

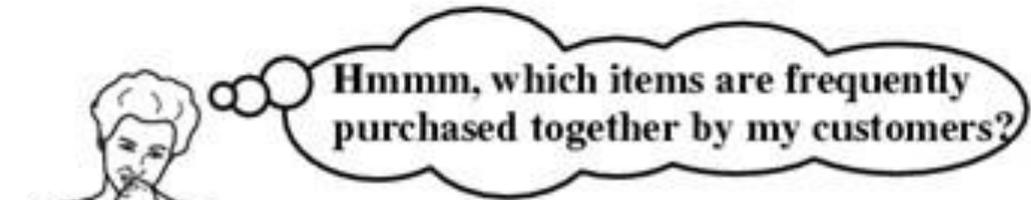
Association Rules

An association rule applied to your sales records may disclose that consumers who buy basmati rice and chicken masala powder also buy ghee (biryani combination).

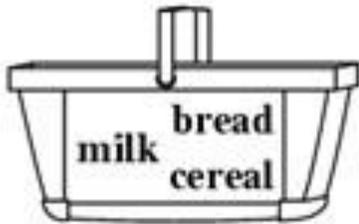
Market Basket Analysis

- One basket tells you about what one customer purchased at one time.
- A loyalty card makes it possible to tie together purchases by a single customer (or household) over time.

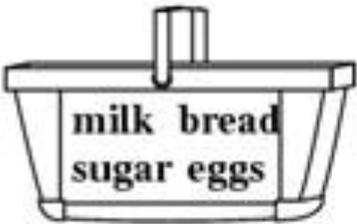




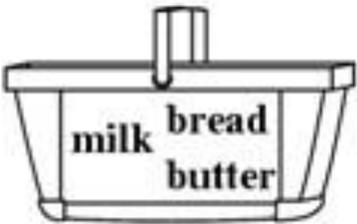
Shopping Baskets



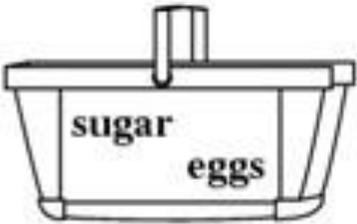
Customer 1



Customer 2

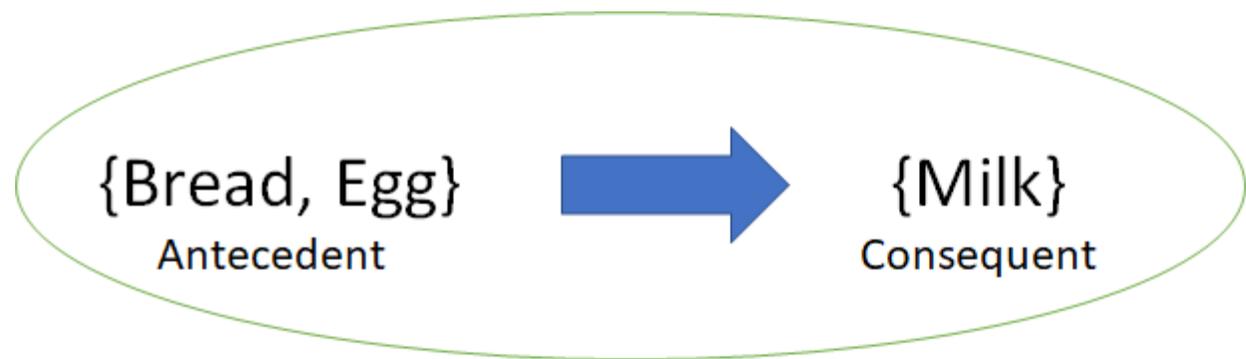


Customer 3



Customer n

Market Analyst



Itemset = {Bread, Egg, Milk}

- Rule form

Antecedent \rightarrow Consequent [**support**, **confidence**]

(support and confidence are user defined measures of interestingness)

- Examples

- $\text{buys}(x, \text{"computer"}) \rightarrow \text{buys}(x, \text{"financial management software"})$ [0.5%, 60%]
- $\text{age}(x, \text{"30..39"}) \wedge \text{income}(x, \text{"42..48K"}) \rightarrow \text{buys}(x, \text{"car"})$ [1%, 75%]

$$A \Rightarrow B [s, c]$$

Criterion for “**Support**”: *Support* of an item % (or number) of transactions in which antecedent (IF) and consequent (THEN) appear in the data

$$\frac{\# \text{ transactions with both antecedent \& consequent item sets}}{\# \text{ Total transactions}}$$

Confidence: % of antecedent (IF) transactions that also have the consequent (THEN) item set

$$\frac{\# \text{ transactions with both antecedent \& consequent item sets}}{\# \text{ transactions with antecedent item set}}$$

What is the support for “if white then blue”? (choose one or more)

1. 4

2. 40%

3. 2

4. 90%

Transaction #	Faceplate	colors	purchased
1	red	white	green
2	white	orange	
3	white	blue	
4	red	white	orange
5	red	blue	
6	white	blue	
7	white	orange	
8	red	white	blue green
9	red	white	blue
10	yellow		

What is the support for “if blue then white”? (choose one or more)

- 1. 4
- 2. 40%
- 3. 2
- 4. 90%

Transaction #	Faceplate	colors	purchased	
1	red	white	green	
2	white	orange		
3	white	blue		
4	red	white	orange	
5	red	blue		
6	white	blue		
7	white	orange		
8	red	white	blue	green
9	red	white	blue	
10	yellow			

Tr1	Shoes, <i>Socks</i> , Tie, Belt
Tr2	Shoes, <i>Socks</i> , Tie, Belt, Shirt, Hat
Tr3	Shoes, Tie
Tr4	Shoes, <i>Socks</i> , Belt

Transaction	Shoes	Socks	Tie	Belt	Shirt	Scarf	Hat
1	1	1	1		0	0	0
2	1	1	1	1	1	0	1
3	1	0	1	0	0	0	0
4	1	1	0	1	0	0	0
...							

What is the support and confidence for the Rule Socks => Tie ???

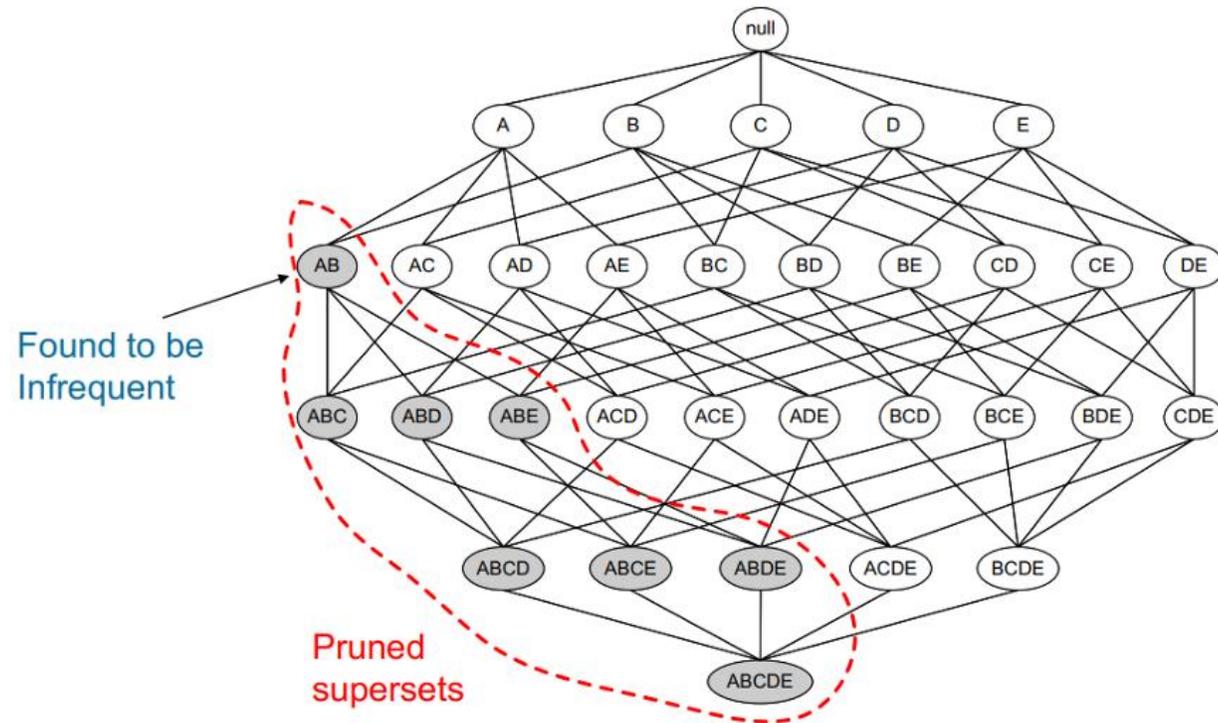
Tr1	Shoes, <i>Socks</i> , Tie, Belt
Tr2	Shoes, <i>Socks</i> , Tie, Belt, Shirt, Hat
Tr3	Shoes, Tie
Tr4	Shoes, <i>Socks</i> , Belt

Transaction	Shoes	Socks	Tie	Belt	Shirt	Scarf	Hat
1	1	1	1		0	0	0
2	1	1	1	1	1	0	1
3	1	0	1	0	0	0	0
4	1	1	0	1	0	0	0
...							

Socks \Rightarrow *Tie*

- Support is 50% (2/4)
- Confidence is 66.67% (2/3)

The Apriori algorithm, like the preceding example, detects the most frequent itemsets or elements in a transaction database and establishes association rules between the items. The method employs a “bottom-up” strategy,



How Apriori algorithm work?

The algorithm examines three crucial aspects while constructing association rules between components or items: support, confidence, and lift. Each of these elements is discussed below:

- 1.Association rule:** For example, $X \rightarrow Y$ is a depiction of discovering Y on a basket that contains X .
- 2.Itemset:** For example, $\{X, Y\}$ is a representation of the list of all objects that comprise the association rule.
- 3.Support:** Transactions containing the itemset as a percentage of total transactions
- 4.Confidence:** Given X , what is the likelihood of Y occurring?
- 5.Lift:** Confidence ratio to baseline likelihood of occurrence of Y

Apriori algorithm – Python library

Because the Apriori algorithm is not included in scikit learn, we must install it externally using the pip install apyori command.

Performance Measure : Lift ratio

Lift ratio = $\text{confidence} / (\text{benchmark confidence})$

Benchmark assumes independence between antecedent and consequent:

$$P(\text{antecedent} \ \& \ \text{consequent}) = P(\text{antecedent}) \times P(\text{consequent})$$

Benchmark confidence:

$$P(C | A) = P(C \ \& \ A) / P(A) = P(C) \times P(A) / P(A) = P(C)$$

=

$$\frac{\text{\# transactions with consequent item sets}}{\text{\# transactions in database}}$$

Interpreting Lift

Lift > 1 indicates a rule that is useful in finding consequent items sets (i.e., more useful than selecting transactions randomly)

Interpretation revisited

- *Lift ratio* shows how effective the rule is in finding consequents vs. random (useful if finding particular consequents is important)
- *Confidence* shows the rate at which consequents will be found (useful in learning costs of promotion)
- *Support* measures overall impact (% transactions affected)

Advantage



- Among association rule learning algorithms, this is the simplest and most straightforward algorithm.
- The resulting rules are simple to understand and express to the end-user.

Disadvantage



- It asserts that all subsets of a frequent itemset must also be frequent and that all supersets of an infrequent itemset must also be uncommon.
- And the main disadvantage is the slow process, to create candidates, the algorithm must continually search the database;
- this operation consumes a significant amount of time and memory, particularly if the pattern is very many and lengthy.

Application of Apriori algorithm



There are in many sectors, Apriori algorithms are used,

1. Education sector
2. Medical sector
3. Forest sector
4. E-Commerce sector
5. Office sector