

Association Rules: Cosmetic Recommendation

1) Analytic Problem

A drug store chain wants to learn more about cosmetics buyers purchase patterns. Specifically, they want to know what items are purchased in conjunction with each other, for purposes of display, point of sale special offers, and to eventually implement a real time recommender system to cross-sell items at time of purchase. The data are in the form of a matrix in which each column represents a product group, and each row a customer transaction.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
		Bag	Blush	Nail Polish	Brushes	Concealer	Eyebrow Pencils	Bronzer	Lip liner	Mascara	Eye shadow	Foundation	Lip Gloss	Lipstick	Eyeliner
Trans.	1	0	1	1	1	1	0	1	1	1	0	0	0	0	1
	2	0	0	1	0	1	0	1	1	0	0	1	1	0	0
	3	0	1	0	0	1	1	1	1	1	1	1	1	1	0
	4	0	0	1	1	1	0	1	0	0	0	1	0	0	1
	5	0	1	0	0	1	0	1	1	1	1	0	1	1	0
	6	0	0	0	0	1	0	0	0	0	0	0	0	0	1
	7	0	1	1	1	1	0	1	1	1	1	1	1	0	0
	8	0	0	1	1	0	0	1	0	1	1	0	1	0	0
	9	0	0	0	0	1	0	0	0	0	0	1	0	1	0
	10	1	1	1	1	0	0	0	0	1	1	0	0	0	0
	11	0	0	1	0	0	0	1	0	0	0	0	1	1	1
	12	0	0	1	1	1	0	1	0	1	1	1	0	0	0
	13	0	1	0	0	1	0	0	1	1	1	0	1	0	0
	14	0	1	1	1	1	0	1	1	1	1	1	1	1	0
	15	0	1	0	0	1	0	1	1	1	1	1	1	1	1

The dataset contains data on the purchases of different cosmetic items at a large chain drugstore. The store wants to analyze associations among purchases of these items for purpose of point-of-sale display, guidance to sales personnel in promoting cross sales.

2) Exploration

Use visualization technique to explore your dataset. Explain the meaning of each transaction.

3) Pre-processing

Pre-process your data to obtain input format for association rule discovery step.

4) Model construction using association rules

- Generate all the frequent itemsets with support = 20%
- What is the total number of frequent itemsets?
- Show all the maximal frequent itemsets.

- Generate all the association rules with minimum support = 20% and minimum confidence = 70%

5) **Model deployment**

- Show rule with highest confidence
- Explain the meaning of this rule
- Explain how support, confidence and lift are calculated?
- Show example of using these rules to increase sales.