

Nama : Akmal Zuhdy Prasetya
NIM : H071191035
Kelas : Machine Learning

Resume Pertemuan 2

A. Association Rules

Dalam istilah dasar, aturan asosiasi menyajikan hubungan antar item. Mereka adalah pernyataan yang membantu untuk menemukan hubungan antara data dalam database. Aturan asosiasi dapat didefinisikan sebagai implikasi dari bentuk $A \rightarrow B$. Di sini 'A' disebut premis, yang mewakili kondisi yang harus benar agar 'B' dapat dipertahankan. 'B' adalah kesimpulan yang terjadi ketika 'A' benar. 'A' disebut *antecedent* dan 'B' disebut *consequent*. Antecedent adalah elemen yang ditemukan dalam data sedangkan konsekuensi ditemukan dalam kombinasi dengan antecedent. Aturan $A \rightarrow B$ dapat diartikan sebagai "Jika A terjadi, B terjadi". Ini adalah interpretasi yang sangat umum, interpretasi yang akurat tergantung pada domain.

Aturan asosiasi dapat diterapkan jika kondisi data memenuhi syarat yaitu data harus memiliki bentuk berupa transaksi. Transaksi berarti sekelompok item logis dalam konteks ini. Transaksi ini tidak terkait dengan transaksi database. Transaksi dapat berupa sekelompok barang belanjaan, daftar film, dan lain-lain.

Dalam data mining, interpretasi aturan asosiasi hanya bergantung pada data apa yang kita tambang. Dalam contoh ini, transaksi berarti isi keranjang. Setiap transaksi memberi kita isi sekeranjang pelanggan. Tabel di bawah ini adalah datanya.

Transaction ID	Items
1	Bread, Milk, Cheese, Butter
2	Bread, Milk, Cheese, Butter, Eggs, Sugar
3	Milk, Butter, Cocoa
4	Bread, Milk, Cheese, Eggs

Kita memiliki data empat pelanggan dan isi keranjang pasar mereka. Berdasarkan data tersebut, kita dapat melihat bahwa transaksi 1, 2 dan 4 mengandung roti dan susu. Ini dapat dikonversi sebagai aturan asosiasi pertama,

yaitu, *Roti* → *Susu*. Dengan cara berpikir yang sama, kita dapat mengetahui bahwa susu dan mentega hadir dalam transaksi 1, 2, dan 3. Ini mengarah ke aturan asosiasi lain yaitu, *Susu* → *Mentega*. Namun, terdapat masalah. Kita belum menemukan aturan mana yang lebih baik? Kita tidak mungkin membandingkan mereka semua. Untuk mengatasi masalah ini, kita dapat menggunakan beberapa *classifier*. Pengklasifikasi ini akan menunjukkan kekuatan aturan. Mereka juga disebut sebagai '*ukuran ketertarikan*', karena kekuatan aturan sebanding dengan daya tariknya.

B. Apriori Algorithm

Algoritma Apriori merupakan algoritma yang digunakan untuk fase pertama aturan asosiasi. Properti dan data algoritma ini dievaluasi dengan aturan asosiasi boolean. Dalam algoritma ini, ada cluster produk yang sering muncul dan kemudian dicari hubungan yang kuat antara produk ini dan produk lainnya. Pentingnya aturan asosiasi dapat ditentukan oleh 3 parameter yang digunakan untuk mengidentifikasi kekuatan algoritma, yaitu *Support*, *Confidence*, dan *Lift*. *Support* memberikan ukuran seberapa sering aturan itu diterapkan, *Confidence* adalah persentase dari semua transaksi, yang berisi item di sebelah kiri dan di sebelah kanan aturan. Sedangkan *Lift* adalah probabilitas semua item yang terjadi bersama-sama dibagi dengan produk anteseden dan konsekuen yang terjadi seolah-olah mereka independen satu sama lain.

Adapun cara menghitung *Support*, *Confidence*, dan *Lift* adalah sebagai berikut:

Misalkan X dan Y mewakili produk di pasar dan N mewakili jumlah total produk, maka dapat dituliskan:

$$RULE: X \Rightarrow Y$$

$$SUPPORT = \frac{frq(X,Y)}{N}$$

$$CONFIDENCE = \frac{frq(X,Y)}{frq(X)}$$

$$LIFT = \frac{SUPPORT}{SUPPORT(X) \times SUPPORT(Y)}$$

Berikut adalah contoh agar kita lebih memahami konsep *support*, *confidence*, dan *lift*.

Transaction ID	Items	Bread + Mayo	Bread -> Mayo	Mayo -> Bread
1	Bread, Mayo, Chicken, Cheese	1	1	1
2	Eggs, Bread, Mayo, Apples	1	1	1
3	Eggs, Bread, Cheese	None	0	None
4	Bread, Eggs	None	0	None
5	Eggs, Apple, Mayo	None	None	0
6	Eggs, Bread, Mayo	1	1	1

Mari kita coba menganalisis data pada tabel di atas. Roti dan mayo keduanya ada di keranjang transaksi 1, 2 dan 6. Jumlah total transaksi adalah enam. Jadi, *support* untuk aturan '*Roti → Mayo*' dan '*Mayo → Roti*' akan menjadi $\frac{3}{6} = 50\%$. *Confidence* dihitung sebagai jumlah transaksi di mana kedua item hadir yaitu, 3 dibagi dengan jumlah total transaksi yaitu 6. Untuk menghitung *confidence*, kita harus melihat jumlah transaksi di mana mayo dan roti ada di keranjang.

Sekarang pertimbangkan hanya aturan '*Roti → Mayo*'. Roti ada dalam keranjang 1, 2, 3, 4 dan 6, jadi totalnya 5 keranjang. Dari jumlah tersebut, mayo ada di keranjang 1, 2 dan 6. Jadi kepercayaan dari aturan '*Roti → Mayo*' adalah $\frac{3}{5} = 60\%$.

Jadi, sekarang kita dapat menghitung Support dan Confidence dari aturan asosiasi. Ini dapat dengan mudah digunakan untuk menyaring aturan yang tidak menarik dengan menetapkan ambang batas (threshold), seperti support minimum harus 50% dan confidence minimum harus 75% untuk menganggap aturan itu menarik. Menetapkan nilai-nilai ukuran ini akan menentukan jumlah aturan yang akan menarik. Jadi, jika nilainya terlalu rendah, mungkin terlalu banyak aturan yang tergolong menarik dan sebaliknya. Jadi nilai yang sesuai harus diberikan.

Kesimpulannya, kita dapat mengatakan bahwa ada berbagai parameter kontrol dalam data mining menggunakan aturan asosiasi yang dapat digunakan untuk mendapatkan output yang diinginkan tergantung pada kasusnya. Pertanyaan yang paling penting adalah apa yang kita tambang. Terlepas dari Support dan Confidence, banyak ukuran menarik lainnya yang ada untuk data mining menggunakan aturan asosiasi yang dapat digunakan dan yang mungkin bekerja lebih baik dalam kasus tertentu. Data mining menggunakan aturan asosiasi memiliki aplikasi dalam penambangan penggunaan web, analisis keranjang pasar,

bioinformatika, perawatan kesehatan, proses aliran berkelanjutan, dan lain-lain. Dan oleh karena itu merupakan konsep baru yang menarik yang dapat membantu meningkatkan efisiensi.