Apriori Algorithm

APRIORI ALGORITHM

Market Basket Analysis

Market Basket Analysis

iPhone X

Walmart

TARGET

Market Basket Analysis

Market Basket Analysis is one of the key techniques used by large retailers to uncover associations between items.

Bread and Jam

Laptop and Bag

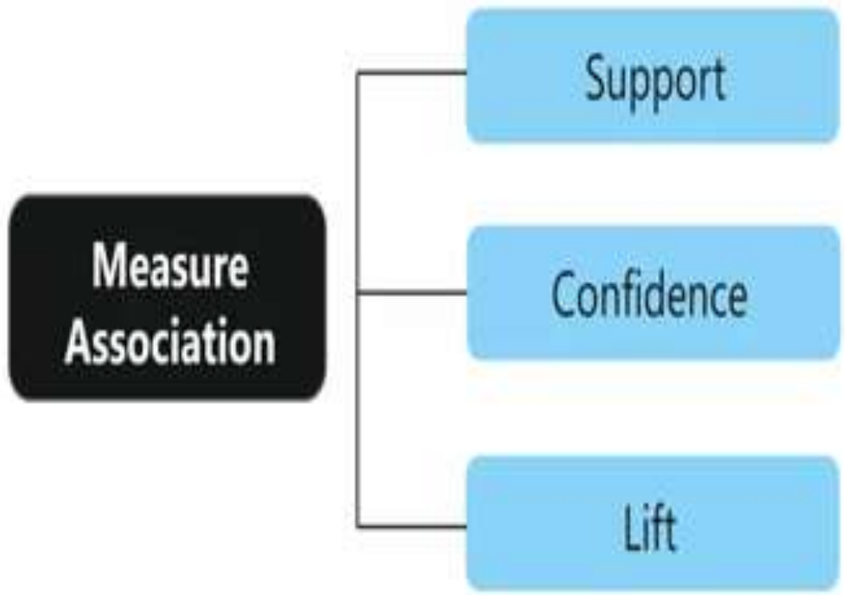
Market Basket Analysis

Market Basket Analysis is one of the key techniques used by large retailers to uncover associations between items.

Bread and Butter

Association Rule Mining

Association Rule Mining



Association Rule Mining

$$A \Rightarrow B$$

$$Support = \frac{freq(A,B)}{N}$$

$$Confidence = \frac{freq(A,B)}{freq(A)}$$

$$Lift = \frac{Support}{Supp(A)\ x\ Supp(B)}$$

Association Rule Mining

Transaction at a Local Market

T1	A	B	C
T2	A	C	D
T3	B	C	D
T4	A	D	E
T5	B	C	E

Association Rule Mining

Association Rule Mining

Rule	Support	Confidence	Lift
A=>D	2/5	2/3	10/9
C=>A	2/5	2/4	5/6
A=>C	2/5	2/3	5/6
B, C=>A	1/5	1/3	5/9

Apriori Algorithm

Apriori Algorithm

Apriori algorithm uses frequent item sets to generate association rules. It is based on the concept that a subset of a frequent itemset must also be a frequent itemset.

Frequent Itemset is an itemset whose support value is greater than a threshold value.

Apriori Algorithm

TID	Items
T1	1 3 4
T2	2 3 5
T3	1 2 3 5
T4	2 5
T5	1 3 5

Min. Support count = 2

Apriori Algorithm - 1st Iteration

TID	Items
T1	1 3 4
T2	2 3 5
T3	1 2 3 5
T4	2 5
T5	1 3 5

C1

Itemset	Support
{1}	3
{2}	3
{3}	4
{4}	1
{5}	4

Apriori Algorithm - 1st Iteration

C1

Itemset	Support
{1}	3
{2}	3
{3}	4
{4}	1
{5}	4

F1

Itemset	Support
{1}	3
{2}	3
{3}	4
{5}	4

Item sets with support value less than min. support value (i.e. 2) are eliminated

Apriori Algorithm – 2nd Iteration

TID	Items
T1	1 3 4
T2	2 3 5
T3	1 2 3 5
T4	2 5
T5	1 3 5

Only Items present in F1

C2

Itemset	Support
{1,2}	1
{1,3}	3
{1,5}	2
{2,3}	2
{2,5}	3
{3,5}	3

F2

Itemset	Support
{1,3}	3
{1,5}	2
{2,3}	2
{2,5}	3
{3,5}	3

Item sets with support value less than min. support value (i.e. 2) are eliminated

Apriori Algorithm – Pruning

C3 ?

TID	Items
T1	1 3 4
T2	2 3 5
T3	1 2 3 5
T4	2 5
T5	1 3 5

Itemset	Support
{1,2,3}	
{1,2,5}	
{1,3,5}	
{2,3,5}	

Apriori Algorithm – Pruning

TID	Items
T1	1 3 4
T2	2 3 5
T3	1 2 3 5
T4	2 5
T5	1 3 5

C3

Itemset	In F2?
{1,2,3}, {1,2}, {1,3}, {2,3}	NO
{1,2,5}, {1,2}, {1,5}, {2,5}	NO
{1,3,5},{1,5}, {1,3}, {3,5}	YES
{2,3,5}, {2,3}, {2,5}, {3,5}	YES

Apriori Algorithm – Pruning

C3

Itemset	In F2?
{1,2,3}, {1,2}, {1,3}, {2,3}	NO
{1,2,5}, {1,2}, {1,5}, {2,5}	NO
{1,3,5},{1,5}, {1,3}, {3,5}	YES
{2,3,5}, {2,3}, {2,5}, {3,5}	YES

F2

Itemset	Support
{1,3}	3
{1,5}	2
{2,3}	2
{2,5}	3
{3,5}	3

If any of the subsets of these item sets are not there in F2 then we remove that itemset

Apriori Algorithm – Pruning

TID	Items
T1	1 3 4
T2	2 3 5
T3	1 2 3 5
T4	2 5
T5	1 3 5

F3

Itemset	Support
{1,3,5}	2
{2,3,5}	2

If any of the subsets of these item sets are not there in F2 then we remove that itemset

Apriori Algorithm – 4th Iteration

TID	Items
T1	1 3 4
T2	2 3 5
T3	1 2 3 5
T4	2 5
T5	1 3 5

F3

Itemset	Support
{1,3,5}	2
{2,3,5}	2

C3

Itemset	Support
{1,2,3,5}	1

Since support of C4 is less than 2, stop and return to the previous itemset, i.e. C3

Apriori Algorithm – Subset Creation

F3

Itemset	Support
{1,3,5}	2
{2,3,5}	2

For I = {1,3,5}, subsets are {1,3}, {1,5}, {3,5}, {1}, {3}, {5}

For I = {2,3,5}, subsets are {2,3}, {2,5}, {3,5}, {2}, {3}, {5}

- For every subsets S of I, output the rule:

S → **(I-S)** (S recommends I-S)

if ***support(I)/support(S) >= min_conf value***

Apriori Algorithm – Applying Rules

Applying Rules to Item set F3

1. **{1,3,5}**

   ✓ Rule 1: **{1,3}** → **({1,3,5} - {1,3})** means 1 & 3 → 5
   Confidence = support(1,3,5)/support(1,3) = 2/3 = 66.66% > 60%
   Rule 1 is selected

   ✓ Rule 2: **{1,5}** → **({1,3,5} - {1,5})** means 1 & 5 → 3
   Confidence = support(1,3,5)/support(1,5) = 2/2 = 100% > 60%
   Rule 2 is selected

   ✓ Rule 3: **{3,5}** → **({1,3,5} - {3,5})** means 3 & 5 → 1
   Confidence = support(1,3,5)/support(3,5) = 2/3 = 66.66% > 60%
   Rule 3 is selected

Apriori Algorithm – Applying Rules

Applying Rules to Item set F3

1. **{1,3,5}**

✓ Rule 4: **{1}** → **({1,3,5} - {1})** means 1 → 3 & 5

Confidence = support(1,3,5)/support(1) = 2/3 = 66.66% > 60%

Rule 4 is selected

✓ Rule 5: **{3}** → **({1,3,5} - {3})** means 3 → 1 & 5

Confidence = support(1,3,5)/support(3) = 2/4 = 50% <60%

Rule 5 is rejected

✓ Rule 6: **{5}** → **({1,3,5} - {5})** means 5 → 1 & 3

Confidence = support(1,3,5)/support(3) = 2/4 = 50% < 60%

Rule 6 is rejected

THANK YOU