

Apriori algorithm steps in data mining are the following-

1. **Define minimum threshold**

* The first step is to decide on the threshold value for the support metric.
* The metric determining the minimum number of times an item should appear in the data to be considered significant is “Support.”
* The support value is based on data size, domain knowledge, or other considerations.

1. **Create a list of frequent items**

* After determining the support threshold, the subsequent step involves scanning the entire dataset to identify items that meet the support threshold.
* The selected itemsets, meeting the support threshold, are called frequent itemsets.

1. **Create candidate item sets**

* The next step is to use the previously identified frequent k-item sets and generate a list of candidate item sets.
* The length of these candidate itemsets is k+1.

1. **Calculate the Support of each candidate**

* The algorithm needs to scan the dataset again and count the frequency of every candidate item, i.e., the number of times each item appeared in the data.

1. **Prune the candidate item sets**

* The minimum threshold is again used to remove itemsets that fail to meet the minimum support threshold.

1. **Iterate**

* This can be considered the most crucial stage of the Apriori algorithm. Steps 3 – 5 are repeated until no frequent itemsets can be generated.

1. **Generate Association Rules**

* The algorithm now generates the association rules using the final frequent item sets identified at the end of the previous step.

1. **Evaluate Association Rules**

* Metrics such as Confidence and Lift can be employed to filter the relevant association rules.
* At the end of this step, you get association rules that indicate the probability of a customer purchasing an item Z if they have already purchased an item Y (here, Y and Z are itemsets).