Mining Data for Rules Underlying User Behavior

Swetha Kolalapudi
CO-FOUNDER, LOONYCORN

www.loonycorn.com

Overview

Understand what association rules are

Mine transaction data for association rules using the apriori algorithm

Implement the apriori algorithm on a bakery sales dataset

Recommendation Algorithms

Content Based Filtering

Find products with “similar” attributes

Collaborative Filtering

Find products liked by “similar” users

Association Rules Learning

Find “complementary” products

Association Rules Learning

What items are bought together in a transaction?

What items are bought by a user in a short period of time?

Market basket analysis

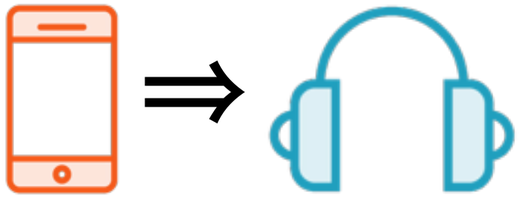
Association Rules Learning

Conditional Probabilities

P(/)

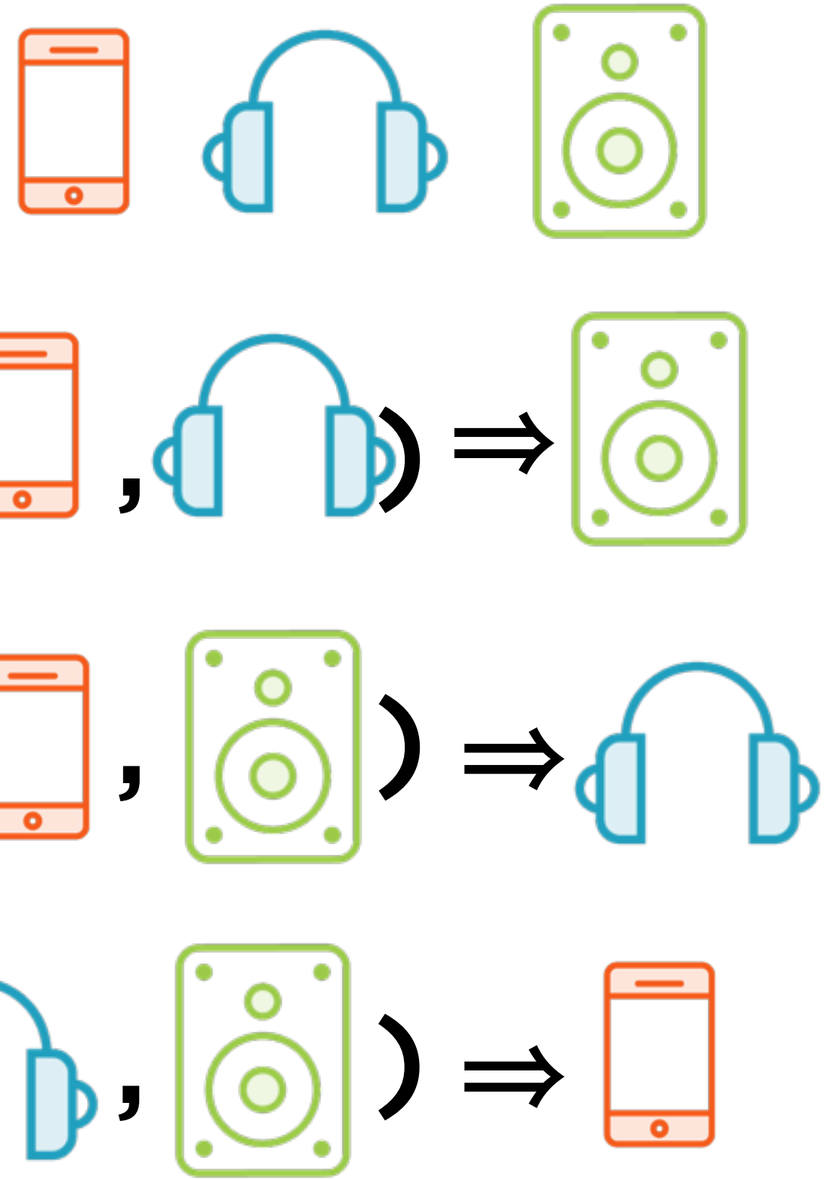
Association Rules Learning

Does a person buying  increase the likelihood of buying  ?

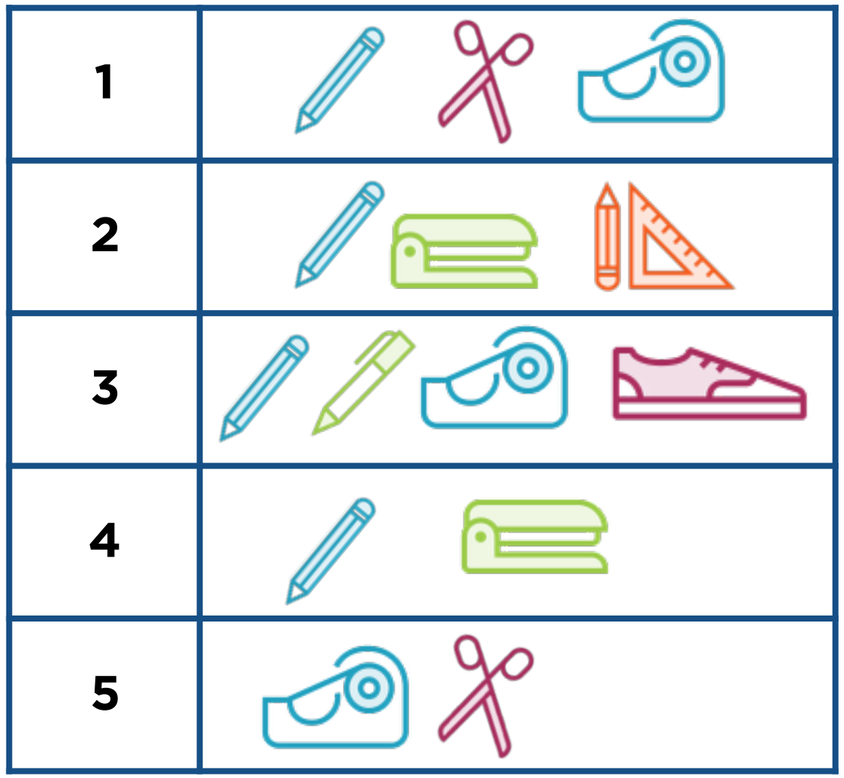


Association rule

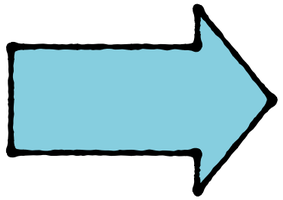
Association Rules Learning

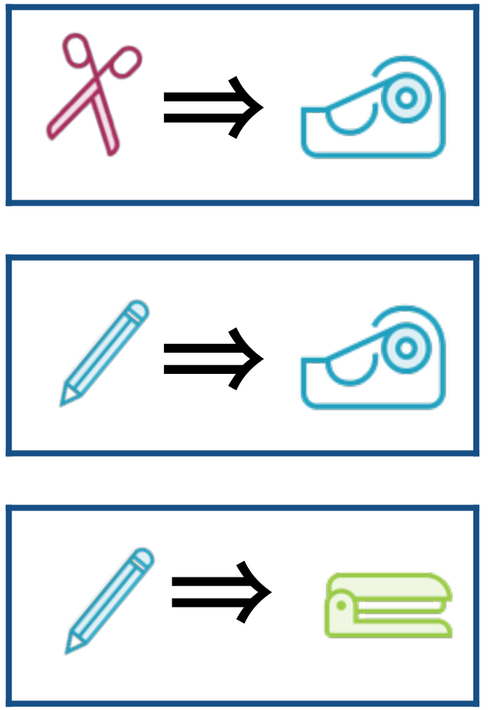


Mining for Association Rules



Transactions

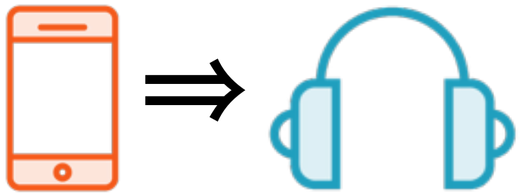


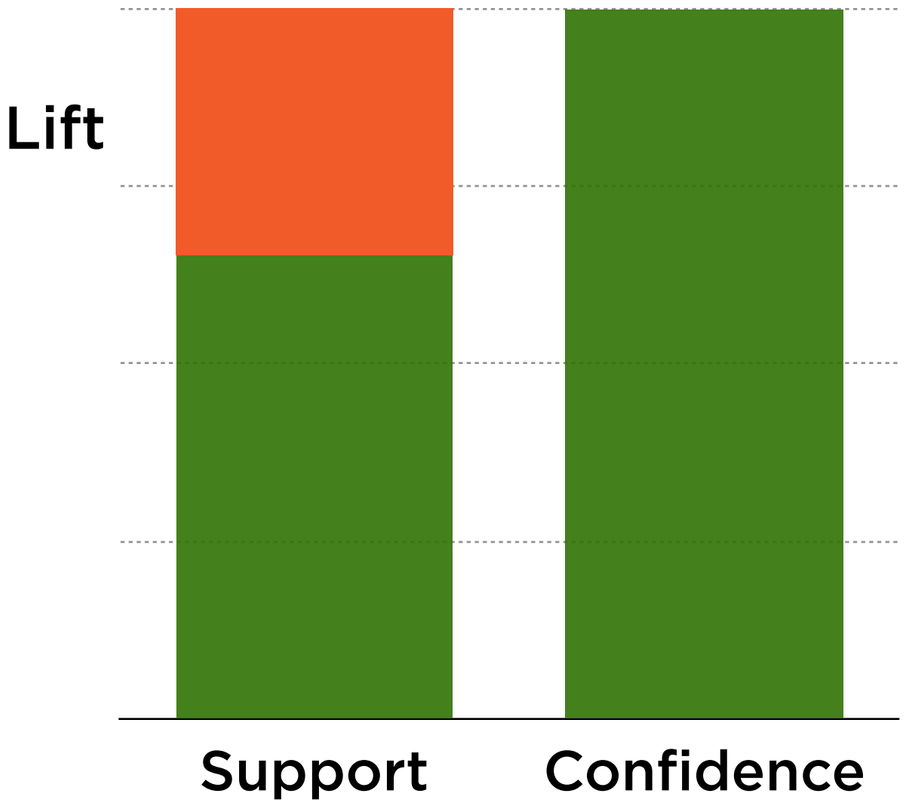


Rules

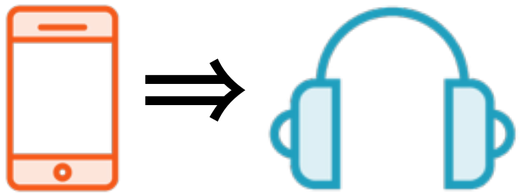
Measuring the Strength of a Rule

Measuring Rule Strength





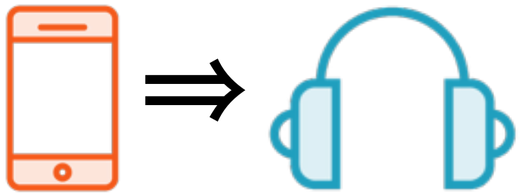
Measuring Rule Strength



What proportion of all transactions contain both items?

Support

Measuring Rule Strength

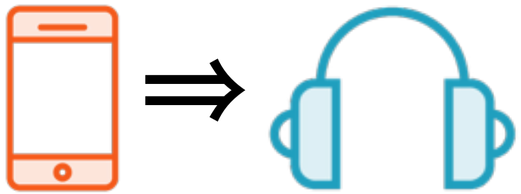


Out of all transactions with  how many include  ?

P( / ) = 5%

Confidence

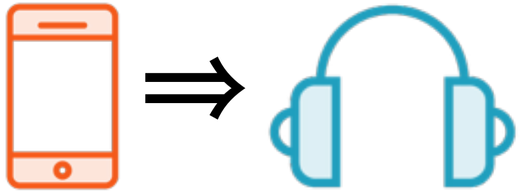
Measuring Rule Strength



$$P(\text{headphones} / \text{phone}) = \frac{P(\text{headphones}, \text{phone})}{P(\text{phone})}$$

Confidence

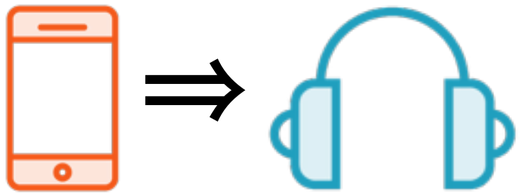
Measuring Rule Strength



$$P(\text{headphones} / \text{phone}) = \frac{Supp(\text{headphones}, \text{phone})}{Supp(\text{phone})}$$

Confidence

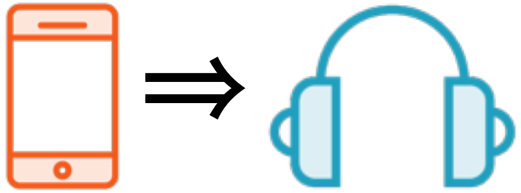
Measuring Rule Strength



How much does the likelihood of buying  increase when  is bought?

Lift

Measuring Rule Strength



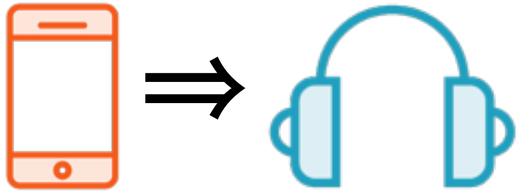
How much does the likelihood of buying  increase when  is bought?

$$P(\text{headphones}) = 3\%$$

$$P(\text{headphones} / \text{phone}) = 5\%$$

$$\text{Lift}(\text{phone} \Rightarrow \text{headphones}) = \frac{5\%}{3\%} = 1.67$$

Measuring Rule Strength

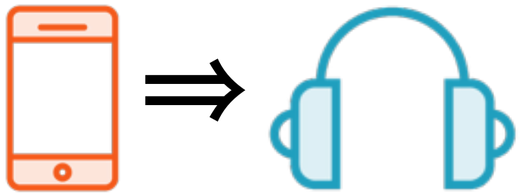


How much does the likelihood of buying  increase when  is bought?

$$\text{Lift}(\text{📱} \Rightarrow \text{🎧}) = \frac{P(\text{🎧}/\text{📱})}{P(\text{📱})}$$

$$= \frac{\text{Conf}(\text{🎧}/\text{📱})}{\text{Supp}(\text{📱})}$$

Measuring Rule Strength



Support

Likelihood of all items in the rule being in a single transaction

Confidence

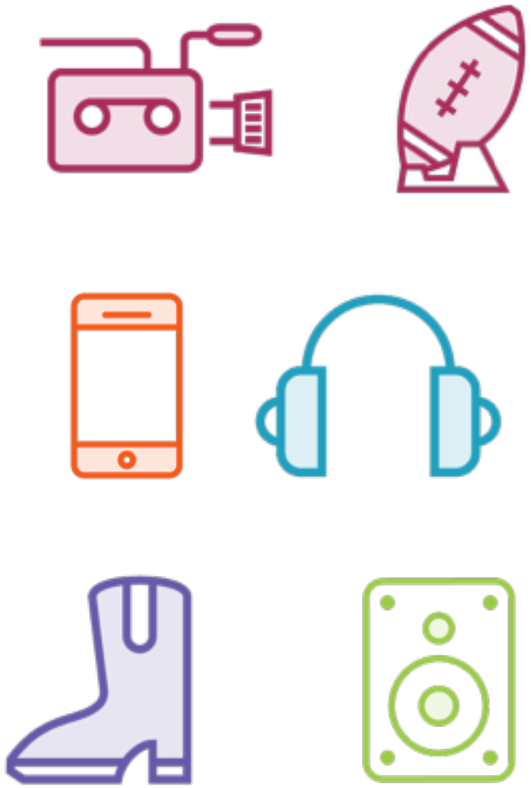
Likelihood of second item being bought once the first is bought

Lift

Change in likelihood of second item being bought once the first is bought

Mining for Rules Using the Apriori Algorithm

Mining for Association Rules



Catalog

Brute Force

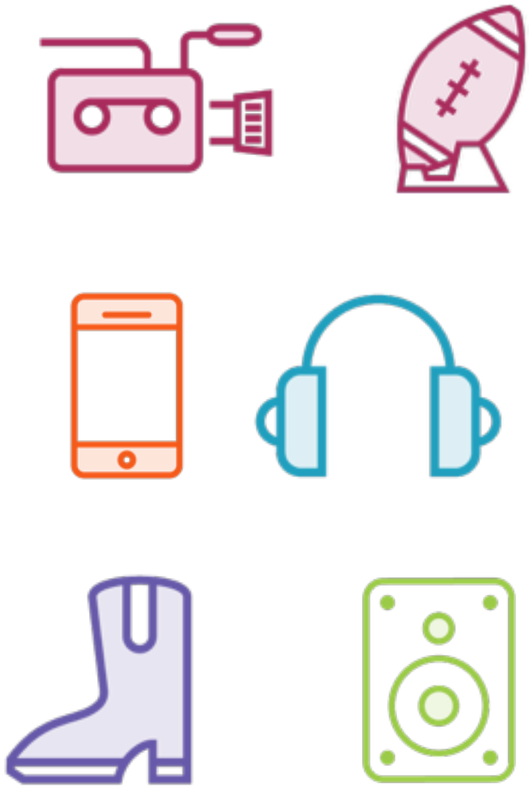
Find all possible 2 item sets x 2

Find all possible 3 item sets x 3

⋮

Find all possible N item sets x N

Mining for Association Rules



Catalog

Apriori Algorithm

Prune the number of items in each stage

Use metrics to check how important an item set is

Support

Confidence

Apriori Algorithm

Apriori Algorithm

Find 1 item sets

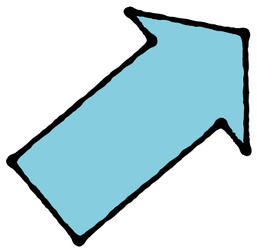
Keep only those with a minimum support

Apriori Algorithm

Find all possible 1 item sets

Compute the support of each set

1	
2	
3	
4	
5	



2/5	1/5	1/5	1/5	2/5	3/5	4/5

Apriori Algorithm

Find all possible 1 item sets

Compute the support of each set

2/5	1/5	1/5	1/5	2/5	3/5	4/5

Drop item sets with support below a minimum threshold

Apriori Algorithm

Find 1 item sets

Keep only those with a minimum support

Find 2 item sets

Use only the items left from previous step

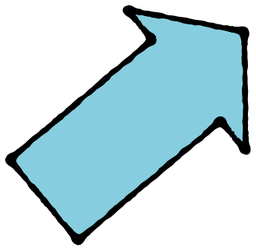
Keep only item sets with minimum support

Apriori Algorithm

Find all possible 2 item sets

Compute the support of each set

1	
2	
3	
4	
5	



0/5	1/5	2/5	2/5	0/5	2/5

Apriori Algorithm

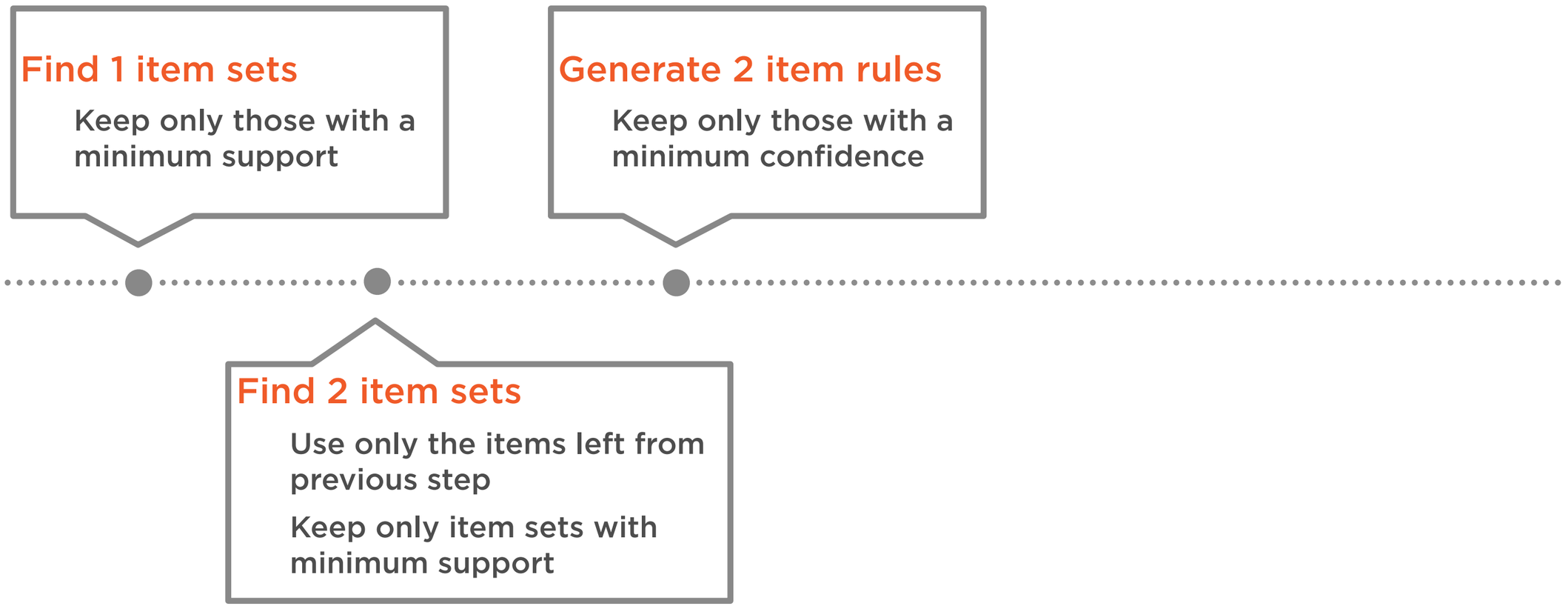
Find all possible 2 item sets

Compute the support of each set

0/5	1/5	2/5	2/5	0/5	2/5

Drop item sets with support below a minimum threshold

Apriori Algorithm



Apriori Algorithm

Find all possible 2 item sets

Compute the support of each set

0/5	1/5	2/5	2/5	0/5	2/5

Drop item sets with support below a minimum threshold

Apriori Algorithm

Drop item sets with support below a minimum threshold

From each item set, generate rules

Keep rules with a minimum confidence

1	
2	
3	
4	
5	

2/2

Apriori Algorithm

Drop item sets with support below a minimum threshold

From each item set, generate rules

Keep rules with a minimum confidence

1	
2	
3	
4	
5	

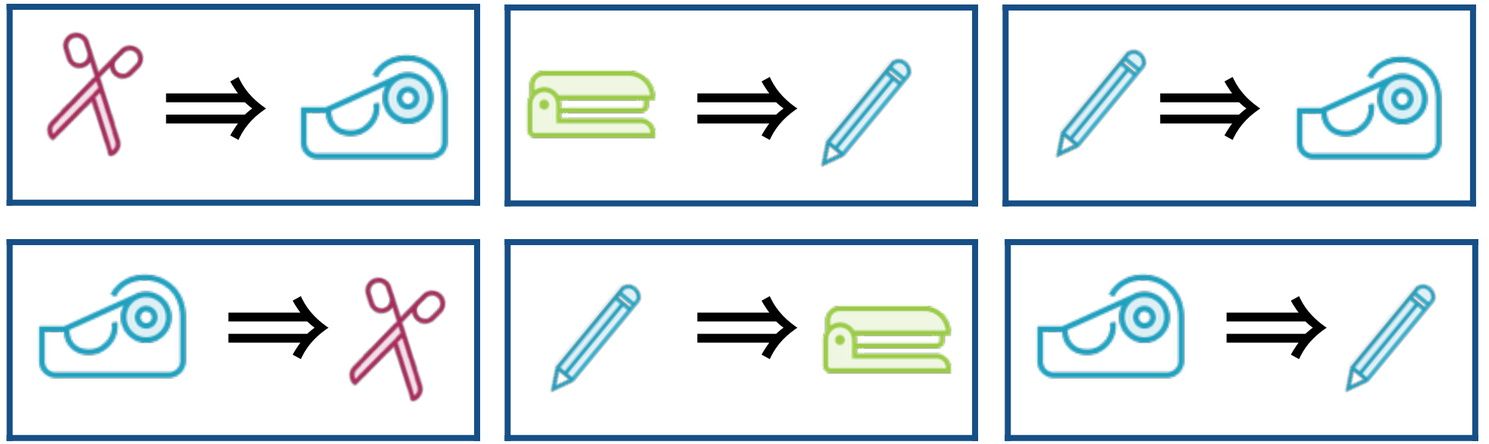
2/2	2/3

Apriori Algorithm

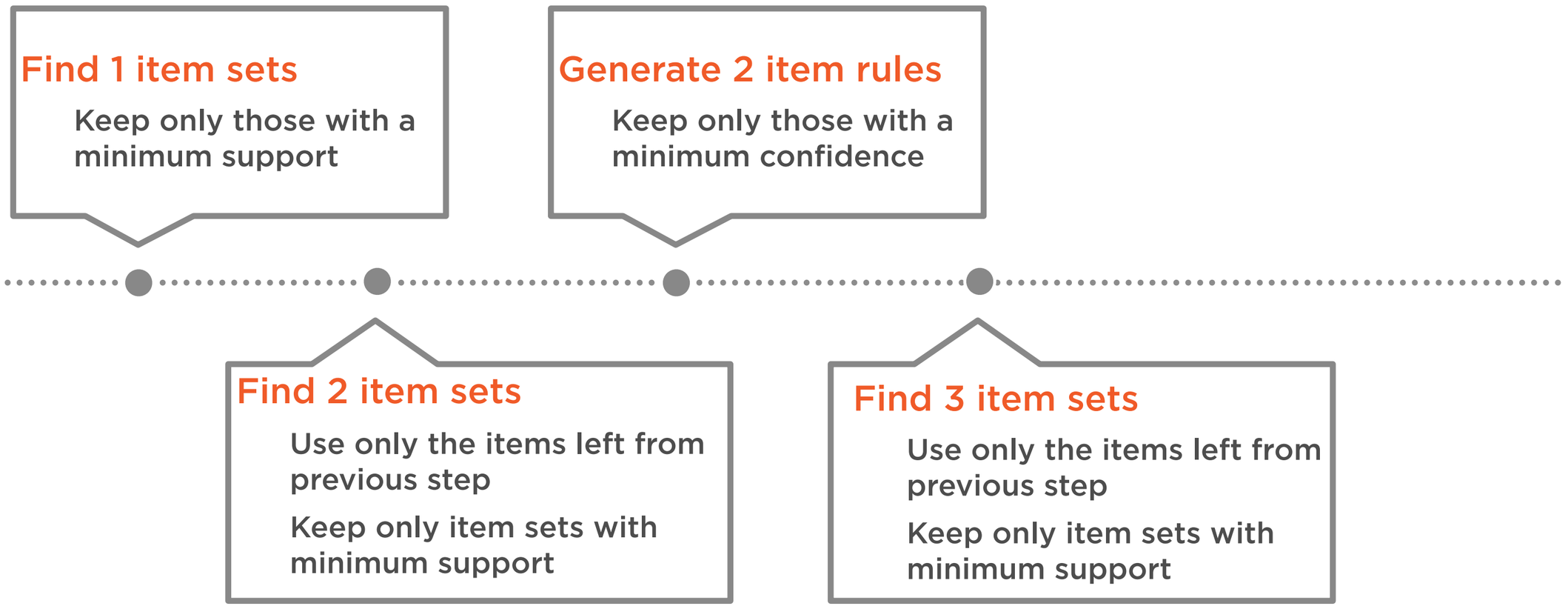
Drop item sets with support below a minimum threshold

From each item set, generate rules

Keep rules with a minimum confidence



Apriori Algorithm

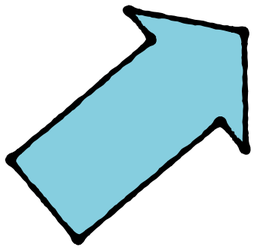


Apriori Algorithm

Find all possible 3 item sets

Compute the support of each set

1	
2	
3	
4	
5	



0/5	1/5	0/5

Apriori Algorithm

Find all possible 3 item sets

Compute the support of each set

0/5	1/5	0/5

Drop item sets with support below a minimum threshold

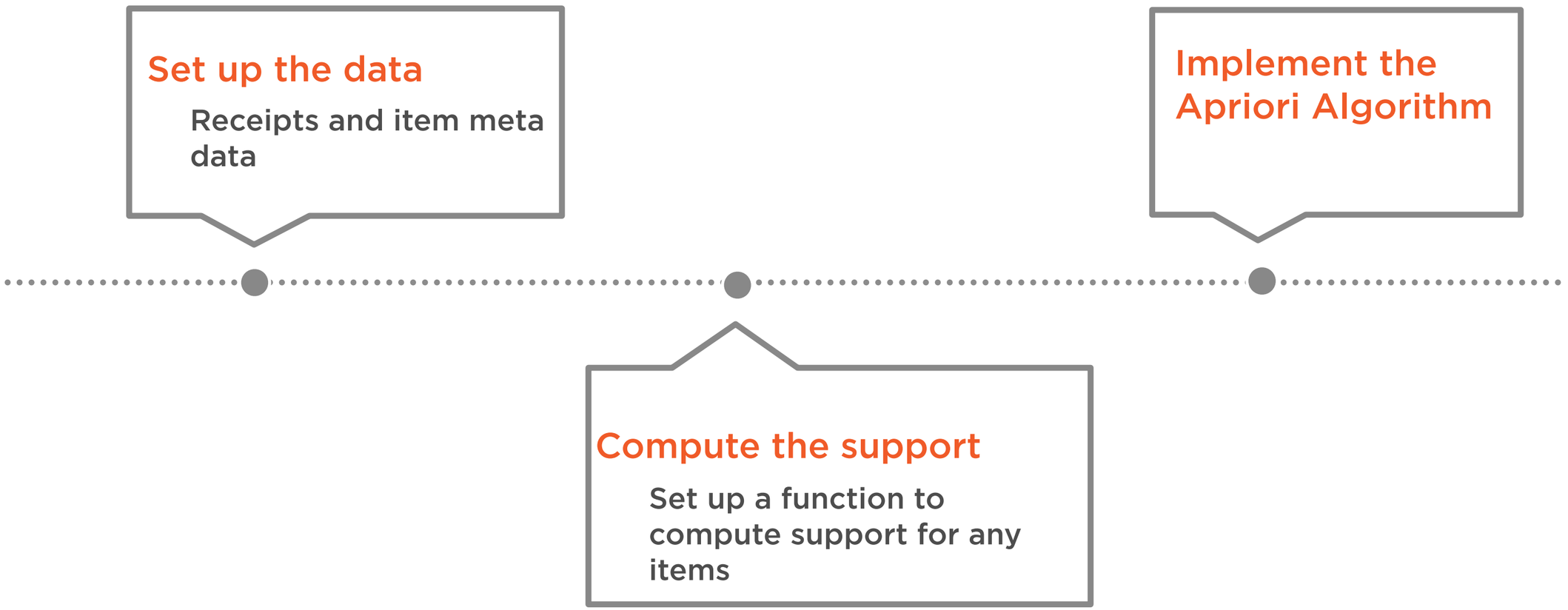
Algorithm stops

Apriori Algorithm

Demo

Find association rules in a bakery dataset

Association Rules for Bakery Items



Summary

Understand what association rules are

Mine transaction data for association rules using the apriori algorithm

Implement the apriori algorithm on a bakery sales dataset