

Apriori Algorithm Tutorial

The **Apriori Algorithm** is used in **association rule mining** to discover frequent itemsets in large datasets. It is widely applied in market basket analysis to find relationships between items purchased together.

Step 1: Define the Dataset

Consider the following transaction dataset:

Transaction ID	Items Purchased
T1	Milk, Bread, Butter
T2	Bread, Butter
T3	Milk, Bread
T4	Milk, Butter
T5	Bread, Butter
T6	Milk, Bread, Butter

Step 2: Set Minimum Support and Confidence

- **Support:** The proportion of transactions that contain an itemset.
- **Confidence:** The likelihood that item **B** is also bought when item **A** is bought.

For this example, let's set:

- **Minimum Support = 50% (3 transactions)**
- **Minimum Confidence = 60%**

Step 3: Generate Frequent Itemsets

Step 3.1: Find 1-itemsets (L1)

Calculate the support count for each item.

Item	Support Count
Milk	4
Bread	5

Item	Support Count
Butter	4

Since all items appear in at least **3 transactions (50% support)**, all are **frequent itemsets (L1)**.

Step 3.2: Generate 2-itemsets (L2)

Form pairs of frequent items:

Itemset	Support Count
{Milk, Bread}	4
{Milk, Butter}	3
{Bread, Butter}	4

All these itemsets meet the **50% support threshold**, so they are **frequent itemsets (L2)**.

Step 3.3: Generate 3-itemsets (L3)

Form a set of three items:

Itemset	Support Count
{Milk, Bread, Butter}	3

This itemset meets the **50% support threshold**, so it is **frequent (L3)**.

Step 4: Generate Association Rules

Now, generate **association rules** using confidence.

Example Rules from L3

1. {Milk, Bread} → Butter

- Confidence = $\text{Support}(\{\text{Milk, Bread, Butter}\}) / \text{Support}(\{\text{Milk, Bread}\})$
- = $3/4 = \mathbf{75\% \text{ (Accepted)}}$

2. {Milk, Butter} → Bread

- Confidence = $3/3 = \mathbf{100\% \text{ (Accepted)}}$

3. {Bread, Butter} → Milk

- Confidence = $3/4 = \mathbf{75\% \text{ (Accepted)}}$

Final Results

Frequent Itemsets:

- **L1:** {Milk}, {Bread}, {Butter}
- **L2:** {Milk, Bread}, {Milk, Butter}, {Bread, Butter}
- **L3:** {Milk, Bread, Butter}

Association Rules:

- **Milk & Bread → Butter (75%)**

- **Milk & Butter** → **Bread (100%)**
 - **Bread & Butter** → **Milk (75%)**
-

Conclusion

The Apriori Algorithm helps identify frequently purchased items and their relationships. It can be extended to large datasets for real-world applications such as **retail, e-commerce, and recommendation systems**.

@SSRoy