



Sohail Akhtar
CS Department
Bahria University, Islamabad Campus

Association Rules

What are association rules?

Association rules are "if-then" statements, that help to show the probability of relationships between data items, within large data sets in various types of databases.

Association rule mining has a number of applications and is widely used to help discover sales correlations in transactional data or in medical data sets.

What are association rules?

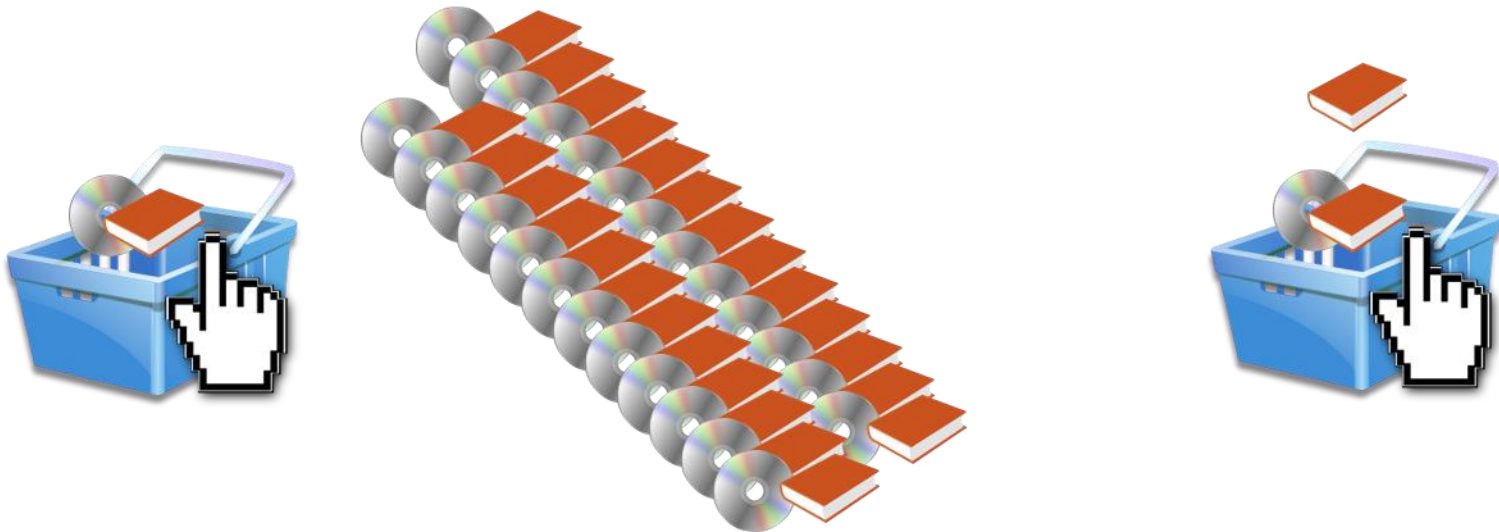
In data science, association rules are used to find correlations and co-occurrences between data sets.

They are ideally used to explain patterns in data from seemingly independent information repositories, such as relational databases and transactional databases. Based on the dependency, it then maps accordingly so that it can be more profitable.

The act of using association rules is sometimes referred to as "association rule mining" or "mining associations."

What are association rules?

Items already in basket + Available items \longrightarrow Item likely to be added



Use cases for association rules

- **Medicine**

By using association rules and machine learning data analysis, doctors can determine the conditional probability of a given illness by comparing symptom relationships in the data from past cases. As new diagnoses get made, machine learning models can adapt the rules to reflect the updated data.

- **Retail**

Retailers can collect data about purchasing patterns, recording purchase data as item barcodes are scanned by point-of-sale systems. Machine learning models can look for co-occurrence in this data to determine which products are most likely to be purchased together. The retailer can then adjust marketing and sales strategy to take advantage of this information.

- **Entertainment**

Services like Netflix and Spotify can use association rules to fuel their content recommendation engines. Machine learning models analyse past user behaviour data for frequent patterns, develop association rules and use those rules to recommend content that a user is likely to engage in future.

How association rules work

Association rule mining, at a basic level, involves the use of machine learning models to analyse data for patterns, or co-occurrences, in a database. It identifies frequent if-then associations, which themselves are the *association rules*.

An association rule has two parts:

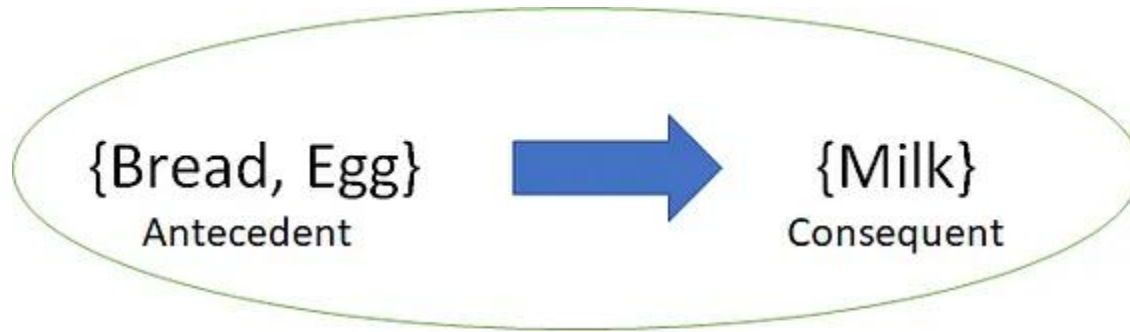
Antecedent (if): An antecedent is an item found within the data.

Consequent (then): A consequent is an item found in combination with the antecedent.

How association rules work

Lets now see what an association rule exactly looks like.

Note that implication here is co-occurrence and not causality. For a given rule, **itemset** is the list of all the items in the antecedent and the consequent.



Itemset = {Bread, Egg, Milk}

How association rules work

Various metrics are in place to help us understand the strength of association between these two. Let us go through them all.

1. Support

This measure gives an idea of how frequent an *itemset* is in all the transactions.

- Consider itemset1 = {bread} and itemset2 = {shampoo}. There will be far more transactions containing bread than those containing shampoo. So as you rightly guessed, itemset1 will generally have a higher support than itemset2.
- Now consider itemset1 = {bread, butter} and itemset2 = {bread, shampoo}. Many transactions will have both bread and butter on the cart but bread and shampoo? Not so much. So in this case, itemset1 will generally have a higher support than itemset2.

Mathematically, support is the fraction of the total number of transactions in which the itemset occurs.

$$\text{Support}(\{X\} \rightarrow \{Y\}) = \frac{\text{Transactions containing both } X \text{ and } Y}{\text{Total number of transactions}}$$

How association rules work

1. Support

$$\text{Support}(\{X\} \rightarrow \{Y\}) = \frac{\text{Transactions containing both } X \text{ and } Y}{\text{Total number of transactions}}$$

Value of support helps us identify the rules worth considering for further analysis.

For example, one might want to consider only the itemsets which occur at least 50 times out of a total of 10,000 transactions i.e. support = 0.005. If an itemset happens to have a very low support, we do not have enough information on the relationship between its items and hence no conclusions can be drawn from such a rule.

How association rules work

2. Confidence

This measure defines the likeliness of occurrence of consequent on the cart given that the cart already has the antecedents. That is to answer the question — of all the transactions containing say, {Tea Sugar}, how many also had {Milk} on them?

We can say by common knowledge that {Tea Sugar} \rightarrow {Milk} should be a high confidence rule.

Technically, confidence is the conditional probability of occurrence of consequent given the antecedent.

$$\text{Confidence}(\{X\} \rightarrow \{Y\}) = \frac{\text{Transactions containing both } X \text{ and } Y}{\text{Transactions containing } X}$$

How association rules work

2. Confidence

$$\text{Confidence}(\{X\} \rightarrow \{Y\}) = \frac{\text{Transactions containing both } X \text{ and } Y}{\text{Transactions containing } X}$$

Let us consider few more examples before moving ahead.

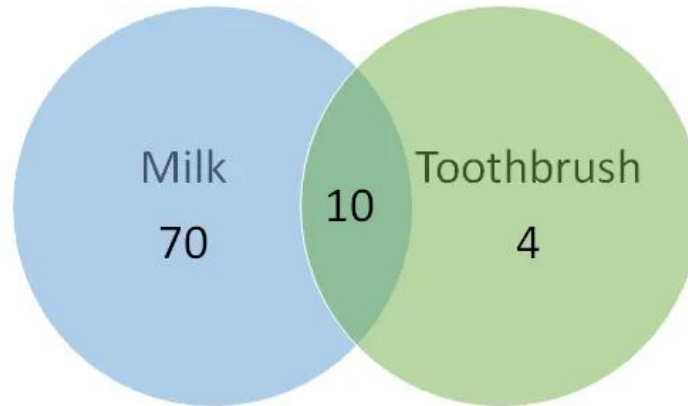
What do you think would be the confidence for $\{\text{Butter}\} \rightarrow \{\text{Bread}\}$? That is, what fraction of transactions having butter also had bread? Very high i.e. a value close to 1?

That's right. What about $\{\text{Yogurt}\} \rightarrow \{\text{Milk}\}$? High again. $\{\text{Toothbrush}\} \rightarrow \{\text{Milk}\}$? Not so sure?

Confidence for this rule will also be high since $\{\text{Milk}\}$ is such a frequent itemset and would be present in every other transaction.

Note: It does not matter what you have in the antecedent for such a frequent consequent. The confidence for an association rule having a very frequent consequent will always be high.

How association rules work



Total transactions = 100. 10 of them have both milk and toothbrush, 70 have milk but no toothbrush and 4 have toothbrush but no milk.

Consider the numbers from figure on the left. Confidence for $\{\text{Toothbrush}\} \rightarrow \{\text{Milk}\}$ will be $10/(10+4) = 0.7$

Looks like a high confidence value. But we know intuitively that these two products have a weak association and there is something misleading about this high confidence value. Lift is introduced to overcome this challenge.

Note: Considering just the value of confidence limits our capability to make any business inference.

How association rules work

3. Lift

Lift is a measure of the strength of the association between two items, taking into account the frequency of both items in the dataset. It is calculated as the confidence of the association divided by the support of the second item. Lift is used to compare the strength of the association between two items to the expected strength of the association if the items were independent.

$$\text{Lift}(\{X\} \rightarrow \{Y\}) = \frac{(\text{Transactions containing both } X \text{ and } Y) / (\text{Transactions containing } X)}{\text{Fraction of transactions containing } Y}$$

How association rules work

3. Lift

$$\text{Lift}(\{X\} \rightarrow \{Y\}) = \frac{(\text{Transactions containing both } X \text{ and } Y) / (\text{Transactions containing } X)}{\text{Fraction of transactions containing } Y}$$

Lift controls for the support (frequency) of consequent while calculating the conditional probability of occurrence of {Y} given {X}.

Lift is a very literal term given to this measure. Think of it as the *lift* that {X} provides to our confidence for having {Y} on the cart. To rephrase, lift is the rise in probability of having {Y} on the cart with the knowledge of {X} being present over the probability of having {Y} on the cart without any knowledge about presence of {X}.

Mathematically,

Lift in Association Rules

As for interpretation,

- If **lift=1**, A and B are **independent** (according to probability theory, $\text{support}(A \cup B) = \text{support}(A) \times \text{support}(B)$ given A and B are independent).
- If **lift > 1**, A and B are **positively correlated**. Namely, the 2 item sets lift the likelihood of each another.
- If **lift < 1**, A and B are **negatively correlated**. Using the supermarket example, the customers tend **NOT** to buy A and B together.

How association rules work

An antecedent is something that's found in data, and a consequent is an item that is found in combination with the antecedent. Have a look at this rule for instance:

“If a customer buys bread, he’s 70% likely of buying milk.”

In the above association rule, bread is the antecedent and milk is the consequent. Simply put, it can be understood as a retail store’s association rule to target their customers better.

If the above rule is a result of a thorough analysis of some data sets, it can be used to not only improve customer service but also improve the company’s revenue.

The Market-Basket Model

- A large set of *items*
 - e.g., things sold in a supermarket
- A large set of *baskets*, each of which is a small set of the items
 - e.g., the things one customer buys on one day
- Can be used to model any many-many relationship, not just in the retail setting
- Find “interesting” connections between items

Applications – (1)

- **Items** = products; **baskets** = sets of products someone bought in one trip to the store
- Suppose many people buy milk and diapers together
 - Run a sale on diapers; raise price of milk
- Only useful if many buy diapers & milk

Applications – (2)

- **Baskets** = sentences; **items** = documents containing those sentences
 - Items that appear together too often could represent plagiarism
 - Notice items do not have to be “in” baskets
- **Baskets** = Web pages; **items** = words.
 - Co-occurrence of relatively rare words may indicate an interesting relationship

Applications – (3)

- **Baskets** = patients; **items** = drugs and side-effects
- **Baskets** = movies; **items** = Oscar nominations and wins in different categories
 - Does being nominated in certain categories boost win likelihood in other categories?

Interesting Association Rules

- Not all high-confidence rules are interesting
 - The rule $X \rightarrow \text{milk}$ may have high confidence for many itemsets X , because milk is just purchased very often (independent of X)
- Interesting rules are those with high positive or negative interest values

An Example:

Suppose an X store's retail transactions database includes the following data:

- Total number of transactions: 600,000
- Transactions containing diapers: 7,500 (1.25 percent)
- Transactions containing milk: 60,000 (10 percent)
- Transactions containing both milk and diapers: 6,000 (1.0 percent)

From the above figures, we can conclude that if there was no relation between milk and diapers (that is, they were statistically independent), then we would have got only 10% of diaper purchasers to buy milk too.

However, as surprising as it may seem, the figures tell us that **80% (=6000/7500) of the people who buy diapers also buy milk.**

This is a significant jump of 8 over what was the expected probability. This factor of increase is known as Lift – which is the ratio of the observed frequency of co-occurrence of our items and the expected frequency.

Another Example:

There are 5 items in our shop including bread, butter, jam, milk, and egg. There are 5 transactions (baskets) in the database.

Basket 1 	    
Basket 2 	  
Basket 3 	  
Basket 4 	 
Basket 5 	 

Another Example:

The computer will translate to a matrix like this.

	Bread	Butter	Jam	Milk	Egg
Basket 1	1	1	1	1	1
Basket 2	1	1		1	
Basket 3	1	1			1
Basket 4	1	1			
Basket 5				1	1

From the transaction dataset, we will calculate the metrics that can guide us to business actions. The metrics are: support, confidence, and lift.

Another Example:

1. Support

How many transaction contain this itemset?

Support gives an idea of how frequent an itemset is in all the transactions.

The support value helps us identify the rules worth considering for further analysis.

$\text{freq}(X)$ is the number of occurrences of X in all transactions.


























N is the total number of transactions.

$$\text{Support}(X) = \frac{\text{freq}(X)}{N}$$

Ex. Support of bread = $4/5 = 0.8$

These tables show the support of all itemset with size 1–2. You can try calculating along.

Another Example:

	Support		Support		Support
	0.8	 	0.8	 	0.4
	0.8	 	0.2	 	0.4
	0.2	 	0.4	 	0.2
	0.6	 	0.4	 	0.2
	0.6	 	0.2	 	0.4

High support means that most customers buy this item, so this item is important for our shop.

We can filter out itemsets with low support since the number of occurrences is too low to find any insight (rule).

Another Example:

2. Confidence

Considering only transaction containing item A. How many transaction contain item B?

Confidence is a conditional probability . It defines the probability of the occurrence of the following item(s) in the same transaction given some item(s) (antecedents) are already in that transaction.

$$Confidence(A \rightarrow B) = \frac{freq(A, B)}{freq(A)}$$











Ex.

Confidence(Bread \rightarrow Milk) = $2/4 = 0.5$

Confidence(Milk \rightarrow Jam) = $1/3 = 0.33$

Another Example:

The below tables show the confidence of some rules.

	Confidence		Confidence
	1		0.5
	0.25		0.5
	0.5		1
	0.5		1
	0.25		0.67

High confidence means that the basket containing A will likely contain B also.

However, considering only the value of confidence is not enough to make any business decision. If B (following itemset) is a very frequent itemset, confidence related to B will always be high. Lift is another metric to be considered together with confidence.

Another Example:

3. Lift

With and without item A is in the transaction, how much it affect item B?

Lift is the ratio of the probability of occurrence of B given A is present and the probability of B occurrence without knowing about A.

$$\text{Lift}(A \rightarrow B) = \frac{\text{Support}(A, B)}{\text{Support}(A) \times \text{Support}(B)} = \frac{\text{Confidence}(A \rightarrow B)}{\text{Support}(B)}$$











Ex.

$\text{Lift}(\text{Bread} \rightarrow \text{Milk}) = 0.5/0.4 = 1.25$

$\text{Lift}(\text{Milk} \rightarrow \text{Jam}) = 0.33/0.2 = 1.65$

Another Example:

The below tables show the lift of some rules.

	Lift		Lift
	1.25		0.83
	1.25		0.83
	0.83		1.67
	0.83		1.67
	1.25		1.11

If Lift is higher than 1, the presence of A in this transaction causes a higher probability of B in the same transaction.

Goals of Association Mining

Goal of Association Rule Mining

When you apply Association Rule Mining on a given set of transactions T your goal will be to find all rules with:

1. Support greater than or equal to min_support
2. Confidence greater than or equal to min_confidence

APRIORI Algorithm

Association Rule Mining is viewed as a two-step approach:

1. Frequent Itemset Generation: Find all frequent item-sets with support \geq pre-determined min_support count

2. Rule Generation: List all Association Rules from frequent item-sets. Calculate Support and Confidence for all rules. Prune rules that fail min_support and min_confidence thresholds.

APRIORI Algorithm: Frequent Items Generation

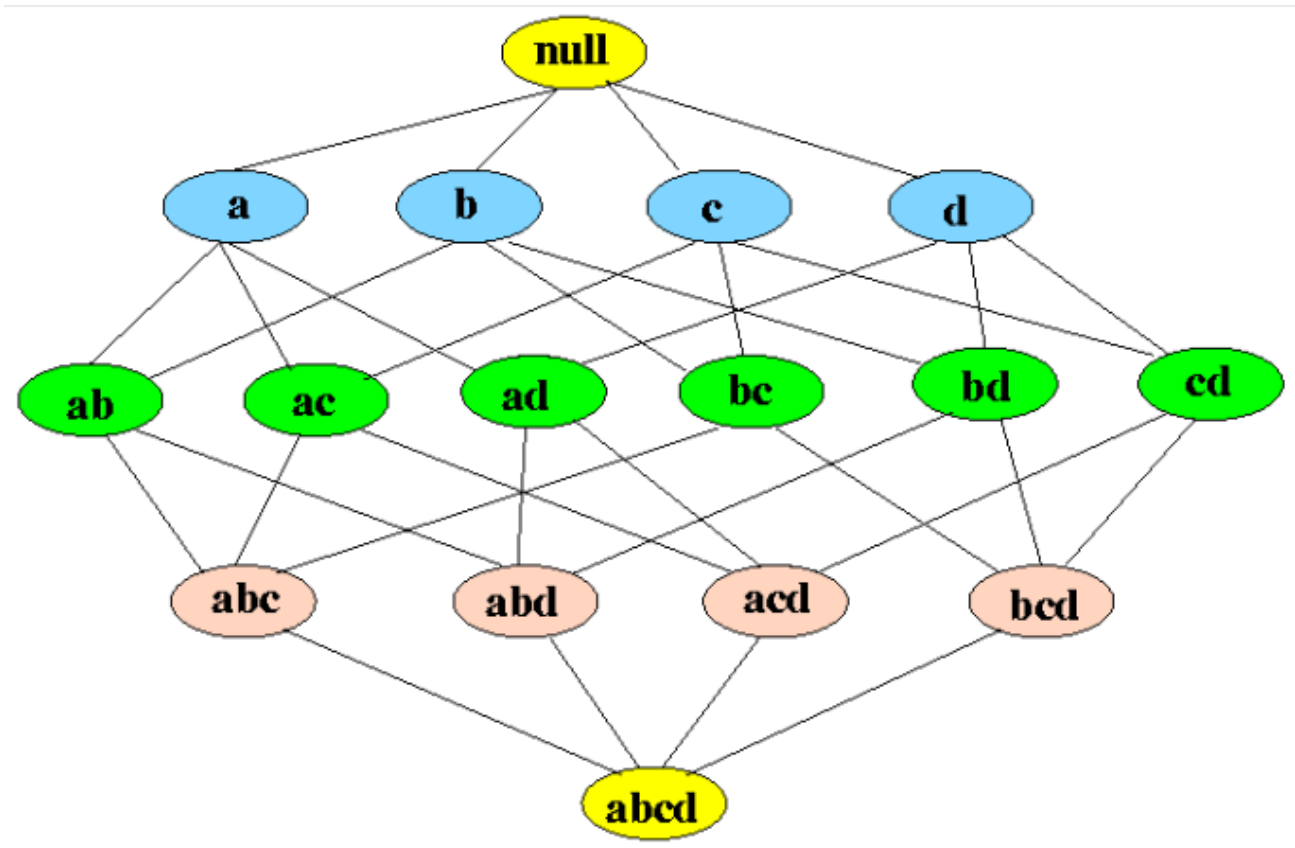
Frequent Itemset Generation is the most computationally expensive step because it requires a full database scan.

In real-world transaction data for retail can exceed up to GB s and TBs of data for which an optimized algorithm is needed to prune out Item-sets that will not help in later steps. For this, APRIORI Algorithm is used. It states:

Any subset of a frequent itemset must also be frequent. In other words, No superset of an infrequent itemset must be generated or tested.

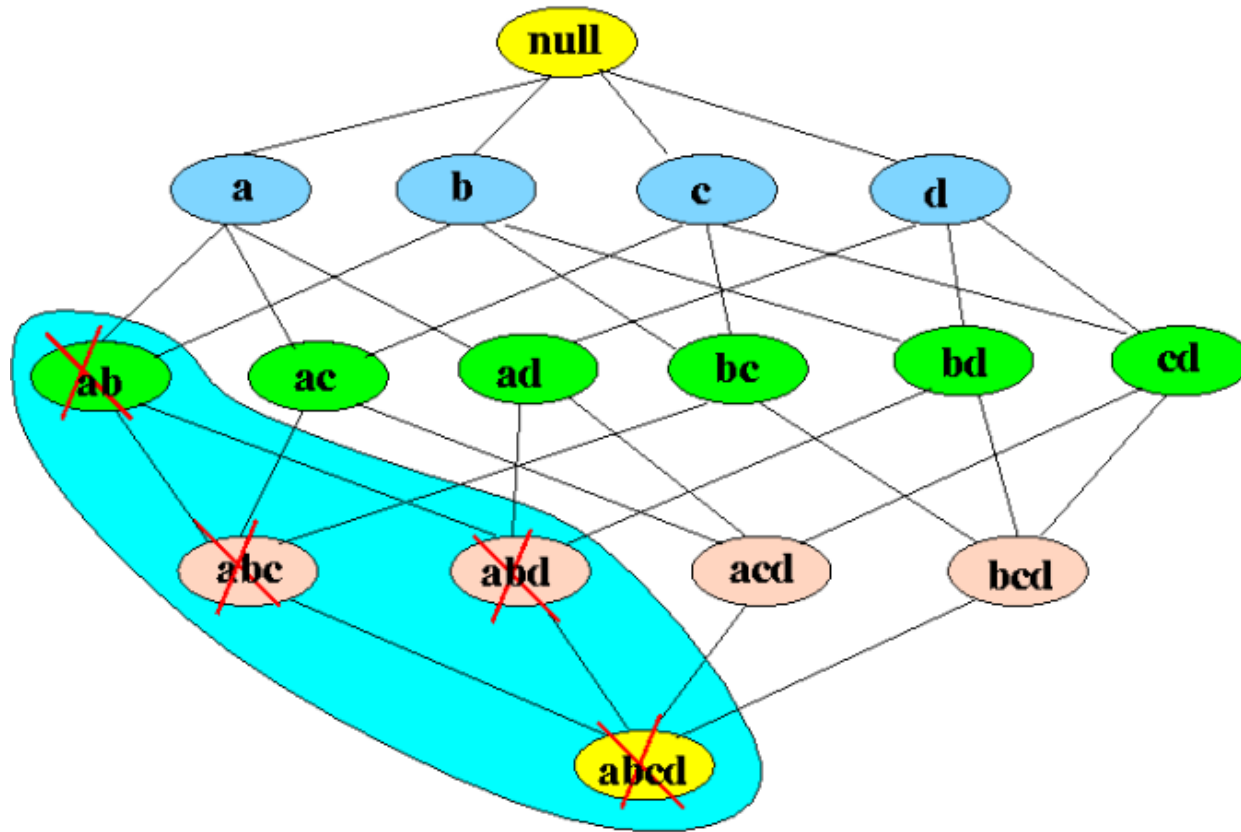
There is an itemset graphical representation of the APRIORI algorithm principle. It consists of k-item-set node and relation of subsets of that k-item-set.

APRIORI Algorithm



You can see in above figure that in the bottom is all the items in the transaction data and then you start moving upwards creating subsets till the null set. The following figure shows how much APRIORI helps to reduce the number of sets to be generated:

APRIORI Algorithm



If item-set $\{a,b\}$ is infrequent then we do not need to take into account all its super-sets.

APRIORI Algorithm: Example

In the following example, you will see why APRIORI is an effective algorithm and also generate strong association rules step by step.



As you can see, you start by creating *Candidate List* for the 1-itemset that will include all the items, which are present in the transaction data, individually. Considering retail transaction data from real-world, you can see how expensive this candidate generation is. Here APRIORI plays its role and helps reduce the number of the Candidate list, and useful rules are generated at the end. In the following steps, you will see how we reach the end of Frequent Itemset generation, that is the first step of Association rule mining.

APRIORI Algorithm:Example

C2= all possible 2-itemset combinations

Itemset	Support Count
I1,I2	2
I1,I3	1
I2,I3	3
I3,I1	1

3. Generate second Candidate list by L1 cross join L1. And note support counts. {I1,I2} appear in 2 transactions together.

L2

Itemset	Support Count
I1,I2	2
I2,I3	3

4. Remove candidates that fail min_sup count.

C3= all possible 2-itemset combinations

Itemset	Support Count
I1,I2,I3	1

5. Generate third Candidate list by L2 cross join L2. And note support counts. {I1,I2,I3} appear in only 1 transactions together.

6. L3 is null. L3={} since Support count for {I1,I2,I3} fails min_sup. Here First step of Association rule mining is completed and there will be no C4 candidate list

Transaction ID	Items
100	I1,I2
200	I2,I3,I4,I5
300	I2,I3
400	I1
500	I1,I2,I3

Your next step will be to list all frequent itemsets. You will take the last non-empty Frequent Itemset, which in this example is **L2={I1, I2},{I2, I3}**. Then make all non-empty subsets of the item-sets present in that Frequent Item-set List. Follow along as shown in below illustration:

APRIORI Algorithm: Example



Given $\text{Min_confidence} = 50\%$. Confidence is calculated by:

$$C(A \Rightarrow B) = P(A \cup B) / P(A) = n(A \cup B) / n(A)$$

Confidence is number of times A and B are together in all transactions containing A

Transaction ID	Items
100	I1,I2
200	I2,I3,I4,I5
300	I2,I3
400	I1
500	I1,I2,I3

You can see above there are four strong rules. For example, take $\{I2\} \Rightarrow I3$ having confidence equal to 75% tells that 75% of people who bought **I2** also bought **I3**.