

Modul 5

Association Rules

1.1. Tujuan Praktikum

Setelah menyelesaikan praktikum ini, mahasiswa mampu:

- Memahami Konsep Association Rules
- Mengimplementasikan Association Rules menggunakan RapidMiner

1.2. Peralatan yang dibutuhkan

Beberapa peralatan yang dibutuhkan dalam menyelesaikan praktikum ini adalah:

- Aplikasi Notepad atau sejenisnya
- Aplikasi RapidMiner versi 7 ke atas
- Aplikasi Microsoft Excel

1.3. Dasar Teori

1.3.1. Association Rules

Aturan asosiasi (Association Rules) adalah teknik dalam analisis data yang digunakan untuk menemukan hubungan dan korelasi antara dua atau lebih item dalam basis data transaksional atau data yang diwakili dalam bentuk tabel transaksi. Tujuan utama dari aturan asosiasi adalah untuk mengungkap pola-pola yang tersembunyi dalam data dan memberikan wawasan tentang hubungan antara item-item tersebut.

Aturan asosiasi digunakan secara luas dalam berbagai bidang seperti analisis belanja ritel, rekomendasi produk, analisis keranjang belanja online, dan lain-lain. Beberapa contoh aplikasi termasuk rekomendasi produk kepada pelanggan berdasarkan pembelian sebelumnya, deteksi pola pembelian yang berpotensi untuk promosi produk, dan analisis pola perilaku pelanggan.

Konsep utama dalam aturan asosiasi adalah sebagai berikut:

1. **Support:** Support adalah ukuran seberapa sering kombinasi item tertentu muncul bersama dalam dataset. Hal ini dihitung sebagai proporsi transaksi yang mengandung semua item dalam aturan dibandingkan dengan total transaksi dalam dataset. Aturan dengan support yang tinggi menunjukkan bahwa kombinasi item tersebut sering muncul bersama dalam dataset.

2. **Confidence:** Confidence mengukur seberapa sering aturan tersebut terbukti benar. Dihitung sebagai proporsi transaksi yang mengandung semua item di sisi kanan aturan dibandingkan dengan transaksi yang mengandung item di sisi kiri aturan saja. Confidence yang tinggi menunjukkan bahwa item-item di sisi kanan aturan cenderung muncul bersamaan dengan item-item di sisi kiri.

3. **Lift:** Lift adalah ukuran seberapa banyak confidence aturan dibandingkan dengan support dari item di sisi kanan. Ini mengindikasikan seberapa banyak aturan memperbaiki prediksi, dibandingkan dengan jika prediksi dilakukan secara acak. Lift yang tinggi menunjukkan bahwa hubungan antara item-item tersebut kuat dan berguna untuk membuat rekomendasi.

Contoh aturan asosiasi yang terkenal adalah "Jika seseorang membeli roti, maka dia juga kemungkinan besar akan membeli mentega." Aturan ini memiliki support yang tinggi karena kedua item (roti dan mentega) sering muncul bersama dalam transaksi, dan confidence yang tinggi karena kebanyakan orang yang membeli roti juga membeli mentega. Dengan menggunakan aturan ini, toko atau supermarket dapat membuat rekomendasi untuk menempatkan roti dan mentega berdekatan untuk meningkatkan penjualan.

Contoh Perhitungan:

Berikut ini contoh perhitungan Association Rules dengan menggunakan data transaksi sederhana. Misalkan kita memiliki data transaksi belanja di sebuah toko kelontong seperti di bawah ini:

Transaksi 1: roti, susu, telur

Transaksi 2: roti, mentega, telur

Transaksi 3: susu, mentega, roti, telur

Transaksi 4: roti, susu

Transaksi 5: roti, mentega, telur

Selanjutnya dihitung support dan confidence dari beberapa aturan asosiasi yang mungkin. Misalnya, kita tertarik untuk menemukan aturan "Jika seseorang membeli roti, maka dia juga kemungkinan besar akan membeli susu".

1. Support (roti -> susu):

- Support adalah proporsi transaksi yang mengandung kedua item, roti dan susu.
- Jadi, jumlah transaksi yang mengandung roti dan susu adalah 2 (Transaksi 1 dan 4), dan total transaksi adalah 5.
- Maka, support untuk aturan "roti -> susu" adalah $2/5 = 0.4$ atau 40%.

2. Confidence (roti -> susu):

- Confidence adalah proporsi transaksi yang mengandung kedua item, roti dan susu, dari semua transaksi yang mengandung roti.
- Jadi, jumlah transaksi yang mengandung roti dan susu adalah 2 (Transaksi 1 dan 4), dan jumlah transaksi yang hanya mengandung roti adalah 4 (Transaksi 1, 2, 4, dan 5).
- Maka, confidence untuk aturan "roti -> susu" adalah $2/4 = 0.5$ atau 50%.

Dengan demikian, aturan "roti -> susu" memiliki support sebesar 40% dan confidence sebesar 50%.

Artinya: "50% dari transaksi di database yang memuat item roti dan mentega juga memuat item susu. Sedangkan 40% dari seluruh transaksi yang ada di database memuat ketiga item itu."

Dapat juga diartikan: "Seorang konsumen yang membeli roti dan mentega punya kemungkinan 50% untuk juga membeli susu. Aturan ini cukup signifikan karena mewakili 40% dari catatan transaksi selama ini."

Kita juga dapat menghitung aturan lainnya dengan cara yang sama. Misalnya, "mentega -> roti":

1. Support (mentega -> roti):

- Jumlah transaksi yang mengandung mentega dan roti adalah 3 (Transaksi 2, 3, dan 5), dan total transaksi adalah 5.
- Support untuk aturan "mentega -> roti" adalah $3/5 = 0.6$ atau 60%.

2. Confidence (mentega -> roti):

- Jumlah transaksi yang mengandung mentega dan roti adalah 3 (Transaksi 2, 3, dan 5), dan jumlah transaksi yang hanya mengandung mentega adalah 3 (Transaksi 2, 3, dan 5).

- Confidence untuk aturan "mentega -> roti" adalah $3/3 = 1$ atau 100%.

Dengan demikian, aturan "mentega -> roti" memiliki support sebesar 60% dan confidence sebesar 100%.

Perhitungan support dan confidence ini memberikan kita gambaran tentang seberapa sering pola belanja tertentu muncul dalam data, dan seberapa kuat hubungan antara item-item tersebut.

Percobaan 1:

Pada percobaan ini kita akan menggunakan dataset **Chapter05Dataset.CSV**

- Roger adalah manajer tingkat kota. Kota ini memiliki keterbatasan sumber daya, dan seperti kebanyakan kota, ada lebih kebutuhan dari sumber yang ada.
- Dia merasa seperti warga di masyarakat cukup aktif dalam berbagai organisasi masyarakat, dan percaya bahwa dia mungkin bisa mendapatkan sejumlah kelompok untuk bekerja sama untuk memenuhi beberapa kebutuhan di masyarakat.
- Dia tahu ada tempat ibadah, klub sosial, penggemar hobi dan jenis-jenis kelompok di masyarakat.
- Apa yang tidak dia tahu adalah apakah ada hubungan antara kelompok-kelompok yang memungkinkan kolaborasi alami antara dua atau lebih kelompok yang bisa bekerja sama dalam proyek di sekitar kota.
- Dia memutuskan bahwa sebelum ia dapat mulai bertanya organisasi masyarakat untuk mulai bekerja bersama-sama dan menerima tanggung jawab untuk proyek-proyek, ia perlu untuk mencari tahu apakah ada asosiasi yang ada antara berbagai jenis kelompok di daerah.

Business Understanding

Tujuan Roger adalah untuk mengidentifikasi dan kemudian mencoba untuk mengambil keuntungan dari koneksi yang ada di masyarakat setempat untuk mendapatkan pekerjaan yang akan bermanfaat bagi seluruh masyarakat. Dia tahu dari banyak organisasi di kota, memiliki informasi kontak untuk mereka dan bahkan terlibat dalam beberapa dari mereka sendiri. Perhatian utamanya adalah menemukan jenis organisasi yang tampaknya akan

terhubung dengan satu sama lain. Mengidentifikasi individu yang aktif di tiap tempat ibadah, klub sosial atau organisasi politik akan menjadi informasi yang penting tanpa terlebih dahulu mengelompokkan organisasi ke dalam kelompok dan mencari asosiasi antara kelompok. Setelah didapatkan korelasi yang dibutuhkan, dia siap untuk memulai menghubungi orang dan meminta mereka untuk menggunakan kontak lintas-organisasi mereka dan menjadi pimpinan program / aktifitas.

Data Understanding:

1. Elapsed_Time: Waktu yang dihabiskan setiap responden untuk menyelesaikan survei (menit)
2. Time_in_Community: Data ini menanyakan orang tersebut apakah mereka pernah tinggal di area rumah tangganya untuk 0-2 tahun, 3-9 tahun, atau 10+ tahun dan dicatat dalam kumpulan data sebagai Pendek, Sedang, atau Panjang.
3. Gender: Jenis kelamin responden survei.
4. Working: Sudah bekerja atau belum
5. Age: Umur
6. Family: Kolom ya/tidak yang menunjukkan apakah responden saat ini menjadi anggota atau tidak dari organisasi komunitas yang berorientasi keluarga, seperti Big Brothers/Big Sisters, anak-anak liga rekreasi atau olahraga, kelompok silsilah, dll.
7. Hobbies: Kolom ya/tidak yang menunjukkan apakah responden saat ini anggota organisasi komunitas yang berorientasi pada hobi, seperti radio amatir, aktifitas outdoor, bersepeda, dll.
8. Social_Club: Kolom ya/tidak yang menunjukkan apakah responden saat ini jadi anggota organisasi sosial masyarakat, seperti Rotary International, Lion's Club, dll.
9. Political: Kolom ya/tidak yang menunjukkan apakah responden saat ini jadi anggota organisasi politik yang mengadakan pertemuan rutin di masyarakat, seperti partai politik, kelompok aksi akar rumput, upaya lobi, dll.
10. Professional: Kolom ya/tidak yang menunjukkan apakah responden saat ini jadi anggota organisasi profesi yang mengadakan pertemuan cabang lokal, seperti komunitas hukum atau medis, kelompok pemilik umkm, dll.
11. Religious: Kolom ya/tidak yang menunjukkan apakah responden saat ini jadi anggota takmir masjid / gereja di areanya.

12. Support_Group: Support_Group: Kolom ya/tidak yang menunjukkan apakah responden saat ini jadi anggota organisasi komunitas yang berorientasi pada dukungan sosial, seperti Alcoholics Anonymous, kelompok manajemen kemarahan, dll.

Langkah-Langkah:

1. Import Data yang dibutuhkan dan jadikan sebagai Input Data

Row No.	Elapsed_Time	Time_in_Comm	Gender	Working	Age	Family	Hobbies	Social_Club	Political	Professional	Religious
1	6.710	Short	M	No	53	1	0	0	0	0	0
2	5.240	Medium	F	No	31	0	0	0	0	0	0
3	4.220	Medium	M	No	42	1	1	0	0	0	0
4	4.810	Long	F	No	30	0	0	0	0	0	0
5	3.950	Long	M	Yes	29	0	0	0	0	1	0
6	9.350	Long	F	No	40	0	0	0	0	0	0
7	2.910	Medium	F	Yes	33	0	0	0	0	0	0
8	4.540	Medium	M	Yes	27	1	1	1	1	0	0
9	4.790	Short	F	No	50	1	1	0	0	0	0
10	3.070	Medium	M	No	28	0	0	0	0	0	0
11	2.200	Medium	F	No	20	1	1	0	0	0	0
12	2.770	Medium	F	Yes	54	1	0	1	1	0	0
13	7.320	Long	M	Yes	48	1	0	0	0	0	0
14	2.230	Short	F	Yes	48	0	0	0	0	0	0
15	2.130	Medium	F	No	25	1	1	1	1	0	0
16	7.150	Short	F	Yes	25	0	0	0	0	0	0

2. Pilih atribut yang dibutuhkan

Select Attributes: attributes
The attribute which should be chosen.

Attributes

Search

Age
Elapsed_Time
Gender
Time_in_Community
Working

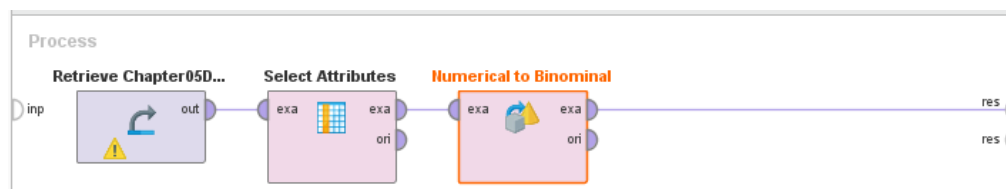
Selected Attributes

Search

Family
Hobbies
Political
Professional
Religious
Social_Club
Support_Group

Apply Cancel

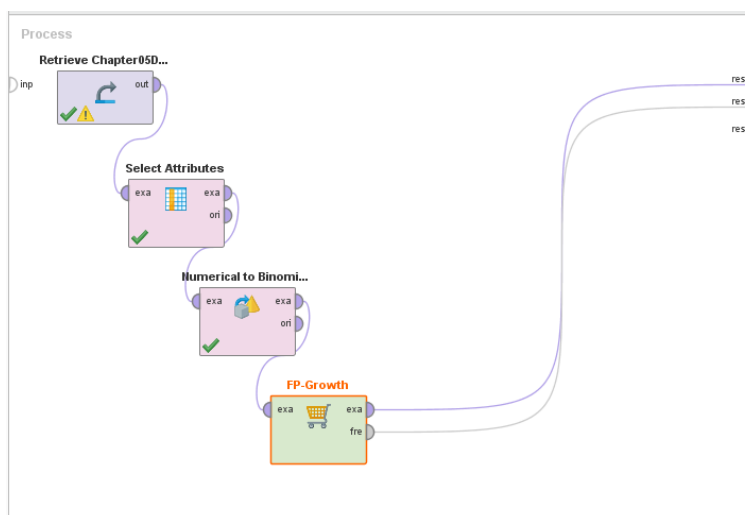
3. Konversi data dari numerical ke binomial. Hal ini dilakukan agar data diubah menjadi data Boolean



ExampleSet (3483 examples, 0 special attributes, 7 regular attributes) Filter (3,483 / 3,483 examples)

Row No.	Family	Hobbies	Social_Club	Political	Professional	Religious	Support_Gr...
1	true	false	false	false	false	false	false
2	false	false	false	false	false	true	true
3	true	true	false	false	true	false	false
4	false	false	false	false	false	false	false
5	false	false	false	true	true	false	true
6	false	false	false	false	true	false	false
7	false	false	false	false	false	false	true
8	true	true	true	false	false	true	false
9	true	true	false	false	true	true	false
10	false	false	false	false	false	true	true
11	true	true	false	false	true	false	false
12	true	false	true	false	false	true	false
13	true	false	false	false	false	false	false
14	false	false	false	false	true	false	false
15	true	true	true	false	false	true	false

4. Gunakan FP Growth. FP (Frequency Pattern). FP Growth dibutuhkan oleh Association Rules. Tanpa nilai frekuensi dari kombinasi atribut, kita tidak dapat menentukan pola apapun dalam data cukup sering muncul dan dapat digunakan sebagai rules. Perhatikan outputnya.



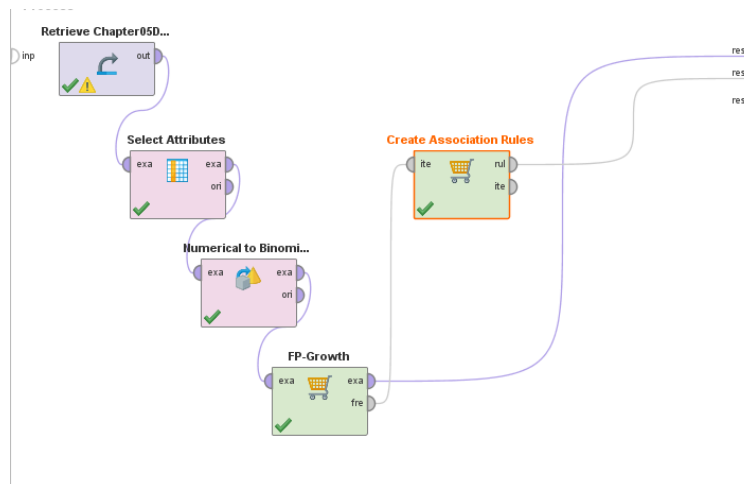
No. of Sets: 6
Total Max. Size: 2

Min. Size:
Max. Size:
Contains Item:

Size	Support	Item 1	Item 2
1	0.419	Religious	
1	0.390	Family	
1	0.324	Professional	
1	0.300	Hobbies	
2	0.225	Religious	Family
2	0.239	Religious	Hobbies

5. Dari hasil keluaran, kita dapat melihat bahwa beberapa atribut tampak sering muncul dan faktanya terlihat bahwa ada tiga atribut yang dimilikinya dan berhubungan satu sama lain. Religious mungkin memiliki hubungan dengan Family dan Hobbies. Kita dapat menyelidiki kemungkinan hubungan ini lebih jauh dengan menambahkan satu hubungan terakhir operator ke model yaitu Association Rules.

6. Tambahkan Association Rules dan perhatikan hasilnya, set min confidence di 0.5



Parameters

Create Association Rules

criterion: confidence

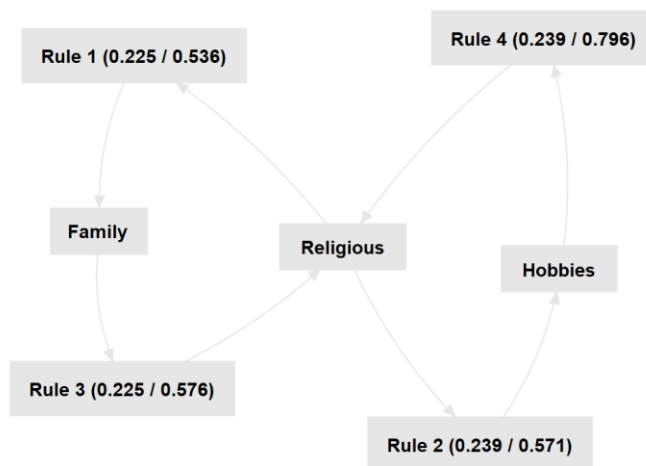
min confidence: 0.5

gain theta: 2.0

laplace k: 1.0

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift
2	Religious	Hobbies	0.239	0.571	0.873	-0.598	0.113	1.902
3	Family	Religious	0.225	0.576	0.881	-0.555	0.061	1.376
4	Hobbies	Religious	0.239	0.796	0.953	-0.361	0.113	1.902

7. Perhatikan poin nomor empat. Tertulis Confidence Level 80% yang ternyata secara aktual terukur 79,6%. Asosiasi yang lain memiliki persentase Confidence yang lebih rendah, namun masih cukup baik. Seandainya ditentukan threshold Confidence persen pada 0,55 (atau 55% persen), asosiasi 1 akan hildang dari hasil keluaran, jadi Family → Religious akan menjadi asosiasi tetapi Religious → Family tidak. Perhitungan lainnya dibagian kanan (LaPlace...Conviction) adalah indikator aritmatika tambahan dari hubungan Association Rules.



Salah satu dokumentasi video dari tutorial RapidMiner dapat dilihat dan dicoba juga untuk lebih memahami penggunaan Association Rules.

<https://academy.rapidminer.com/learn/video/association-rules-demo>

I. Latihan Praktikum

1. Pilih satu diantara data-data berikut dan lakukan Association Rules. Tuliskan kesimpulan dari hasil percobaan anda!
 - a. <https://www.kaggle.com/datasets/shazadudwadia/supermarket>
2. Cari hubungan yang menarik dari dataset **Restaurant.CSV** yang diberikan. Buat kesimpulan dari hasil percobaan anda. Contoh hubungan yang dimaksud “Ice Cream Gelato Yogurt Ices adalah jenis makanan yang banyak disajikan pada restoran grade A”

--- SELAMAT BELAJAR ---