

# Association Rules Mining

Association rule mining can be viewed as a two-step process:-

- Find all frequent item sets
  1. Apriori Method
  2. FP-Growth
- Generate strong association rules from the frequent item sets
  1. Must Satisfy min support
  2. Must satisfy min confidence

# Apriori Algorithm in Data Mining




Minimum Support = 30%

Threshold Confidence = 80%

Transaction ID (TID)	Items
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5

Step 1:

Items	Support
1	2/4 = 50%
2	3/4 = 75%
3	3/4 = 75%
 4	1/4 = 25%
5	3/4 = 75%

So, Item set  $\neq \{1, 2, 3, 5\}$

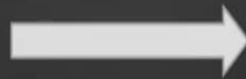
# Apriori Algorithm in Data Mining



## Step 2:

Item set	Support
{1,2}	$1/4 = 25\%$
{1,3}	$2/4 = 50\%$
{1,5}	$1/4 = 25\%$
{2,3}	$2/4 = 50\%$
{2,5}	$3/4 = 75\%$
{3,5}	$2/4 = 50\%$

Transaction ID (TID)	Items
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5



## Step 2:

Item set	Support
{1,3}	$2/4 = 50\%$
{2,3}	$2/4 = 50\%$
{2,5}	$3/4 = 75\%$
{3,5}	$2/4 = 50\%$



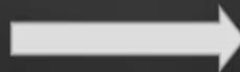
# Apriori Algorithm in Data Mining



## Step 2:

Item set	Support
$\{1,3\}$	$2/4 = 50\%$
$\{2,3\}$	$2/4 = 50\%$
$\{2,5\}$	$3/4 = 75\%$
$\{3,5\}$	$2/4 = 50\%$

Transaction ID (TID)	Items
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5



## Step 3:

Item set	Support
$\{1,3,5\}$	$1/4 = 25\%$
$\{2,3,5\}$	$2/4 = 50\%$
$\{1,2,3\}$	$1/4 = 25\%$

Final Item Set:  $\{2,3,5\}$



# Apriori Algorithm in Data Mining



## Step 4:

Rules	Support
$(2, 3) \rightarrow 5$	2
$(3, 5) \rightarrow 2$	2
$(2, 5) \rightarrow 3$	2
$3 \rightarrow (2, 5)$	2
$2 \rightarrow (3, 5)$	2
$5 \rightarrow (2, 3)$	2



Transaction ID (TID)	Items
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5

## Step 3:

Item set	Support
$\{1,3,5\}$	$1/4 = 25\%$
$\{2,3,5\}$	$2/4 = 50\%$
$\{1,2,3\}$	$1/4 = 25\%$

Final Item Set:  $\{2,3,5\}$



# Apriori Algorithm in Data Mining



## Step 4:

Rules	Support
$(2, 3) \rightarrow 5$	2
$(3, 5) \rightarrow 2$	2
$(2, 5) \rightarrow 3$	2
$3 \rightarrow (2, 5)$	2
$2 \rightarrow (3, 5)$	2
$5 \rightarrow (2, 3)$	2

## Step 2:

Item set	Support
$\{1,3\}$	$2/4 = 50\%$
$\{2,3\}$	$2/4 = 50\%$
$\{2,5\}$	$3/4 = 75\%$
$\{3,5\}$	$2/4 = 50\%$

**Confidence  $(A \rightarrow B) = \text{Support}(A \cup B) / \text{Support}(A)$**

So,

$$\begin{aligned} (2, 3) \rightarrow 5 &= S((2, 3) \cup 5) / S(2, 3) \\ &= 2 / 2 \\ &= 100\% \end{aligned}$$

$$\begin{aligned} (3, 5) \rightarrow 2 &= S((3, 5) \cup 2) / S(3, 5) \\ &= 2 / 2 \\ &= 100\% \end{aligned}$$

$$\begin{aligned} (2, 5) \rightarrow 3 &= S((2, 5) \cup 3) / S(2, 5) \\ &= 2 / 3 \\ &= 67\% \end{aligned}$$

$$\begin{aligned} 3 \rightarrow (2, 5) &= S(3 \cup (2, 5)) / S(3) \\ &= 2 / 3 \\ &= 67\% \end{aligned}$$

$$\begin{aligned} 2 \rightarrow (3, 5) &= S(2 \cup (3, 5)) / S(2) \\ &= 2 / 3 \\ &= 67\% \end{aligned}$$

$$\begin{aligned} 5 \rightarrow (2, 3) &= S(5 \cup (2, 3)) / S(5) \\ &= 2 / 3 \\ &= 67\% \end{aligned}$$



# Apriori Algorithm in Data Mining

## Step 4:

Rules	Support
$(2, 3) \rightarrow 5$	2
$(3, 5) \rightarrow 2$	2
$(2, 5) \rightarrow 3$	2
$3 \rightarrow (2, 5)$	2
$2 \rightarrow (3, 5)$	2
$5 \rightarrow (2, 3)$	2

Minimum Support = 30%

Threshold Confidence = 80%

Transaction ID (TID)	Items
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5

## Step 5:

Rules	Support	Confidence
$(2, 3) \rightarrow 5$	2	$2/2 = 100\%$
$(3, 5) \rightarrow 2$	2	$2/2 = 100\%$
$(2, 5) \rightarrow 3$	2	$2/3 = 67\%$
$3 \rightarrow (2, 5)$	2	$2/3 = 67\%$
$2 \rightarrow (3, 5)$	2	$2/3 = 67\%$
$5 \rightarrow (2, 3)$	2	$2/3 = 67\%$

After compare with threshold confidence (80%):

Final rules are,  $(2, 3) \rightarrow 5$  &  $(3, 5) \rightarrow 2$