



Association Rule Mining

Learning Association Rule Mining made easy!

- [HOME](#)
- [ASSOCIATION](#)
- [APRIORI](#)
- [FP GROWTH](#)
- [RULE GENERATION](#)
- [VISUALIZATION](#)
- [FAQ](#)
- [REFERENCES](#)

Association Rule Mining Algorithms ?

- [Apriori Algorithm.](#)
- [FP Growth Algorithm.](#)

What is Apriori Algorithm ?

It is a classic algorithm used in data mining for finding association rules based on the principle "Any subset of a large item set must be large". It uses a generate-and-test approach – generates candidate itemsets and tests if they are frequent.

Frequent Itemset Generation:

Given the [minimum threshold support](#), Generating large item sets (only keep [frequent item sets](#) – large item sets with enough support).

Illustration:

Consider the below transaction in which B = Bread, J = Jelly, P = Peanut Butter, M = Milk and E = Eggs. Given that minimum threshold support = 40% and minimum threshold confidence = 80% [13].

TID	Items Bought
1	{ B, J, P }
2	{ B, P }
3	{ B, M, P }
4	{ E, B }
5	{ E, M }

Step-1: Count the number of transactions in which each item occurs (Bread B occurs in 4 transactions and so on).

Items	Support	Support (in percentage) = (Support * 100) / No. of trans
B	4	$(4 * 100) / 5 = 80\%$
J	1	$(1 * 100) / 5 = 20\%$
P	3	$(3 * 100) / 5 = 60\%$
M	2	$(2 * 100) / 5 = 40\%$
E	2	$(2 * 100) / 5 = 40\%$

Step-2: As minimum threshold support = 40%, So in this step we will remove all the items that are bought less than 40% of support or support less than 2.

Items	Support
B	4 (80%)
P	3 (60%)
M	2 (40%)
E	2 (40%)

The above table has single items that are bought frequently. Now let's find a pair of items that are bought frequently. We continue from the above table (Table in step 2)

Step-3: We start making pairs from the first item and below items like {B,P} , {B,M} , {B,E} and then we start with the second item and below items like {P,M} , {P,E}. We do not make pair {P,B} because we already made {P,B} pair when we were making pairs of B. As buying a bread and Peanut Butter together is same as buying Peanut Butter and bread together. After making all the pairs we get,

Items	Support	Support (in percentage) = (Support * 100) / No. of trans
{B, P}	3	$(3 * 100) / 5 = 60\%$
{B, M}	1	$(1 * 100) / 5 = 20\%$
{B, E}	1	$(1 * 100) / 5 = 20\%$
{P, M}	1	$(1 * 100) / 5 = 20\%$
{P, E}	0	$(0 * 100) / 5 = 0\%$
{M, E}	1	$(1 * 100) / 5 = 20\%$

Step-4: As minimum threshold support = 40%, So in this step we will remove all the items that are bought less than 40% of support and we are left with

Items	Support
{B, P}	3

The above table has two items {B, P} that are bought together frequently.

Association Rule Generation :

Step-5: As we cannot generate large frequent item (itemset of 3) further because we are left with 1 frequent item set. We will start generating association rules from the frequent item set. As we have frequent item set of two, only two association rules will be generated which is shown below :

Items	Association Rule	Confidence $(X \Rightarrow Y) =$ Occurrence (X U Y) / Occurrence (X)
{B, P} (60%)	B \Rightarrow P	$(3/4) * 100 = 75\%$
	P \Rightarrow B	$(3/3) * 100 = 100\%$

As P \Rightarrow B has confidence 100% which is greater than minimum confidence threshold 80%, thus P \Rightarrow B is a **Strong Association Rule**.

[Try Examples](#)

Disadvantages of Apriori Algorithm ?

- Generation of itemsets is expensive(in both space and time)
- Support counting is expensive