prayitno. 2016)

II. TINJAUAN PUSTAKA

A. Data Mining

Data Mining adalah proses penggalian informasi pada data dalam jumlah yang besar. Data Mining merupakan serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data yang menghasilkan suatu pengetahuan yang selama ini tidak diketahui secara manual.

Pada Data Mining terdapat 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 (Gunadi G and Sensuse I., 2012).

Ada beberapa macam pendekatan yang berbeda yang diklasifikasikan sebagai teknik pencarian informasi/pengetahuan dalam Data Mining. Ada pendekatan kuantitatif, seperti pendekatan probabilistik, logika induktif, pencarian pola, dan analisis pohon keputusan. Pendekatan yang lain meliputi deviasi, analisis kecenderungan, algoritma genetik, jaringan saraf tiruan, dan pendekatan

campuran dua atau lebih dari beberapa pendekatan yang ada.

Pada dasarnya ada enam elemen yang paling esensial dalam teknik pencarian informasi/pengetahuan dalam Data Mining yaitu:

1. Mengerjakan sejumlah besar data
2. Diperlukan efisiensi berkaitan dengan volume data
3. Mengutamakan ketetapan/keakuratan
4. Membutuhkan pemakaian bahasa tingkat tinggi.
5. Menggunakan beberapa bentuk dari pembelajaran otomatis
6. Menghasilkan hal yang menarik

Kemajuan luar biasa yang terus berlanjut dalam bidang data mining didorong oleh beberapa faktor, antara lain:

1. Pertumbuhan yang cepat dalam kumpulan data.
2. Penyimpanan data dalam data warehouse, sehingga seluruhnya perusahaan memiliki akses kedalam database yang andal.
3. Adanya peningkatan akses data melalui navigasi web dan internet.
4. Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.
5. Perkembangan teknologi perangkat lunak untuk data mining (ketersediaan teknologi).
6. Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.

Pada umumnya data mining terdiri dari komponen-komponen berikut ini:

1. *Database, data warehouse*, dan media penyimpanan informasi.
Media pada komponen ini dapat berupa *database, data warehouse, spreadsheet*, atau jenis media penampung lainnya. *Data cleaning* dan *data integration* dapat dilakukan pada data tersebut.
2. *Database* atau *data warehouse server*.
Database atau *data warehouse server* bertanggung jawab untuk menyediakan data yang relevan berdasarkan permintaan dari pengguna data mining.
3. Basis pengetahuan (*knowledge base*)
Merupakan basis pengetahuan yang digunakan sebagai panduan dalam pencarian pola.
4. *Data mining engine*
Yaitu bagian dari *software* yang menjalankan program berdasarkan algoritma yang ada.
5. *Pattern evaluation module*
Yaitu bagian dari *software* yang berfungsi untuk menemukan *pattern* atau pola-pola

yang terdapat dalam *database* yang diolah sehingga nantinya proses data mining dapat menemukan *knowledge* yang sesuai.

6. *Graphical user interface*

Bagian ini merupakan sarana antara *user* dan sistem *data mining* untuk berkomunikasi, dimana *user* dapat berinteraksi dengan sistem melalui *data mining query*, untuk menyediakan informasi yang dapat membantu dalam pencarian *knowledge*. Lebih jauh lagi, bagian ini memungkinkan *user* untuk melakukan *browsing* pada *database* dan *data warehouse*, mengevaluasi *pattern* yang telah dihasilkan, dan menampilkan *pattern* tersebut dengan tampilan yang berbeda-beda.

B. Algoritma *Frequent Pattern Growth (FP-Growth)*

Algoritma *Frequent Pattern Growth* adalah pengembangan dari algoritma Apriori. Algoritma *FP-Growth* merupakan salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah kumpulan data. 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 dari pada algoritma Apriori (Meilani D.B. and Azinar A.W, 2015).

Algoritma *FP-Growth* dibagi menjadi tiga tahapan utama, yaitu:

1. Tahap pembangkitan *Conditional Pattern Base*

Tahap ini merupakan *subdatabase* yang berisi *prefix path* dan *suffix pattern* (pola akhiran).

2. Tahap pembangkitan *Conditional FP-Tree*, Pada tahap ini *Support Count* dari setiap item untuk *Conditional pattern base* dijumlahkan.

3. Tahap pencarian *Frequent Itemset*

Tahap ini merupakan lintasan tunggal (*single path*), kemudian didapatkan *Frequent itemset* dengan melakukan kombinasi item untuk *conditional FP-Tree*.

C. Promosi

Promosi adalah aktivitas yang dilakukan untuk mengkomunikasikan keunggulan produk, serta membujuk pelanggan sasaran untuk membeli dan memanfaatkan produk-produk yang dijual oleh suatu perusahaan. Aktivitas ini merupakan suatu proses

sosial dan manajerial yang individu atau kelompok untuk memperoleh apa yang mereka butuhkan dan inginkan, lewat penciptaan dan pertukaran timbal balik produk dan nilai dengan orang lain.

Promosi dapat berupa iklan yakni segala bentuk penyajian dan promosi secara non pribadi dari ide barang dan pelayanan yang dibayar oleh sponsor tertentu bertujuan menyampaikan berita kepada konsumen melalui berbagai media seperti: surat kabar, majalah, surat langsung dan melalui media lainnya.

III. HASIL DAN PEMBAHASAN

Sebagai tahap awal *data mining* adalah dengan melakukan pengumpulan data penjualan pada CV. Sumber Segar Utama. Data tersebut akan digunakan sebagai sampel dan akan dianalisis dengan menggunakan metode *Frequent pattern growth (Fp-Growth)*. Data tersebut adalah berupa data transaksi dalam satu bulan pada bulan Maret 2017, seperti dapat dilihat pada tabel 1 berikut ini:

Tabel 1. Data Awal Transaksi

No.	Jenis Buah	Inisialisasi	Jumlah Transaksi
1	Anggur	AN	3
2	Apel	AP	7
3	Pisang	PG	10
4	Kiwi	KW	5
5	Mangga	MG	8
6	Buah Naga	BN	7
7	Jeruk	JR	11
8	Lemon	LM	10
9	Pear	PR	14
10	Semangka	SM	3
11	Buah Atep	BA	3
12	Melon	ML	6
13	Pepaya	PY	4
14	Lengkeng	LK	3
15	Kurma	KM	3
16	Salak	SL	2
17	Nenas	NN	3
18	Tomat	TM	4
19	Blueberry	BB	2
20	Blewah	BL	4
21	Alpukat	AK	2

Sumber : CV. Sumber Segar Utama

Data yang ada pada tabel 1 dipilih untuk menentukan item-item yang berfungsi dalam praproses penentuan dan perhitungan *FP-Growth*. Yang dipilih sebagai itemnya adalah jenis buah beserta jumlah transaksi. Dari 21 data buah yang ada, 10 data buah dan jumlah transaksi yang akan dijadikan *sample* atau contoh dalam perhitungan algoritma *FP-Growth* yang dapat dilihat pada tabel 2 berikut ini:

Tabel 2. Data Sampel Proses

No.	Jenis Buah	Inisialisasi	Jumlah Transaksi	Minggu ke-
1	Salak	SL	2	II dan III
2	Anggur	AN	3	I, III dan V
3	Apel	AP	7	I, II, III dan IV
4	Pisang	PG	10	I, II, III, IV dan V
5	Kiwi	KW	5	I, III dan V
6	Mangga	MG	8	I, III, IV dan V
7	Pepaya	PY	4	II, III, IV dan V
8	Melon	ML	6	II, III dan IV
9	Jeruk	JR	11	I, II, III, IV dan V
10	Pear	PR	14	I, II, III, IV dan V

Di bawah ini akan ditampilkan tabel data transaksi sampel dalam kurun waktu satu bulan:

Tabel 3. Data Transaksi Sampel

Minggu ke-	Rate Penjualan Buah									
	1	2	3	4	5	6	7	8	9	10
I	-	AN	AP	PG	KW	MG	-	-	JR	PR
II	SL	-	AP	PG	-	-	PY	ML	JR	PR
III	SL	AN	AP	PG	KW	MG	PY	ML	JR	PR
IV	-	-	AP	PG	-	MG	PY	ML	JR	PR
V	-	AN	-	PG	KW	MG	PY	-	JR	PR

Data Transaksi Sampel tersebut adalah data sampel yang akan diolah dengan menggunakan algoritma *FP-Growth*. Frekuensi kemunculan tiap item dari data transaksi dapat dilihat pada tabel 4 berikut ini:

Tabel 4. Frekuensi Kemunculan Data

Inisialisasi Buah	Frekuensi	Support (Frekuensi/5)	Support (%)
SL	2	2/5	40%
AN	3	3/5	60%
AP	4	4/5	80%
PG	5	5/5	100%
KW	3	3/5	60%
MG	4	4/5	80%
PY	4	4/5	80%
ML	3	3/5	60%
JR	5	5/5	100%
PR	5	5/5	100%

Pada tabel 4 dapat dilihat hasil frekuensi kemunculan item penjualan pada CV. Sumber Segar Utama melalui transaksi setiap minggu. Kemudian dilakukan penentuan *support*. Pada kasus ini diambil nilai *support count* = 75%. Nilai tersebut akan mempengaruhi *item* yang akan dianalisa ke tahap pembuatan *FP-Tree*. Kita dapat mengetahui nilai *support* tertinggi, yaitu nilai *support* yang memenuhi *support account* 75%. Nilai *support* tersebut merupakan nilai yang menjadi kebutuhan perusahaan dalam melihat persentase rata-rata transaksi. Item yang akan digunakan untuk selanjutnya adalah item yang memenuhi frekuensi $\geq 75\%$, yaitu pada tabel 5 berikut ini:

Tabel 5. Frekuensi Kemunculan *support account*

Inisialisasi Buah	Frekuensi	Support (Frekuensi/5)	Support (%)
AP	4	4/5	80%
PG	5	5/5	100%
MG	4	4/5	80%
PY	4	4/5	80%
JR	5	5/5	100%
PR	5	5/5	100%

Data tersebut akan dipindai berdasarkan tabel frekuensi kemunculan data. Sedangkan item lain yang tidak memenuhi frekuensi kemunculan *support account* yakni SL dengan nilai *support* 40%, AN dengan nilai *support* 60%, KW dengan nilai *support* 60% dan ML dengan nilai *support* 60% masing-masing tidak memenuhi nilai minimum *support* 75% sehingga tidak dipindai. Setelah dilakukan proses pemindaian data antara tabel frekuensi kemunculan data dan tabel frekuensi kemunculan *support account*, maka proses data mining dapat dilihat pada tabel 6 berikut ini:

Tabel 6. Tabel Transaksi Pemindaian

Minggu Ke-	Transaksi
I	AP, PG, MG, JR, PR
II	AP, PG, PY, JR, PR
III	AP, PG, MG, PY, JR, PR
IV	AP, PG, MG, PY, JR, PR
V	PG, MG, PY, JR, PR

1. Penerapan *FP-Growth*

Setelah dilakukan pembangunan *FP-Tree* dari sekumpulan transaksi yang telah dilakukan, pada tahapan diatas maka dilanjutkan dengan tahap *FP-Growth* untuk mencari *frequent itemset* yang memenuhi syarat. Langkah *FP-Growth* ini sebanyak 3 langkah utama yaitu *Conditional Pattern Base*, pembangkitan *Conditional FP-Tree*, dan pencarian *Frequent Itemset*.

a. Pembangkitan *Conditional Pattern Base*

Pada tahap pembangkitan *conditional pattern base* dilakukan dengan cara *up-pohon* atau ditentukan dengan cabang pohon dengan *support count* terkecil, namun dalam *sample* data ini nilai *support* terkecil sama semua sehingga penulis menentukan secara random, yakni PY dengan item pepaya.

b. Pembangkitan *Conditional FP-Tree*

Pada tahap ini juga dilakukan dengan melihat *FP-Tree* yang sudah dibentuk sebelumnya. *Conditional FP-Tree* dimaksudkan untuk mencari *frequent itemset* yang berakhiran *item* tertentu atau mengandung *suffix* yang sama. Langkah pertama yang dilakukan adalah membangkitkan *conditional FP-Tree* untuk lintasan yang mengandung *suffix* PR, dimana semua simpul yang berakhiran PR dibuang sehingga selanjutnya adalah membangun *conditional FP-Tree* setiap lintasan yang mengandung *suffix* JR.

c. Pencarian *Frequent Itemset*

Setelah melakukan tahapan pembentukan *Frequent Itemset* untuk akhiran (*suffix*), maka hasil dapat dilihat pada tabel 7 berikut:

Tabel 7. Tabel Hasil *Frequent Itemset*

Suffix	Frequent Itemset
PR	{JR}, {JR,PY}, {JR,MG}, {JR,PY,MG}, {JR,PY,PG}, {JR,MG,PG}, {JR,PY,MG,PG}, {JR,PY,PG,AP}, {JR,MG,PG,AP}, {JR,PY,MG,PG,AP}
JR	{MG}, {PY}, {MG,PG}, {PY,MG}, {PY,PG}, {MG,PG,AP}, {PY,MG,PG}, {PY,PG,AP}, {PY,MG,PG,AP}
PY	{MG}, {PG}, {MG,PG}, {PG,AP}, {MG,PG,AP}
MG	{PG}, {PG,AP}
PG	{AP}
AP	Null

Pencarian *Association Rules* dilakukan melalui dua tahap yaitu pencarian *frequent itemset* dan penyusunan *rules*. Penting tidaknya suatu *Association Rules* dapat diketahui dengan dua parameter, yaitu *support* (nilai penunjang) dengan nilai min *support* 75% dan *confidence* (nilai kepastian) dengan minimum *confidence* 75%. *Support* adalah ukuran menunjukkan tingkat dominasi *itemset* keseluruhan transaksi.

Setelah didapat *frequent itemset*, selanjutnya membuat *rules* dengan cara menghitung *confidence* dari tiap kombinasi *rule*. Sebelum membuat *rules* maka dilakukan pemilihan nilai kombinasi keseluruhan dari *frequent itemset* dimana nilai A mengandung nilai B. Baik yang memenuhi nilai *support* maupun yang tidak memenuhi nilai *support*.

Hanya kombinasi yang lebih besar atau sama dengan *minimum confidence* yang akan diambil atau *strong association rule* saja. Rumus untuk menghitung *Support* dan *confidence* menggunakan rumus:

$$\text{Support (A,B)} = \frac{\text{Jumlah Transaksi Mengandung A dan B}}{\text{Total Transaksi}} \times 100\%$$

$$\text{Confidence (A} \rightarrow \text{B)} = \frac{\text{Jumlah Transaksi yang Mengandung A dan B}}{\text{Jumlah Total Transaksi yang Mengandung A}} \times 100\%$$

Dari perhitungan nilai *Support* setiap *association rule* yang dilalui dan yang memenuhi nilai *Support* $\geq 75\%$, hanya kombinasi yang lebih kecil atau sama dengan minimum *confidence* atau *low Association Rule* yang akan diambil sebagai *itemset* promosi penjualan. Jika minimum *confidence* adalah 50%, maka *frequent* yang diperoleh dapat dilihat pada tabel 8 sebagai berikut:

Tabel 8. Tabel *Low Association Rule*

Jika Membeli	Maka akan membeli	Confidence
--------------	-------------------	------------

JR	PY,MG,PG,AP	40%
JR,PY	MG,PG,AP	50%
JR,MG	PY,PG,AP	50%
JR,PG	PY,MG,AP	40%
JR,AP	PY,MG,PG	50%
JR,PY,PG	MG,AP	50%
JR,MG,PG	PY,AP	50%
JR,PG,AP	PY,MG	50%
PY	MG,PG,AP	50%
PY,PG	MG,AP	50%

Dari tabel tersebut dapat kita simpulkan bahwa jika pelanggan membeli Jeruk (JR), maka probabilitas pelanggan akan membeli Pepaya (PY), Mangga (MG), Pisang (PG), Apel (AP) berdasarkan pola pembelian pelanggan sebelumnya adalah 40% atau sangat lemah. Oleh karena itu pihak manajemen perlu melakukan promosi terhadap pelanggan untuk membeli Pepaya, Mangga, Pisang, dan Apel dan begitu selanjutnya. Dengan tujuan agar proses penjualan dapat mengoptimalkan keuntungan.

IV. PENUTUP

A. Kesimpulan

Berdasarkan hasil analisis dan perhitungan terhadap implementasi algoritma *FP-Growth* untuk memprediksi *itemset* promosi penjualan pada CV. Sumber Segar Utama, maka dapat disimpulkan sebagai berikut:

1. Data Mining dapat diimplementasikan untuk memprediksi *itemset* promosi penjualan pada CV. Sumber Segar Utama karena dapat menemukan kecenderungan pola kombinasi *itemsets* sehingga dapat dijadikan sebagai informasi yang sangat berharga dalam proses penjualan.
2. Penerapan Algoritma *FP-Growth* pada teknik Data Mining sangat efisien dan dapat mempercepat proses pembentukan pola kombinasi *itemset* hasil penjualan buah pada CV. Sumber Segar Utama.

B. Saran

Adapun saran-saran yang dapat diberikan untuk pengembangan selanjutnya adalah sebagai berikut:

1. Analisa yang dihasilkan pada penelitian ini masih merupakan analisa mendasar dan belum sempurna, sehingga perlu dilakukan pengembangan yang lebih spesifik oleh penelitian selanjutnya.
2. Untuk penelitian selanjutnya, data yang digunakan dapat disesuaikan dengan kebutuhan peneliti.
3. Data yang diolah merupakan data-data dalam jumlah yang besar, sehingga perlu dilakukan pengolahan data yang lebih baik dan teliti, agar tidak terjadi kesalahan dan kekeliruan dalam proses pengolahan data.

DAFTAR PUSTAKA

- [1] Andrian Ody. 2016. Aplikasi Data Mining Metode K-Nearest Neighbor untuk Pemilihan laptop, STMIK AKAKOM, Yogyakarta
- [2] Fitriyah, N., 2016, Penggunaan Association Rule Mining Algoritma CT-Pro dalam Proses Penentuan Pola Tindak Kejahatan Perdagangan Manusia (Strudi Kasus: Data Kepolisian Daerah Jawa Barat), Tesis, Jurusan Ilmu Komputer FMIPA UGM, Yogyakarta.
- [3] Gunadi G dan Sensuse I. 2012. Penerapan Metode Data Mining Market Base Analysis terhadap Data Penjualan Produk Buku dengan Menggunakan Algoritma Apriori dan Frequent Pattern-Growth (FP-Growth): Studi Kasus Percetakan PT. Gramedia. Volume 4 No. 1, Halaman 121.
- [4] Ikhwan, A., Nofriansyah, D. dan Sriani, 2015, Penerapan Data Mining dengan Algoritma Fp-Growth untuk Mendukung Strategi Promosi Pendidikan (Studi Kasus Kampus STMIK Triguna Dharma), Jurnal Ilmiah SAINTIKOM, 14, 211-226.
- [5] Meilani B.D. dan Azinar A.W., 2015. Penentuan Pola Yang Sering Muncul Untuk Penerima Kartu Jaminan Kesehatan Masyarakat (Jamkesmas) Menggunakan Metode FP-Growth. Surabaya : Sistem Informasi Fakultas Teknologi Informasi.
- [6] Santosa, B., 2007, Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis, Cetakan Pertama, Graha Ilmu, Yogyakarta
- [7] Sukamsi, R. S. 2016. Aplikasi Data Mining Metode K-MEANS untuk Mengklasifikasi-kan Data Mahasiswa Akakom, STMIK AKAKOM, Yogyakarta.