

## Apriori Algorithm

### Introduction:

**Apriori Algorithm:** The Apriori algorithm is a data mining technique employed to identify frequent patterns or associations within a large data set. Its main objective is to discover sets of items that frequently occur together in the dataset.

### Dataset:

Transaction No	Products
1	beer, wine, cheese
2	beer, potato chips
3	eggs, flour, butter, cheese
4	eggs, flour, butter, beer, potato chips
5	wine, cheese
6	potato chips
7	eggs, flour, butter, wine, cheese
8	eggs, flour, butter, beer, potato chips
9	wine, beer
10	beer, potato chips
11	butter, eggs
12	beer, potato chips
13	flour, eggs
14	beer, potato chips
15	eggs, flour, butter, wine, cheese
16	beer, wine, potato chips, cheese
17	wine, cheese
18	beer, potato chips
19	wine, cheese
20	beer, potato chips

### Understanding of the dataset:

The dataset is about studying how people buy things together. It shows which items are often purchased at the same time. This helps businesses understand customer behavior and make better decisions to improve their sales and customer satisfaction.

Now, Applying Apriori Algorithm to the data and finding the solution for each question given.

- Computing the support for each individual item
- Use 7 as the minimum support.
- Generate Association Rules and compute confidence.
- Use min confidence as 85%.

## Steps to Apply Apriori Algorithm:

### Step 1: Calculating the support for each item.

The initial step of the algorithm involves computing the support for each individual item. This involves counting how many times each product appears in the transactions and determining the frequency of occurrence for each product.

Items	Support (Freq of occurrence)
Beer	11
Wine	8
Cheese	8
Potato Chips	10
Eggs	7
Butter	6
Flour	6

### Step 2: Comparing the support count with the Minimum Support.

Here we are comparing each item's support count with the minimum support count given. If the support count is less than the minimum support count, then we are removing all those items.

Given minimum support = 7

The items with support count > minimum support is listed below.

Items	Support
Beer	11
Wine	8
Cheese	8
Potato Chips	10
Eggs	7

### Step 3: Combining 2 items.

In this step, we are considering 2 item sets and counting the number of times they are repeated in the given dataset.

Item set	Support
Beer, Wine	3
Beer, Cheese	2
Beer, Potato Chips	9
Beer, Eggs	2
Wine, Cheese	7
Wine, Potato Chips	1

Wine, Eggs	2
Cheese, Eggs	3
Cheese, Potato Chips	1
Potato Chips, Eggs	2

**From the table, we can clearly find that there are only two item sets which have support value greater than or equal to 7. Now, we are considering only those 2 and discarding the rest of the item sets.**

Item set	Support
Beer, Potato Chips	9
Wine, Cheese	7

#### **Step 4: Combining 3 items.**

In this step, we follow the same method as in step 2, but we consider 3 item sets.

Item set	Support
Beer, Potato Chips, Wine	1
Beer, Potato Chips, Cheese	1
Wine, Cheese, Beer	2
Wine, Cheese, Potato Chips	1

There is no frequent item set if we consider 3 item sets.

Therefore, we consider **Beer, Potato Chips and Wine, Cheese** which we achieved in Step 3 as our large datasets and continue with the next steps.

#### **Step 5: Generating Association rules and computing confidence.**

**(Given min confidence = 85%)**

To generate association rules from the frequent item sets found. We must calculate the confidence of each rule. In the Apriori algorithm, confidence is a measure of the strength of the association between two items in an itemset.

**Confidence = Support in (Item set) / Support of each item**

**OR**

**Confidence (A=>B) = Support (XandY) / Support(X)**

Item set	Association Rules	Confidence
Beer, Potato chips	Beer => Potato chips	9/11 = 0.818 – 81.8 %
Potato chips, Beer	Potato Chips => Beer	9/10 = 0.9 – 90%
Wine, Cheese	Wine => Cheese	7/8 = 0.875 – 87.5 %
Cheese, Wine	Cheese => Wine	7/8 = 0.875 – 87.5%

Here, we are not considering Beer => Potato chips because the confidence level is 81.8% which is less than given min confidence 85%

We can consider the below association rules, as they have a confidence greater than 85%:

- **Potato chips => Beer (90% confidence):**

This means that when customers purchase potato chips, there is a 90% chance that they will also buy beer.

- **Wine => Cheese (87.5% confidence):**

When customers buy wine, there is an 87.5% chance that they will also purchase cheese. This suggests a strong relationship between wine and cheese.

- **Cheese => Wine (87.5% confidence):**

When customers purchase cheese, there is an 87.5% chance that they will also buy wine.

References: <https://t4tutorials.com/apriori-algorithm-in-data-mining-with-examples/>

<https://athena.ecs.csus.edu/~mei/associationcw/Apriori.html>

<https://www.scaler.com/topics/apriori-algorithm-in-data-mining/>

<https://www.javatpoint.com/apriori-algorithm-in-machine-learning>

<https://www.educative.io/answers/what-is-the-apriori-algorithm>