# **Apriori Algorithm**

Detailed Implementation:

Support = frequency of the itemset in data base/total number of itemsets in database

Confidence (A, B) = frequency of B/frequency of AUB, where A and B are itemsets.

Steps:

1. First find all the 1-itemset frequent items by scanning the database file and if the support of the itemset is greater than minimum support, we add it in a dictionary with its support.
2. Now, join the itemsets to itself (suppose the size of itemsets is ‘L’) by taking union and extract those union of itemsets, whose size is L+1. Now, find the support of those itemsets whose support is more than minimum support.
3. Take the union of itemsets like before again, till all the bigger itemsets have support less than minimum support.
4. Add all those itemsets with their support in the dictionary.
5. Now, we have all the frequent items.
6. To find association rules, scan all the frequent items and find association rules of all the frequent items. Below is explanation of how to find association rules from a frequent Itemset.
7. Suppose we have to find association rules from Itemset S. Find all the subsets of S except the empty set and S itself. Now find the confidence(subset, S - subset). If the confidence is greater than minimum confidence, add it in the set of Association rules.

Comparison:

We used database file (retail.dat) using Apriori and FP-Tree based algorithm for different Minimum Support and Minimum confidence.

|  |  |  |  |
| --- | --- | --- | --- |
| Minimum Support | Min Confidence | Time (Apriori) | Time (FP-Tree) |
| 2000/88000 | 0.4 | 84.531 seconds |  |
| 5000/88000 | 0.4 | 30.253 seconds |  |
| 5000/88000 | 0.6 | 29.569 seconds |  |
| 10000/88000 | 0.4 | 10.714 seconds |  |
| 20000/88000 | 0.4 | 4.354 seconds |  |