# Association Rule Mining

Learning Association Rule Mining made easy!



## What is Association Rule Mining ?

Association Rule Mining is an important component of data mining. Association Rules Mining is used for finding **frequent item set**, associations, correlations, or causal structures among sets of items or objects.

**Following are the steps for Association Rule Mining:**

- Frequent Itemset Generation :
  - Apriori Algorithm.
  - FP Growth Algorithm.
- Rule Generation.

## Applications of association rules

- Market basket data analysis, which aims to discover how items purchased by customers in a supermarket are associated.
- Shopping centres use association rules to place the items next to each other so that users buy more items.
- Amazon use association mining to recommend you the items based on the current item you are browsing/buying.
- Medical Application to make decision about medical diagnose should be assigned to this patient?

## What is Frequent Item Set ?

It refers to set of items that frequently appear together and satisfies both **minimum support** threshold and **minimum confidence** threshold.

**For example**

Consider the below transaction where items A, B, C are brought together in first transaction (TID 1) , items A,B,D are brought together in second transaction (TID 2) and so on.

Given that minimum threshold support is 60% and minimum threshold confidence is 70%.

We can clearly see that items { B, C } has support (60%) >= minimum threshold support and confidence (75%) >= minimum threshold confidence . So { B, C } is a frequent item set.

| TID | Items | Support = Occurrence /Total Support | Given X => Y Confidence = Occurrence( X U Y )/ Occurrence ( X ) |
|---|---|---|---|
| 1 | A,B,C | Total Support = 5 | |
| 2 | A,B,D | Support { A,B } = 2/5 = 40% | Confidence { A => B } = 2/3 = 66% |
| 3 | B,C | Support { B,C } = 3/5 = 60% | Confidence { B => C } = 3/4 = 75% |
| 4 | A,C | Support { A,B,C} = 1/5 = 20% | Confidence { A,B => C } = 1/2 = 50% |
| 5 | B,C,D | | |

## What is support ?

It is the percentage of the transaction, in which all the items in the item set is bought together.

$$\text{SUPPORT} = \frac{\text{How many times the items are bought together}}{\text{Number of Transactions}} * 100\ \%$$

**For example**

Consider the itemset { B, C } in the above transaction, as items B and C are bought together in 3 out of 5 transactions.

**Support (B, C)** = ( 3 / 5 ) * 100 % = 60%

**Formulae:** Support (A , B) = Probability(A U B) [7].

# What is confidence ?

The rule X => Y holds with confidence C if in C% of the transaction, customers who purchased a X also bought the Y.

$$\text{Confidence (X -> Y)} = \frac{\text{How many times the items X , Y are bought together}}{\text{How many times an item X is brought}} * 100\ \%$$

**For example**

Consider the itemset { B, C } in the above transaction, As both items B and C are bought together in 3 transactions and item B is bought in 4 transactions.

**Confidence (B -> C)** = ( 3 / 4 ) * 100 % = 75%

**Formulae:** Confidence (A -> B) = Probability(A U B)/Probability(A) [7].

# What is minimum support/confidence threshold ?

Association rules are considered interesting if they satisfy minimum support threshold and minimum confidence threshold which is set by users or domain experts. If an itemset I does not satisfy the minimum support threshold, then I is not frequent