

PENERAPAN DATA MINING DALAM PENENTUAN ATURAN ASOSIASI ANTAR JENIS ITEM

Yogi Yusuf W., F. Rian Pratikto, Gerry T.

Jurusan Teknik Industri, Universitas Katolik Parahyangan

Jalan Ciembuleuit 94 Bandung

E-mail: yogi@home.unpar.ac.id

ABSTRAKSI

Data tidak dapat dipisahkan dari kehidupan sehari-hari. Suatu organisasi baik besar maupun kecil dapat dibanjiri dengan berbagai macam data, tidak terkecuali dengan CV JP Bogor. Saat ini pemanfaatan data-data yang dimiliki belum maksimal, baru sebatas untuk pembuatan laporan. Data transaksi penjualan yang terkumpul dan tersimpan dapat memberikan pengetahuan yang bermanfaat bagi manajemen dalam melakukan usaha-usaha yang terkait dengan peningkatan penjualan dengan melakukan promosi yang tepat dan membekali tenaga penjual dengan pengetahuan mengenai kebiasaan customer. Pengetahuan mengenai asosiasi antar jenis produk, jenis-jenis produk yang muncul bersamaan pada tiap transaksi, dapat menjadi masukan penting dalam melakukan usaha peningkatan penjualan tersebut. Salah satu teknologi yang dapat digunakan untuk mengolah data menjadi pengetahuan adalah data mining. Hasil penambangan terhadap data transaksi penjualan memberikan tiga aturan yang memenuhi batas confidence sebesar 80%. Aturan-aturan tersebut adalah: CD2R \Rightarrow RT dengan confidence 87.321% dan support 15.93%, CD5R \Rightarrow DM dengan confidence 98.734% dan support 2.91%, dan G16 dan FF364 \Rightarrow BAA dengan confidence 100% dan support 2.015%.

Kata kunci: Data Mining, Aturan Asosiasi, Market Basket Analysis, Algoritma A Priori, Confidence, Support

1. PENDAHULUAN

Data tidak dapat dipisahkan dari kehidupan sehari-hari. Suatu organisasi baik besar maupun kecil dapat dibanjiri dengan berbagai macam data, tidak terkecuali dengan CV JP Bogor. Khusus untuk transaksi pembelian, setiap hari CV JP melayani sekitar 20 transaksi penjualan yang melibatkan 109 jenis item. Banyak data yang terlibat dalam setiap transaksi penjualan ini seperti data item yang dibeli, waktu pembelian, jumlah item yang dibeli, harga item, dan data customer.

Saat ini pemanfaatan data-data yang dimiliki belum maksimal, baru sebatas untuk pembuatan laporan. Data transaksi penjualan yang terkumpul dan tersimpan dapat memberikan pengetahuan yang bermanfaat bagi pengambil keputusan dalam melakukan usaha-usaha yang terkait dengan peningkatan penjualan dengan melakukan promosi yang tepat dan membekali tenaga penjual dengan pengetahuan mengenai kebiasaan customer.

Pengetahuan mengenai asosiasi antar jenis item, jenis-jenis item yang muncul bersamaan pada tiap transaksi, dapat menjadi masukan penting dalam melakukan usaha peningkatan penjualan. Misalkan dalam melakukan promosi untuk pembelian satu jenis item dapat membeli jenis item lain dengan harga lebih murah, pengetahuan asosiasi antar jenis item ini dapat memberikan pengetahuan mengenai kombinasi item yang harus tercakup dalam promosi tersebut.

Permasalahan yang muncul adalah bagaimana memproses data-data transaksi penjualan tersebut sehingga dapat menampilkan suatu pengetahuan yang berguna bagi manajemen dalam mengambil

keputusan. Salah satu teknologi yang dapat digunakan untuk mengolah data menjadi pengetahuan adalah data mining. Dalam makalah ini akan disajikan penggunaan data mining dalam menemukan pengetahuan mengenai asosiasi antar jenis item di CV JP Bogor.

2. DATA MINING

Data mining membantu perusahaan untuk mendapatkan pola dari data-data yang tersimpan di dalam basis data perusahaan. Pengetahuan yang diperoleh tersebut akan menjadi pedoman dalam mengambil tindakan-tindakan bisnis sebagai upaya pemeliharaan dan peningkatan tingkat kompetitif bisnis perusahaan. Walaupun sudah banyak perangkat lunak yang menawarkan kemampuan dalam proses data mining, keterlibatan manusia sangat dibutuhkan dalam setiap fase proses data mining itu sendiri. Pemahaman terhadap model statistik dan matematik yang digunakan dalam perangkat lunak sangat dituntut.

Banyak pihak yang telah mendefinisikan data mining. Berikut beberapa definisi data mining:

“Data mining merupakan suatu proses pencarian pola dari data-data dengan jumlah yang sangat banyak yang tersimpan dalam suatu tempat penyimpanan dengan menggunakan teknologi pengenalan pola, teknik statistik, dan matematik.” (Gartner Group)

“Data mining merupakan analisis dari sekumpulan data yang diamati (sangat besar) untuk menemukan hubungan yang tidak terduga dan merangkum data dengan cara yang baru yang dapat dipahami dan berguna bagi pemilik data.” (Hand et al.)

“Data mining merupakan proses eksplorasi dan analisis data yang berjumlah sangat besar untuk menemukan pola dan aturan yang berarti.” (Berry and Linoff)

2.1 Fungsi Data Mining

Data mining dapat menjalankan fungsi-fungsi berikut:

Deskripsi

Pola dan trend data sering dideskripsikan. Deskripsi tersebut sangat membantu dalam menjelaskan pola dan trend yang terjadi. Model data mining harus setransparan mungkin, dimana hasilnya dapat mendeskripsikan pola dengan jelas.

Estimasi

Estimasi mirip dengan klasifikasi kecuali variabel target-nya numerik ketimbang kategorikal. Model yang dibangun menggunakan record yang lengkap, yang menyediakan nilai variabel target dan predictor. Untuk observasi yang baru, estimasi nilai variabel target ditentukan, berdasarkan nilai-nilai predictor.

Prediksi

Prediksi mirip dengan klasifikasi dan estimasi, kecuali dalam prediksi, hasil terjadi di masa datang.

Klasifikasi

Dalam klasifikasi, variabel target-nya merupakan kategorikal. Model data mining memeriksa set record yang besar, tiap record mempunyai informasi variabel target dan set input atau variabel predictor.

Clustering

Clustering merupakan pengelompokkan record, observasi, atau kasus ke dalam kelas-kelas objek yang mirip. Clustering berbeda dengan klasifikasi dimana dalam clustering tidak terdapat variabel target. Clustering mencoba menyegmentasi seluruh set data ke dalam subgroup atau cluster yang relatif homogen, dimana kemiripan antar record dalam cluster dimaksimalkan dan kemiripan record di luar cluster diminimasi.

Asosiasi

Asosiasi merupakan suatu tugas untuk menemukan atribut-atribut yang “terjadi” bersamaan. Tugas asosiasi mencoba untuk menemukan aturan untuk mengkuantifikasi hubungan antara dua atau lebih atribut. Aturan asosiasi berbentuk “If antecedent, then consequent”, bersama-sama dengan ukuran support dan confidence yang berhubungan dengan aturan.

2.2 Data Preprocessing

Data mentah yang tersimpan di dalam basis data banyak yang tidak dipraproses, tidak lengkap, contoh:

- a. Field yang usang atau redundan

- b. Nilai yang hilang
- c. Outlier
- d. Bentuk data tidak sesuai dengan model data mining.
- e. Nilai tidak konsisten dengan common sense.

Agar data-data yang tersimpan di dalam basis data berguna bagi keperluan data mining, basis data yang ada harus menjalani praproses, dengan cara data cleaning dan data transformation.

2.3 Aturan Asosiasi

Affinity analysis melakukan studi atribut atau karakteristik yang berjalan bersamaan. Metode affinity analysis, yang dikenal sebagai market basket analysis, mencoba menemukan asosiasi antar atribut; mendapatkan aturan yang mengkuantifikasi hubungan antara dua atribut atau lebih. Aturan asosiasi berbentuk “If antecedent, then consequent”, dilengkapi dengan tingkat support dan confidence aturan tersebut.

Ada dua metode utama dalam merepresentasikan tipe data dalam market basket analysis: format data transaksi dan format data tabular. Format data transaksi membutuhkan dua field, field ID dan field isi, dimana tiap record merepresentasikan hanya satu item. Format data transaksi untuk contoh di atas.

Field ID	Field Isi
1	Item 1
1	Item 2
2	Item n
.	.
.	.
.	.

Gambar 1. Format Data Transaksi

Dalam format data tabular, tiap record merepresentasikan transaksi terpisah, dimana flag field 0/1 sebanyak itemnya.

Transaksi	Item 1	Item 2	...	Item n
1	1	1	0	0
2	0	0	0	1
.
.
.

Gambar 2. Format Data Tabular

2.3.1 Support, Confidence, Frequent Itemset

Misalkan D adalah set transaksi, dimana tiap transaksi T di D merepresentasikan set item yang ada di I. Aturan asosiasi berbentuk if A, then B ($A \Rightarrow B$), dimana antecedent A dan consequent B subset dari I, dan A dan B mutually exclusive.

Support s untuk aturan asosiasi tertentu $A \Rightarrow B$ adalah proporsi transaksi di D yang berisi baik A dan B.

$$\text{support} = P(A \cap B) = \frac{\text{number of transactions containing both A and B}}{\text{total number of transactions}}$$

Confidence c aturan asosiasi tertentu $A \Rightarrow B$ adalah ukuran akurasi aturan, ditentukan dengan

persentase transaksi di D yang berisi A yang juga berisi B.

$$\text{confidence} = P(B | A) = \frac{P(A \cap B)}{P(A)}$$

= $\frac{\text{number of transactions containing both A and B}}{\text{number of transactions containing A}}$

Aturan yang diinginkan adalah aturan yang mempunyai *support* yang tinggi atau *confidence* yang tinggi, dan biasanya keduanya. Aturan yang kuat adalah aturan yang memenuhi atau melebihi kriteria *support* dan *confidence* minimum.

Itemset adalah set item yang berisi di dalam I, dan k-itemset adalah itemset yang berisi k items. Itemset frequency adalah jumlah transaksi yang berisi itemset tertentu. **Frequent itemset** adalah itemset yang terjadi paling sedikit pada jumlah tertentu, mempunyai itemset frequency $\geq \Phi$. Misalkan, $\Phi = 4$, itemset yang terjadi lebih dari 4 disebut frequent. Notasi set of frequent k-itemsets adalah F_k .

2.3.2 Penambangan Aturan Asosiasi

Penambangan aturan asosiasi dari basis data melalui dua tahap:

1. Cari seluruh frequent itemset; Cari seluruh itemset dengan frekuensi $\geq \Phi$.

Ada set transaksi D. Misalkan $\Phi = 4$, itemset frequent jika itemset tersebut terjadi empat kali atau lebih di dalam D.

Pertama temukan F_1 , 1-itemsets yang frequent, yang merepresentasikan item individual. Berikutnya cari frequent 2-itemsets. Secara umum untuk menemukan F_k , algoritma a priori pertama-tama membangun set C_k dari kandidat k-itemsets dengan menggabungkan F_{k-1} dengan dirinya sendiri. Kemudian algoritma memangkas C_k dengan menggunakan properti a priori. Itemsets dalam C_k yang bertahan membentuk F_k .

2. Dari frequent itemsets, bangkitkan aturan asosiasi yang memenuhi kondisi support dan confidence minimum.

Setelah frequent itemsets ditentukan, aturan asosiasi dapat dibangkitkan. Ada dua tahap dalam pembangkitan aturan asosiasi ini, untuk tiap frequent itemset s:

Bangkitkan seluruh subset dari s.

Misalkan ss merepresentasikan nonempty subset dari s. Aturan asosiasi R: $ss \Rightarrow (s - ss)$, dimana $(s - ss)$ mengindikasikan set s tanpa ss. Bangkitkan (dan output) R jika R memenuhi persyaratan confidence minimum. Lakukan untuk setiap subset ss dari s. Untuk kemudahan, single-item consequent sangat diinginkan.

Berikutnya aturan asosiasi untuk single antecedent/single consequent dapat dibangkitkan. Tabel berikut menunjukkan kandidat aturan asosiasi untuk single antecedent/single consequent.

Algoritma a priori mengambil keuntungan dari properti a priori untuk menyusutkan ruang pencarian. Properti a priori menyatakan jika suatu itemset Z bukan frequent, menambahkan item yang lain A ke itemset Z tidak akan membuat Z lebih frequent. Jika Z tidak frequent, $Z \cup A$ tidak akan frequent. Tidak ada superset dari Z (itemset yang berisi Z) akan frequent.

3. PENAMBANGAN DATA TRANSAKSI PENJUALAN

3.1 Pemilihan Variabel

Transaksi penjualan menyimpan data-data no faktur, tanggal transaksi, kode item, jumlah item yang dibeli, dan harga. Karena tujuan utama dari penelitian ini adalah pencarian pola asosiasi antar item, maka date field yang dibutuhkan adalah no faktur dan kode item. Berikut tampilan dari sebagian data dalam format tabular.

NoTransaksi	BAA	C16R	C24R	C2R	C4R	C5R	C8R
1080506	1.00	.00	.00	.00	.00	.00	.00
1080507	.00	.00	.00	.00	.00	.00	.00
1080508	.00	.00	.00	.00	1.00	.00	.00
1080509	.00	.00	.00	.00	.00	.00	.00
1080510	1.00	.00	.00	.00	.00	.00	.00
1080511	.00	.00	.00	.00	1.00	.00	.00
1080512	.00	.00	.00	.00	.00	.00	.00
1080513	.00	.00	.00	.00	.00	.00	.00
1090501	1.00	.00	.00	.00	.00	.00	.00
1090502	.00	.00	.00	.00	.00	.00	.00
1090503	.00	.00	.00	.00	.00	.00	.00
1090504	.00	.00	.00	.00	1.00	.00	.00
1090505	1.00	.00	.00	.00	.00	.00	.00
1090506	.00	.00	.00	.00	.00	.00	.00
1090507	.00	.00	.00	1.00	.00	1.00	.00

Gambar 3. Contoh Tampilan Data

3.2 Pemodelan Asosiasi

Pemodelan asosiasi ini menggunakan algoritma A Priori dan pemrosesannya dibantu dengan software Clementine. Berikut hasil pemrosesan data dengan algoritma A Priori.

Consequent	Antecedent	Support %	Confidence %
BAA	G16 FF364	2.015	100.000
DM	CD5R	2.948	98.734
RT	CD2R	18.246	87.321
BAA	FF242	3.955	75.472
CD2R	RT	21.231	75.044
F8R1206	CC24	2.425	72.308
BAA	FF364	6.940	70.430
DM	CD4R	11.306	67.327
BAA	G16	6.231	65.269

Gambar 4. Asosiasi Antar Item

Dari gambar 4 di atas terlihat ada 9 aturan yang dihasilkan dengan batas confidence sebesar 60%. Support sesungguhnya untuk masing-masing aturan harus dikalikan antara support dengan confidence yang dihasilkan clementine.

Aturan dapat dipilih berdasarkan batas confidence ataupun support tertentu, dalam

penelitian ini batas confidence ditetapkan sebesar 80%. Dari 9 aturan tersebut, terdapat tiga aturan yang memenuhi batas yang ditetapkan. Aturan-aturan tersebut adalah:

1. Jika CD2R, maka RT dengan confidence 87.321% dan support 15.93%.
2. Jika CD5R, maka DM dengan confidence 98.734% dan support 2.91%.
3. Jika G16 dan FF364, maka BAA dengan confidence 100% dan support 2.015%.

3.3 Evaluasi Model

Evaluasi model bertujuan untuk menilai apakah bermanfaat aturan yang dihasilkan dengan membandingkan confidence aturan dengan prior probability.

Confidence aturan 1: 87.321%
Prior probability RT: 5.9%

Aturan 1 ini sangat diinginkan dimana nilai confidence-nya lebih besar dari prior probability. Hal ini dapat meningkatkan peluang terjadinya RT. Misalkan manajemen dapat menginstruksikan kepada tenaga penjual untuk selalu mengingatkan customer yang membeli CD2R untuk juga membeli RT ketimbang mengingatkan seluruh customer untuk membeli RT. Cara yang pertama yang hanya mengingatkan customer yang membeli CD2R memberikan peluang sebesar 87.321% customer tersebut akan membeli RT. Cara yang kedua hanya memberikan peluang sebesar 5.9% customer yang diingatkan akan membeli RT.

Confidence aturan 2: 98.734%
Prior probability DM: 4.6%

Confidence aturan 3: 100%
Prior probability: 3.6%

Aturan 2 dan aturan 3 juga sama mempunyai confidence yang lebih besar dari prior probability, sehingga kedua aturan tersebut memberikan manfaat bagi manajemen.

4. USULAN TINDAKAN

Pengetahuan yang diperoleh dari aturan asosiasi ini dapat dijadikan pedoman bagi manajemen untuk menentukan upaya yang dapat dilakukan untuk meningkatkan penjualan. Manajemen dapat membuat paket-paket produk yang berisi kombinasi dari produk-produk yang diperoleh dari asosiasi item. Paket yang dapat ditawarkan adalah paket yang berisi produk CD2R dan RT, CD5R dan DM, dan G16, FF364, dan BAA.

Manajemen juga dapat memberikan pelatihan kepada karyawan khususnya yang berada di bagian depan penjualan untuk selalu mengingatkan customer jika customer tersebut membeli produk CD2R maka tawarkan produk RT; CD5R maka tawarkan produk DM; G16, FF364 maka tawarkan produk BAA.

5. PENUTUP

Data mining dapat membantu manajemen dalam pengambilan tindakan-tindakan bisnis dengan membekali pengetahuan berupa pola yang berasal dari data-data masa lalu. Dalam penelitian ini diperoleh pengetahuan berupa pola asosiasi antara satu produk dengan produk yang lain. Aturan yang dihasilkan harus dievaluasi terlebih dahulu sebelum diaplikasikan. Hasil dari evaluasi memberikan hasil yang memuaskan, di mana aturan asosiasi memberikan manfaat yang lebih besar ketimbang tidak menggunakan aturan sama sekali.

DAFTAR PUSTAKA

- Berry, Michael J.A., Linoff, G., *Data Mining Techniques: For Marketing, Sales, and Customer Support*, John Wiley & Sons, New York, 1997.
- Jawei Han and Micheline Kamber, *Data Mining Concepts and Techniques*, Morgan Kaufmann, San Fransisco, CA, 2001.
- Larose, Daniel T., *Discovering Knowledge in Data: An Introduction to Data Mining*, Wiley-Interscience, New Jersey, 2005.