

PENERAPAN DATA MINING MENGGUNAKAN METODE ASSOCIATION RULE DENGAN ALGORITMA APRIORI UNTUK ANALISA POLA PENJUALAN BARANG

Yori Apridonal M^{1*}, Wirdah Choiriah², Akmal¹

¹Sistem Informasi, STMIK Royal Kisaran

²Sistem Informasi, Universitas Lancang Kuning

*email: yori@royal.ac.id

Abstract: Fantasy Kids is a children's clothing distribution in the Bangkinang area, Kampar Regency, Riau. In its operations, distros sell their products to the general public, including the sale of children's shirts, children's shirts, jackets or children's sweaters which are usually sold in other distros. These distributions carry out product updates at certain events. Data Mining is the development or discovery of new information by looking for certain patterns or rules of a large amount of data expected to overcome these conditions. The method that will be used in the construction of this application is the Association Rule method with the Apriori Algorithm. Association Rule method is a procedure to find relationships between items in a specified data set. In determining a Association Rule, there is a measure of trust obtained from the results of processing data with certain calculations. Apriori Algorithm is an alternative Algorithm that can be used to determine the frequent itemset in a data set.

Keywords : Data Mining, Algoritma, *Apriori*, *Association Rule*, Sales, Distro

Abstrak: Fantasy Kids merupakan sebuah distro baju anak-anak di kawasan Bangkinang, Kabupaten Kampar, Riau. Dalam operasionalnya, distro menjual produknya kepada masyarakat umum meliputi penjualan kaos anak, kemeja anak, bag, jaket atau sweater anak yang biasa dijual di distro-distro lainnya. Distro ini melakukan pembaruan produk pada event tertentu. Data Mining merupakan pengembangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data dalam jumlah besar diharapkan dapat mengatasi kondisi tersebut. Metode yang akan digunakan dalam pembangunan aplikasi ini adalah metode *Association Rule* dengan *Algoritma Apriori*. Metode *Association Rule* adalah suatu prosedur untuk mencari hubungan antara item dalam suatu kumpulan data yang ditentukan. Dalam menentukan suatu *Association Rule*, terdapat suatu ukuran kepercayaan yang di dapatkan dari hasil pengolahan data dengan perhitungan tertentu. *Algoritma Apriori* merupakan salah satu alternatif Algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam suatu kumpulan data.

Kata kunci: Data Mining, Algoritma, *Apriori*, *Association Rule*, Penjualan, Distro

PENDAHULUAN

Perkembangan teknologi komputer

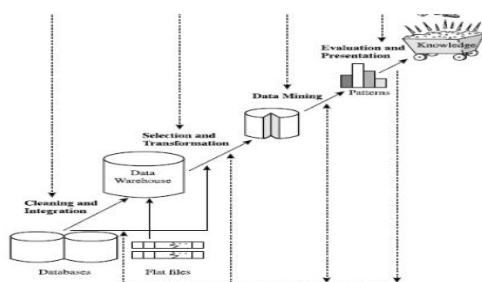
saat ini sudah semakin pesat, yang mengakibatkan hampir seluruh aktivitas kehidupan manusia menggunakan

bantuan komputer, hal ini berdampak pada peningkatan data komputer secara signifikan.

Dalam operasionalnya, distro menjual produknya kepada masyarakat umum meliputi penjualan kaos anak, kemeja anak, bag, jaket atau sweater anak yang biasa dijual di distro-distro lainnya. Distro ini melakukan pembaruan produk pada event tertentu.

Data Mining

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakut dari berbagai basis data besar [1]. Data mining, sering juga disebut Knowledge Discovery in Database atau disingkat menjadi KDD, adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar [2]. Gambar tahapan pembuatan aplikasi data mining ditunjukkan pada gambar 2.1.



Gambar 1. Tahapan KDD

Tahapan proses KDD yaitu:

1. Pembersihan data (*data cleaning*)
Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh dari basis data suatu perusahaan,

memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki.

2. Integrasi data (*data integration*)
Integrasi data merupakan penggabungan data dari berbagai basis data ke dalam suatu basis data baru. Data yang diperlukan untuk data mining tidak hanya berasal dari satu basis data tetapi juga berasal dari beberapa basis data. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan, dan lainnya.
3. Seleksi data (*data selection*)
Data yang ada pada basis data seringkali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk proses analisis yang akan diambil dari basis data. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus analisis keranjang belanja, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan.
4. Transformasi data (*data transformation*)
Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan.
5. Proses mining
Proses mining merupakan proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.
6. Evaluasi pola (*pattern evaluation*)

Evaluasi pola bertujuan untuk menemukan pola-pola menarik ke dalam basis pengetahuan yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai.

Association Rules

Association rules adalah suatu prosedur untuk mencari hubungan antar item suatu dataset yang telah ditentukan [3]. Association rules mining mencari dan menemukan hubungan antar item yang ada pada suatu dataset. Penerapan data mining dengan aturan asosiasi bertujuan menemukan informasi item-item yang saling berhubungan dalam bentuk aturan/rule [4]. Aturan asosiasi adalah teknik data mining untuk menemukan aturan asosiasi antara suatu kombinasi item. Dalam menentukan suatu aturan asosiasi, terdapat suatu ukuran ketertarikan (interestingness measure) yang didapatkan dari hasil pengolahan data dengan data perhitungan tertentu. Pada umumnya terdapat dua ukuran ketertarikan dalam aturan asosiasi, yaitu :

1. Support adalah probabilitas konsumen membeli beberapa produk secara bersamaan dari jumlah seluruh transaksi. Ukuran ini menentukan apakah suatu item/itemset layak untuk dicari nilai confidence-nya (misal dari keseluruhan transaksi yang ada, seberapa besar tingkat dominasi yang menunjukkan bahwa item X dan Y dibeli bersamaan).
2. Confidence atau tingkat kepercayaan merupakan probabilitas kejadian beberapa produk yang dibeli bersamaan dimana salah satu produk sudah pasti dibeli (misal, seberapa sering item Y dibeli apabila konsumen membeli item X).

Kedua ukuran (support dan confidence) berguna dalam menentukan aturan asosiasi, yaitu untuk dibandingkan dengan batasan (threshold) yang ditentukan oleh pengguna. Batasan tersebut umumnya terdiri atas minimum support sebagai batasan minimum dari nilai support dan minimum confidence sebagai batasan minimum dari nilai confidence. Langkah-langkah dalam pembentukan aturan asosiasi meliputi dua tahap, yaitu :

1. Analisis pola frekuensi tinggi Tahap ini mencari kombinasi item yang memenuhi syarat minimum dari nilai support dalam basis data. Support untuk aturan “ $X \Rightarrow Y$ ” adalah probabilitas atribut atau kumpulan atribut X dan Y yang terjadi bersamaan dalam suatu transaksi. Bentuk persamaan matematika dari nilai support adalah : $\text{Support} (X \Rightarrow Y) = P (X \cap Y)$ Dengan keterangan : $X \Rightarrow Y =$ item yang muncul bersamaan $P (X \cap Y) =$ probabilitas transaksi yang mengandung X dan Y dibagi dengan jumlah transaksi seluruhnya.
2. Pembentukan Aturan Asosiasi Setelah semua pola frekuensi tinggi ditemukan, kemudian dicari aturan asosiasi yang memenuhi syarat minimum untuk confidence dengan menghitung confidence dari aturan if X then Y. Bentuk rumus matematika dari confidence adalah : $\text{Confidence} (X \Rightarrow Y) = P (Y | X)$ Dengan keterangan : $X \Rightarrow Y =$ item yang muncul bersamaan $P (Y | X) =$ probabilitas jumlah transaksi yang mengandung X dan Y dibagi dengan jumlah transaksi yang mengandung X.

Algoritma Apriori

Apriori adalah algoritma yang digunakan dalam melakukan pencarian frequent itemset untuk mendapatkan

aturan asosiasi [5]. Sesuai dengan namanya, algoritma ini menggunakan prior knowledge mengenai frequent itemset properties yang telah diketahui sebelumnya untuk memproses informasi selanjutnya. Apriori menggunakan pendekatan secara iterative yang disebut juga sebagai level-wish search dimana kitemset digunakan untuk mencari (k+1)-itemset. Pertama-tama dicari set dari frequent 1-itemset, set ini dinotasikan sebagai L1. L1 yaitu large itemset pertama yang digunakan untuk menemukan L2, kemudian set dari frequent 2 itemset digunakan untuk menemukan L3, dan seterusnya sampai tidak ada lagi frequent k-itemset yang dapat ditemukan. Large itemset adalah itemset yang sering terjadi atau itemset itemset yang sudah melewati batas minimum support yang telah ditentukan.

METODE

Tahapan yang dilakukan :

- a. Analisa pola frekuensi tinggi
- b. Pembentukan aturan assosiatif
 - i. Seleksi data
 - ii. Menentukan frequent itemset
 - iii. Menentukan nilai *antecedent* dan *concequent*
 - iv. Membuat *rule*
 - v. Menghiung nilai *support* dan *confidence*
 - vi. Perkalian *support & confidence*

HASIL DAN PEMBAHASAN

Contoh data transaksi penjualan distro fantasy kids Bangkinang

Tabel 1. Data Transaksi

Transaksi	Barang Yang Dijual
1	Little M, Gymboree, The GAP

2	The GAP, Cool Kids
3	The GAP, Jumping Beans, Little M
4	Cool Kids, The GAP, Gymboree, Little M
5	Gymboree, Little M
6	Gymboree, The GAP, Jumping Beans
7	Jumping Beans, Little M, Cool Kids

Tabel 2. Seleksi Data

Barang Yang Dijual
Cool Kids
Jumping Beans
The GAP
Little M
Gymboree

Tabel 3. Frequent Itemset

T	Cool Kids	Jumping Beans	The GAP	Little M	Gymboree
1	0	0	1	1	1
2	1	0	1	0	0
3	0	1	1	1	0
4	1	0	1	1	1
5	0	0	0	1	1
6	0	1	1	0	1
7	1	1	0	1	0
Σ	3	3	5	5	4

Frekuent itemset = 3 maka himpunan yang mungkin terbentuk adalah: {Cool Kids, Jumping Beans}, {Cool Kids, The GAP}, {Cool Kids, Little M}, {Cool Kids, Gymboree}, {Jumping Beans, The GAP}, {Jumping Beans, Little M}, {Jumping Beans, Little M}, {The GAP, Little M}, {The GAP, Gymboree}, {Little M, Gymboree}.

Tabel 4. Tabel 2 Itemset

T	Cool Kids	Jumping Beans	f
1	0	0	s
2	1	0	s

3	0	1	s
4	1	0	s
5	0	0	s
6	0	1	s
7	1	1	p
Σ			1

T	Little M	Gymboree	f
1	1	1	p
2	0	0	s
3	1	0	s
4	1	1	p
5	1	1	p
6	0	1	s
7	1	0	s
Σ			3

Dari tabel-tabel 2 unsur di atas, P artinya item-item yang dijual bersamaan, sedangkan S berarti tidak ada item yang dijual bersamaan atau tidak terjadi transaksi. Σ melambangkan jumlah Frekuensi item set. Jumlah frekuensi item set harus lebih besar atau sama dengan jumlah Frekuensi item set ($\Sigma \geq \Phi$). Dari tabel diatas, maka didapat: $F2 = \{\{\text{The GAP, Little M}\}, \{\text{The GAP, Gymboree}\}, \{\text{Little M, Gymboree}\}\}$ Kombinasi dari itemset dalam $F2$, dapat kita gabungkan menjadi calon 3 itemset.

Tabel 5. Tabel 3 Itemset

T	The GAP	Little M	Gymboree	f
1	1	1	1	p
2	1	0	0	s
3	1	1	0	s
4	1	1	1	p
5	0	1	1	s
6	1	0	1	s
7	1	1	0	s
Σ				2

Itemset-itemset yang dapat digabungkan adalah itemset-itemset yang memiliki

kesamaan dalam k-1 item pertama. Untuk $k = 3$ (3 unsur), himpunan yang mungkin terbentuk adalah: $\{\text{The GAP, Little M, Gymboree}\}$

Dari tabel-tabel di atas, didapat $F3 = \{ \}$, karena tidak ada $\Sigma \geq \Phi$ sehingga $F4, F5, F6$ dan $F7$ juga merupakan himpunan kosong.

Tentukan (ss-s) sebagai antecedent dan s sebagai consequent dari F_k yang telah didapat. Pada $F2$ didapat himpunan $F2 = \{\{\text{The GAP, Little M}\}, \{\text{The GAP, Gymboree}\}, \{\text{Little M, Gymboree}\}\}$

Dari langkah di atas, didapat 6 rule yaitu:

- Jika pembeli membeli barang merek THE GAP maka pembeli juga membeli barang merek Little M.
- Jika pembeli membeli barang merek LITTLE M maka pembeli juga membeli barang merek THE GAP
- Jika pembeli membeli barang merek THE GAP maka pembeli juga membeli barang merek GYMBOREE
- Jika pembeli membeli barang merek GYMBOREE maka pembeli juga membeli barang merek THE GAP
- Jika pembeli membeli barang merek LITTLE M maka pembeli juga membeli barang merek GYMBOREE
- Jika pembeli membeli barang merek GYMBOREE maka pembeli juga membeli barang merek LITTLE M

Hitung support dan confidence dengan ketentuan support = 40% dan confidence 60%

Tabel 6. Nilai Support & Confidence

Condition	Support	Confidence
If buy The GAP then buy Little M	$(3/7) \times 100\% = 40\%$	$(3/5) \times 100\% = 60\%$

If buy Little M then buy The GAP	$(3/7) \times 100\% = 40\%$	$(3/5) \times 100\% = 60\%$
If buy The GAP then buy Gymboree	$(3/7) \times 100\% = 40\%$	$(3/5) \times 100\% = 60\%$
If buy Gymboree then buy The GAP	$(3/7) \times 100\% = 40\%$	$(3/5) \times 100\% = 60\%$
If buy Little M then buy Gymboree	$(3/7) \times 100\% = 40\%$	$(3/5) \times 100\% = 60\%$
If buy Gymboree then buy Little M	$(3/7) \times 100\% = 40\%$	$(3/5) \times 100\% = 60\%$

Tabel 7. Perkalian Support & Confidence

Condition	Support	Confidence	Support x Confidence
If buy Gymboree then buy The GAP	40%	60%	0.32145
If buy Gymboree then buy Little M	40%	60%	0.32145

Hasil paling besar dari perkalian perkalian tersebut merupakan rule yang dipakai pada saat menjual. Karena hasil perkalian dari ke-2 penjualan di atas bernilai sama, maka semuanya bisa dijadikan rule. Maka :

- a. Jika menjual Gymboree maka akan menjual The GAP dengan tingkat kepercayaan 60% dan didukung oleh 40% dari data keseluruhan.
- b. Jika menjual Gymboree maka akan menjual Little M tingkat kepercayaan 60% dan didukung oleh 40% dari data keseluruhan.

SIMPULAN

Berdasarkan uraian sebelumnya, dapat ditarik beberapa kesimpulan, yaitu:

1. Algoritma *Apriori* yang digunakan dan diterapkan pada distro fantasy kids Bangkinang dapat menampilkan informasi pembelian barang oleh pelanggan.
2. Pada algoritma *Apriori* berdasarkan nilai minimum support 40% dan *minimum confidence* 60% yang diberikan maka didapat hasil *association rule* sebanyak 6 rule.

3. Penggunaan algoritma *Apriori* untuk mendapatkan *frequent itemset* dapat melihat barang apa saja yang harus tersedia pada fantasy kids Bangkinang.

DAFTAR PUSTAKA

- [1] Virgiawan Mitra Donny, I. M. (2013). *Aplikasi Association Rule Mining Untuk Menentukan Pola Pada Data Nilai Mahasiswa Matematika ITS Surabaya*.
- [2] Larose, D. . (2013). *Data Mining Methods and Models*.
- [3] Yuliana, D. (2015). Penerapan Metode *Market Basket Analisis* Dengan Algoritma
- [4] Kusrini & Andi. (2007). *Perancangan Sistem Informasi dan Aplikasinya*. Andi Yogyakarta.
- [5] Nurchalifatun, F. (n.d.). Penerapan Metode Asosiasi Data Mining Menggunakan Algoritma *Apriori* Untuk Mengetahui Kombinasi Antar Item Set Pada Pondok Kopi