

CS 422
DATA MINING
HOME WORK – 3

SUBMITTED BY :
SACHIN KRISHNA MURTHY
CWID : A20354077

Definitions :

- **Association Rule Mining :**

Association rule mining is finding frequent patterns, associations, correlations, or causal structures among sets of objects in transactional or relational databases, and other information repositories.

- **Apriori Algorithm :**

Apriori is an algorithm for frequent item set mining and association rule mining over transactional databases. It works by identifying the frequent individual items in the database and extending them to larger item sets as long as they appear sufficiently in the database.

- **FPGrowth Algorithm :**

The FP-Growth Algorithm, is an efficient and scalable method for mining the complete set of frequent patterns by pattern fragment growth, using extended prefix-tree named frequent-pattern tree (FP-tree).

- **Support :**

Support Count is an important property of an itemset which refers to the number of transactions that contain a particular itemset.

- **Confidence :**

Confidence determines how frequently items in Y appear in transactions that contain X.

Part 1

Chapter 6

2. Consider the data set shown in Table 6.1.

Table 6.1. Example of market basket transactions.

Customer ID	Transaction ID	Items Bought
1	0001	{a, d, e}
1	0024	{a, b, c, e}
2	0012	{a, b, d, e}
2	0031	{a, c, d, e}
3	0015	{b, c, e}
3	0022	{b, d, e}
4	0029	{c, d}
4	0040	{a, b, c}
5	0033	{a, d, e}
5	0038	{a, b, e}

(a) Compute the support for itemsets {e}, {b,d}, and {b,d,e} by treating each transaction ID as a market basket.

Solution :

- Support Count is an important property of an itemset which refers to the number of transactions that contain a particular itemset.
- The support count $\sigma(X)$ for an itemset X can be given by :
$$\sigma(X) = | \{ t_i | X \subseteq t_i, t_i \in T \} |$$

where

t_i : Transaction

X : Itemset

T : Set of all transactions

$$\text{Support, } s(X \rightarrow Y) = \sigma(X \cup Y) / N$$

From the above table given,

- Number of transactions that contain {e} : 8
- Number of transactions that contain {b,d} : 2
- Number of transactions that contain {b,d,e} : 2
- Total number of transactions : 10

Therefore,

$$s(\{e\}) = 8 / 10 = 0.8$$

$$s(\{b,d\}) = 2 / 10 = 0.2$$

$$s(\{b,d,e\}) = 2 / 10 = 0.2$$

- (b) Use the results in part (a) to compute the confidence for the association rules $\{b,d\} \rightarrow \{e\}$ and $\{e\} \rightarrow \{b,d\}$. Is confidence a symmetric measure?

Solution :

Confidence determines how frequently items in Y appear in transactions that contain X and is given by :

$$\text{Confidence, } c(X \rightarrow Y) = \sigma(X \cup Y) / \sigma(X)$$

Therefore,

$$\begin{aligned} c(bd \rightarrow e) &= 0.2/0.2 \\ &= 1 \\ &= 100\% \end{aligned}$$

$$\begin{aligned} C(e \rightarrow bd) &= 0.2/0.8 \\ &= 0.25 \\ &= 25\% \end{aligned}$$

6. Consider the market basket transactions shown in Table 6.2.

Table 6.2. Market basket transactions.

Transaction ID	Items Bought
1	{Milk, Beer, Diapers}
2	{Bread, Butter, Milk}
3	{Milk, Diapers, Cookies}
4	{Bread, Butter, Cookies}
5	{Beer, Cookies, Diapers}
6	{Milk, Diapers, Bread, Butter}
7	{Bread, Butter, Diapers}
8	{Beer, Diapers}
9	{Milk, Diapers, Bread, Butter}
10	{Beer, Cookies}

- a) What is the maximum number of association rules that can be extracted from this data (including rules that have zero support)?

Number of possible rules extracted from a dataset, that contains d items

$$R = 3^d - 2^{d+1} + 1$$

Number of items = 6

$$d = 6$$

$$R = 3^6 - 2^7 + 1$$

$$R = 729 - 128 + 1$$

R = 602

Number of rules = 602

- (c) Write an expression for the maximum number of size-3 itemsets that can be derived from this data set.

Maximum number of size-3 itemsets that can be derived from above set = ${}^6C_3 = 20$

- d) Find an itemset (of size 2 or larger) that has the largest support.

I = {Milk, Beer, Butter, Diapers, Cookies, Bread}

$$s(\{ \text{Milk, Beer} \}) = \sigma(\text{Milk} \cup \text{Beer}) / \sigma(N) = 1/10 = 0.1$$

$$s(\{ \text{Milk, Butter} \}) = \sigma(\text{Milk} \cup \text{Butter}) / \sigma(N) = 3/10 = 0.3$$

$$s(\{ \text{Milk, Diapers} \}) = \sigma(\text{Milk} \cup \text{Diapers}) / \sigma(N) = 4/10 = 0.4$$

$$s(\{ \text{Milk, Cookies} \}) = \sigma(\text{Milk} \cup \text{Cookies}) / \sigma(N) = 1/10 = 0.1$$

$$s(\{ \text{Milk, Bread} \}) = \sigma(\text{Milk} \cup \text{Bread}) / \sigma(N) = 3/10 = 0.3$$

$$s(\{ \text{Beer, Butter} \}) = \sigma(\text{Beer} \cup \text{Butter}) / \sigma(N) = 0/10 = 0$$

$$s(\{ \text{Beer, Diapers} \}) = \sigma(\text{Beer} \cup \text{Diapers}) / \sigma(N) = 3/10 = 0.3$$

$$s(\{ \text{Beer, Cookies} \}) = \sigma(\text{Beer} \cup \text{Cookies}) / \sigma(N) = 2/10 = 0.2$$

$$s(\{ \text{Beer, Bread} \}) = \sigma(\text{Beer} \cup \text{Bread}) / \sigma(N) = 0/10 = 0$$

$$s(\{ \text{Butter, Diapers} \}) = \sigma(\text{Butter} \cup \text{Diapers}) / \sigma(N) = 3/10 = 0.3$$

$$s(\{ \text{Butter, Cookies} \}) = \sigma(\text{Butter} \cup \text{Cookies}) / \sigma(N) = 1/10 = 0.1$$

$$s(\{ \text{Butter, Bread} \}) = \sigma(\text{Butter} \cup \text{Bread}) / \sigma(N) = 5/10 = 0.5$$

$$s(\{ \text{Diapers, Cookies} \}) = \sigma(\text{Diapers} \cup \text{Cookies}) / \sigma(N) = 2/10 = 0.2$$

$$s(\{ \text{Diapers, Bread} \}) = \sigma(\text{Diapers} \cup \text{Bread}) / \sigma(N) = 3/10 = 0.3$$

$$s(\{ \text{Cookies, Bread} \}) = \sigma(\text{Cookies} \cup \text{Bread}) / \sigma(N) = 1/10 = 0.1$$

$$s(\{ \text{Milk, Beer, Butter} \}) = \sigma(\text{Milk} \cup \text{Beer} \cup \text{Butter}) / \sigma(N) = 0/10 = 0.2$$

$$s(\{ \text{Milk, Beer, Diapers} \}) = \sigma(\text{Milk} \cup \text{Beer} \cup \text{Diapers}) / \sigma(N) = 1/10 = 0.1$$

$$s(\{ \text{Milk, Beer, Cookies} \}) = \sigma(\text{Milk} \cup \text{Beer} \cup \text{Cookies}) / \sigma(N) = 0/10 = 0$$

$$s(\{ \text{Milk, Beer, Bread} \}) = \sigma(\text{Milk} \cup \text{Beer} \cup \text{Bread}) / \sigma(N) = 0/10 = 0$$

$$s(\{ \text{Milk, Butter, Diapers} \}) = \sigma(\text{Milk} \cup \text{Butter} \cup \text{Diapers}) / \sigma(N) = 2/10 = 0.2$$

$$s(\{ \text{Milk, Butter, Cookies} \}) = \sigma(\text{Milk} \cup \text{Butter} \cup \text{Cookies}) / \sigma(N) = 0/10 = 0$$

$$s(\{ \text{Milk, Butter, Bread} \}) = \sigma(\text{Milk} \cup \text{Butter} \cup \text{Bread}) / \sigma(N) = 3/10 = 0.3$$

$$s(\{ \text{Milk, Diapers, Cookies} \}) = \sigma(\text{Milk} \cup \text{Diapers} \cup \text{Cookies}) / \sigma(N) = 1/10 = 0.1$$

$$s(\{ \text{Milk, Diapers, Bread} \}) = \sigma(\text{Milk} \cup \text{Diapers} \cup \text{Bread}) / \sigma(N) = 2/10 = 0.2$$

$$s(\{ \text{Milk, Cookies, Bread} \}) = \sigma(\text{Milk} \cup \text{Cookies} \cup \text{Bread}) / \sigma(N) = 0/10 = 0$$

$$s(\{ \text{Beer, Butter, Diapers} \}) = \sigma(\text{Beer} \cup \text{Butter} \cup \text{Diapers}) / \sigma(N) = 0/10 = 0$$

$$s(\{ \text{Beer, Butter, Cookies} \}) = \sigma(\text{Beer} \cup \text{Butter} \cup \text{Cookies}) / \sigma(N) = 0/10 = 0$$

$$s(\{ \text{Beer, Butter, Bread} \}) = \sigma(\text{Beer} \cup \text{Butter} \cup \text{Bread}) / \sigma(N) = 0/10 = 0$$

$$s(\{ \text{Beer, Diapers, Cookies} \}) = \sigma(\text{Beer} \cup \text{Diapers} \cup \text{Cookies}) / \sigma(N) = 1/10 = 0.1$$

$$s(\{ \text{Beer, Diapers, Bread} \}) = \sigma(\text{Beer} \cup \text{Diapers} \cup \text{Bread}) / \sigma(N) = 0/10 = 0$$

$$s(\{ \text{Beer, Cookies, Bread} \}) = \sigma(\text{Beer} \cup \text{Cookies} \cup \text{Bread}) / \sigma(N) = 0/10 = 0$$

$$s(\{ \text{Butter, Diapers, Cookies} \}) = \sigma(\text{Butter} \cup \text{Diapers} \cup \text{Cookies}) / \sigma(N) = 0/10 = 0$$

$$s(\{ \text{Butter, Diapers, Bread} \}) = \sigma(\text{Butter} \cup \text{Diapers} \cup \text{Bread}) / \sigma(N) = 3/10 = 0.3$$

$$s(\{ \text{Butter, Cookies, Bread} \}) = \sigma(\text{Butter} \cup \text{Cookies} \cup \text{Bread}) / \sigma(N) = 1/10 = 0.1$$

$$s(\{ \text{Diapers, Cookies, Bread} \}) = \sigma(\text{Diapers} \cup \text{Cookies} \cup \text{Bread}) / \sigma(N) = 0/10 = 0$$

$$s(\{ \text{Milk, Butter, Diapers, Bread} \}) = \sigma(\text{Milk} \cup \text{Butter} \cup \text{Diapers} \cup \text{Bread}) / \sigma(N) = 2/10 = 0.2$$

Maximum number of items in a transaction set is 4 and above set is the only one itemset with transaction size 4

Pair with highest support of 0.5 is { Butter, Bread }.

e) Find a pair of items, a and b, such that the rules $\{a\} \rightarrow \{b\}$ and $\{b\} \rightarrow \{a\}$ have the same confidence.

$I = \{\text{Milk, Beer, Butter, Diapers, Cookies, Bread}\}$

$$c(\{ \text{Milk} \rightarrow \text{Beer} \}) = \sigma(\text{Milk} \cup \text{Beer}) / \sigma(\text{Milk}) = 1/5 = 0.2$$

$$c(\{ \text{Beer} \rightarrow \text{Milk} \}) = \sigma(\text{Beer} \cup \text{Milk}) / \sigma(\text{Beer}) = 1/4 = 0.25$$

$$c(\{ \text{Milk} \rightarrow \text{Butter} \}) = \sigma(\text{Milk} \cup \text{Butter}) / \sigma(\text{Milk}) = 3/5 = 0.6$$

$$c(\{ \text{Butter} \rightarrow \text{Milk} \}) = \sigma(\text{Butter} \cup \text{Milk}) / \sigma(\text{Butter}) = 3/5 = 0.6$$

$$c(\{ \text{Milk} \rightarrow \text{Diapers} \}) = \sigma(\text{Milk} \cup \text{Diapers}) / \sigma(\text{Milk}) = 4/5 = 0.8$$

$$c(\{ \text{Diapers} \rightarrow \text{Milk} \}) = \sigma(\text{Diapers} \cup \text{Milk}) / \sigma(\text{Diapers}) = 4/7 = 0.57$$

$$c(\{ \text{Milk} \rightarrow \text{Cookies} \}) = \sigma(\text{Milk} \cup \text{Cookies}) / \sigma(\text{Milk}) = 1/5 = 0.2$$

$$c(\{ \text{Cookies} \rightarrow \text{Milk} \}) = \sigma(\text{Cookies} \cup \text{Milk}) / \sigma(\text{Cookies}) = 1/4 = 0.25$$

$c(\{ \text{Milk} \rightarrow \text{Bread} \}) = \sigma(\text{Milk} \cup \text{Bread}) / \sigma(\text{Milk}) = 3/5 = 0.6$
 $c(\{ \text{Bread} \rightarrow \text{Milk} \}) = \sigma(\text{Bread} \cup \text{Milk}) / \sigma(\text{Bread}) = 3/5 = 0.6$
 $c(\{ \text{Beer} \rightarrow \text{Butter} \}) = \sigma(\text{Beer} \cup \text{Butter}) / \sigma(\text{Beer}) = 0/4 = 0$
 $c(\{ \text{Butter} \rightarrow \text{Beer} \}) = \sigma(\text{Butter} \cup \text{Beer}) / \sigma(\text{Butter}) = 0/5 = 0$
 $c(\{ \text{Beer} \rightarrow \text{Diapers} \}) = \sigma(\text{Beer} \cup \text{Diapers}) / \sigma(\text{Beer}) = 3/4 = 0.75$
 $c(\{ \text{Diapers} \rightarrow \text{Beer} \}) = \sigma(\text{Diapers} \cup \text{Beer}) / \sigma(\text{Diapers}) = 3/7 = 0.43$
 $c(\{ \text{Beer} \rightarrow \text{Cookies} \}) = \sigma(\text{Beer} \cup \text{Cookies}) / \sigma(\text{Beer}) = 2/4 = 0.5$
 $c(\{ \text{Cookies} \rightarrow \text{Beer} \}) = \sigma(\text{Cookies} \cup \text{Beer}) / \sigma(\text{Cookies}) = 2/4 = 0.5$
 $c(\{ \text{Beer} \rightarrow \text{Bread} \}) = \sigma(\text{Beer} \cup \text{Bread}) / \sigma(\text{Beer}) = 0/4 = 0$
 $c(\{ \text{Bread} \rightarrow \text{Beer} \}) = \sigma(\text{Bread} \cup \text{Beer}) / \sigma(\text{Bread}) = 0/5 = 0$
 $c(\{ \text{Butter} \rightarrow \text{Diapers} \}) = \sigma(\text{Butter} \cup \text{Diapers}) / \sigma(\text{Butter}) = 3/5 = 0.6$
 $c(\{ \text{Diapers} \rightarrow \text{Butter} \}) = \sigma(\text{Diapers} \cup \text{Butter}) / \sigma(\text{Diapers}) = 3/7 = 0.43$
 $c(\{ \text{Butter} \rightarrow \text{Cookies} \}) = \sigma(\text{Butter} \cup \text{Cookies}) / \sigma(\text{Butter}) = 1/5 = 0.2$
 $c(\{ \text{Cookies} \rightarrow \text{Butter} \}) = \sigma(\text{Cookies} \cup \text{Butter}) / \sigma(\text{Cookies}) = 1/4 = 0.25$
 $c(\{ \text{Butter} \rightarrow \text{Bread} \}) = \sigma(\text{Butter} \cup \text{Bread}) / \sigma(\text{Butter}) = 5/5 = 1$
 $c(\{ \text{Bread} \rightarrow \text{Butter} \}) = \sigma(\text{Bread} \cup \text{Butter}) / \sigma(\text{Bread}) = 5/5 = 1$
 $c(\{ \text{Diapers} \rightarrow \text{Cookies} \}) = \sigma(\text{Diapers} \cup \text{Cookies}) / \sigma(\text{Diapers}) = 2/7 = 0.29$
 $c(\{ \text{Cookies} \rightarrow \text{Diapers} \}) = \sigma(\text{Cookies} \cup \text{Diapers}) / \sigma(\text{Cookies}) = 2/4 = 0.5$
 $c(\{ \text{Diapers} \rightarrow \text{Bread} \}) = \sigma(\text{Diapers} \cup \text{Bread}) / \sigma(\text{Diapers}) = 3/7 = 0.43$
 $c(\{ \text{Bread} \rightarrow \text{Diapers} \}) = \sigma(\text{Bread} \cup \text{Diapers}) / \sigma(\text{Bread}) = 3/5 = 0.6$
 $c(\{ \text{Cookies} \rightarrow \text{Bread} \}) = \sigma(\text{Cookies} \cup \text{Bread}) / \sigma(\text{Cookies}) = 1/4 = 0.25$
 $c(\{ \text{Bread} \rightarrow \text{Cookies} \}) = \sigma(\text{Bread} \cup \text{Cookies}) / \sigma(\text{Bread}) = 1/5 = 0.25$
 Pair of two items, a and b, such that confidence $\{a \rightarrow b\} = \text{confidence } \{b \rightarrow a\}$
 $\{\text{Milk, Bread}\}, \{\text{Beer, Cookies}\}$ and $\{\text{Butter, Bread}\}$

8. The Apriori algorithm uses a generate-and-count strategy for deriving frequent itemsets. Candidate itemsets of size $k+1$ are created by joining a pair of frequent itemsets of size k (this is known as the candidate generation step). A candidate is discarded if any one of its subsets is found to be infrequent during the candidate pruning step. Suppose the Apriori algorithm is

applied to the data set shown in Table 6.3 with minsup = 30%, i.e., any itemset occurring in less than 3 transactions is considered to be infrequent.

Table 6.3. Example of market basket transactions.

Transaction ID	Items Bought
1	{a, b, d, e}
2	{b, c, d}
3	{a, b, d, e}
4	{a, c, d, e}
5	{b, c, d, e}
6	{b, d, e}
7	{c, d}
8	{a, b, c}
9	{a, d, e}
10	{b, d}

From the lattice structure given and after marking each node as 'N', 'F' and 'I' we have :

Number of itemsets marked 'F' : 16

Number of itemsets marked 'N' : 11

Number of itemsets marked 'I' : 5

b) What is the percentage of frequent itemsets (with respect to all itemsets in the lattice)?

Solution :

From the lattice structure given :

Number of frequent itemsets (Number of 'F' nodes) : 16

Total Number of itemsets : 32

Percentage of frequent itemsets : $16/32$

$$= 1/2$$

$$= 0.5$$

$$= 50\%$$

c) What is the pruning ratio of the Apriori algorithm on this data set? (Pruning ratio is defined as the percentage of itemsets not considered to be a candidate because (1) they are not generated during candidate generation or (2) they are pruned during the candidate pruning step.)

Solution :

Pruning Ratio is the ratio of 'N' itemsets to the total number of itemsets.

Number of 'N' itemsets : 11

Total number of itemsets : 32

Pruning Ratio : $11/32$

$= 0.3437$

$= 34.37\%$

d) What is the false alarm rate (i.e, percentage of candidate itemsets that are found to be infrequent after performing support counting)?

Solution :

False Rate is the ration of 'I' itemsets to the total number of itemsets.

Number of 'I' itemsets : 5

Total number of itemsets : 32

False alarm rate : $5/32$

$= 0.1562$

$= 16.62\%$

9. The Apriori algorithm uses a hash tree data structure to efficiently count the support of candidate itemsets. Consider the hash tree for candidate 3itemsets shown in Figure 6.2.

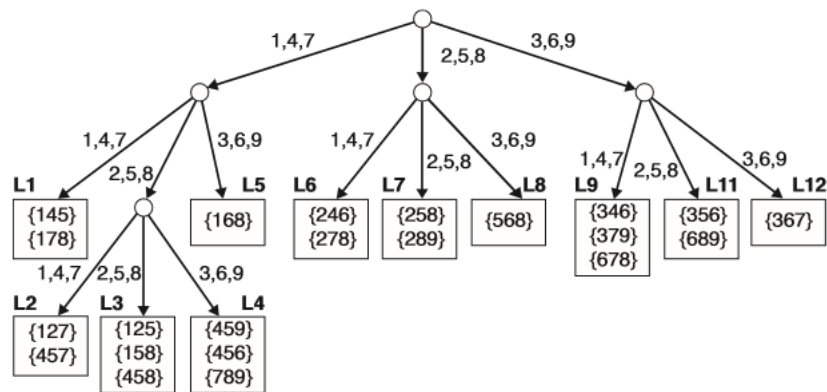


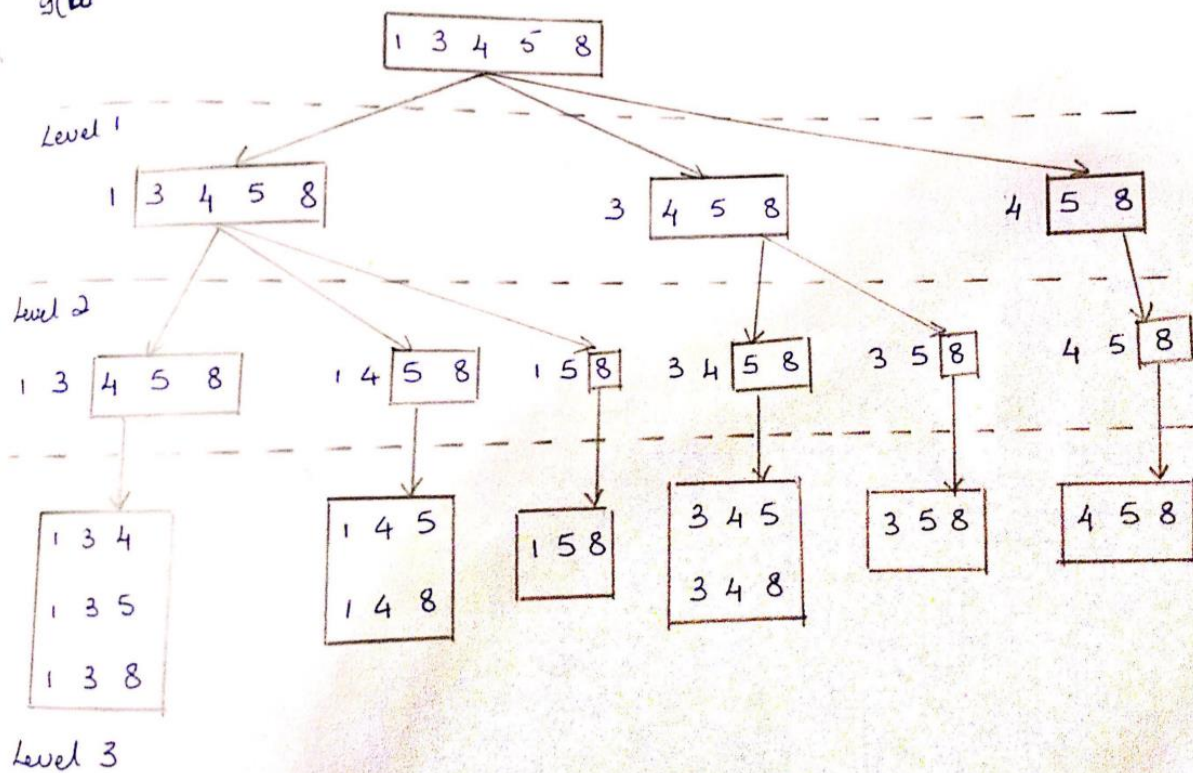
Figure 6.2. An example of a hash tree structure.

a) Given a transaction that contains items{1,3,4,5,8}, which of the hash tree leaf nodes will be visited when finding the candidates of the transaction?

9(a)

Candidate Sets	Leaf Nodes
$\{1, 3, 4\}$	L5
$\{1, 3, 5\}$	L5
$\{1, 3, 8\}$	L5
$\{1, 4, 5\}$	L1
$\{1, 4, 8\}$	L1
$\{1, 5, 8\}$	L3
$\{3, 4, 5\}$	L9
$\{3, 4, 8\}$	L9
$\{3, 5, 8\}$	L11
$\{4, 5, 8\}$	L3

9(b)



The leaf nodes visited are L1, L3, L5, L9 and L11.

- b) Use the visited leaf nodes in part (b) to determine the candidate itemsets that are contained in the transaction {1,3,4,5,8}.

The candidates contained in the transaction are {1,4,5}, {1,5,8} and {4,5,8}

Part -2

➤ **Parameters and their Default Values of Apriori Algorithm**

Parameter	Value
car	False
classIndex	-1
delta	0.05
lowerBoundMinSupport	0.1
metricType	Confidence
minMetric	0.9
numRules	10
outputItemSets	False
removeAllMissingCols	False
significanceLevel	-1.0
upperBoundMinSupport	1.0
Verbose	False

➤ **Description of the Parameters :**

car -- If enabled class association rules are mined instead of (general) association rules.

classIndex -- Index of the class attribute. If set to -1, the last attribute is taken as class attribute.

delta -- Iteratively decrease support by this factor. Reduces support until min support is reached or required number of rules has been generated.

lowerBoundMinSupport -- Lower bound for minimum support.

metricType -- Set the type of metric by which to rank rules. Confidence is the proportion of the examples covered by the premise that are also covered by the consequence(Class association rules can only be mined using confidence). Lift is confidence divided by the proportion of all examples that are covered by the consequence. This is a measure of the importance of the association that is independent of support. Leverage is the proportion of additional examples covered by both the premise and consequence above those expected if the premise and consequence were independent of each other. The total number of examples that this represents is presented in brackets following the leverage. Conviction is another measure of departure from independence. Conviction is given by $\frac{P(\text{premise})P(\text{!consequence})}{P(\text{premise}, \text{!consequence})}$.

minMetric -- Minimum metric score. Consider only rules with scores higher than this value.

numRules -- Number of rules to find.

outputItemSets -- If enabled the itemsets are output as well.

removeAllMissingCols -- Remove columns with all missing values.

significanceLevel -- Significance level. Significance test (confidence metric only).

upperBoundMinSupport -- Upper bound for minimum support. Start iteratively decreasing minimum support from this value.

verbose -- If enabled the algorithm will be run in verbose mode.

- **Data set : Supermarket**
- **Using Apriori Association Algorithm**

```
=== Run information ===  
  
Scheme:      weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1  
Relation:    supermarket  
Instances:   4627  
Attributes:  217  
[list of attributes omitted]  
=== Associator model (full training set) ===
```

- Total number of instances : 4627
- Total number of attributes : 217

Associator Model :

```
Apriori
=====

Minimum support: 0.15 (694 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 44
Size of set of large itemsets L(2): 380
Size of set of large itemsets L(3): 910
Size of set of large itemsets L(4): 633
Size of set of large itemsets L(5): 105
Size of set of large itemsets L(6): 1
```

Best Rules Found :

```
Best rules found:

1. biscuits=t frozen foods=t fruit=t total=high 788 ==> bread and cake=t 723    conf:(0.92)
2. baking needs=t biscuits=t fruit=t total=high 760 ==> bread and cake=t 696    conf:(0.92)
3. baking needs=t frozen foods=t fruit=t total=high 770 ==> bread and cake=t 705    conf:(0.92)
4. biscuits=t fruit=t vegetables=t total=high 815 ==> bread and cake=t 746    conf:(0.92)
5. party snack foods=t fruit=t total=high 854 ==> bread and cake=t 779    conf:(0.91)
6. biscuits=t frozen foods=t vegetables=t total=high 797 ==> bread and cake=t 725    conf:(0.91)
7. baking needs=t biscuits=t vegetables=t total=high 772 ==> bread and cake=t 701    conf:(0.91)
8. biscuits=t fruit=t total=high 954 ==> bread and cake=t 866    conf:(0.91)
9. frozen foods=t fruit=t vegetables=t total=high 834 ==> bread and cake=t 757    conf:(0.91)
10. frozen foods=t fruit=t total=high 969 ==> bread and cake=t 877    conf:(0.91)
```

Observations :

- We can observe that minimum support is given as 0.15 that is 694 instances.
- It means that for Apriori algorithm it attempts to generate 10 rules by starting with the minimum support of 100% and iteratively decreasing the support by delta factor until it reaches minimum non zero support or the required number of rules with the least minimum confidence.
- By specifying minimum metric as a constraint we can say that a minimum support of 0.15 hints the minimum support reached in the process of generating the 10 rules.
- We can also observe that a list of rules are generated by the algorithm along with the itemsets.
- The number preceding the → indicates that the number of cases covered by the list of item sets on the left-hand side (LHS) and the value following the rule is the number of cases covered by the item sets on the right hand side (RHS).
- In the first best rule LHS consists of attributes fruit, frozen food, biscuits, high total and have a value of 788. RHS consists of attributes bread and cake and a value of 723. In this case the confidence is 0.92.

- In the second best rule LHS consists of attributes fruit, baking needs, biscuits, high total and have a value of 760. RHS consists of attributes bread and cake and a value of 696. In this case the confidence is 0.92.
- In the third best rule LHS consists of attributes fruit, baking needs, frozen food, high total and have a value of 770. RHS consists of attributes bread and cake and a value of 705. In this case the confidence is 0.92.
- In the fourth best rule LHS consists of attributes fruit, vegetables, biscuits, high total and have a value of 815. RHS consists of attributes bread and cake and a value of 746. In this case the confidence is 0.92.

➤ **Modifying Support and Confidence Values :**

Considering,

Low Support Value = 0.1

Medium Support Value = 0.5

High Support Value = 0.9

- For the same data set, experiments were done by taking low, medium and high values for both support and confidence.
- 9 different observations were seen by taking different pairs of support and confidence values. The observations can be seen below.

➤ **Considering Low Support Value :**

- Support : 0.1
- Confidence : 0.1

Associator Model :

```
Apriori
=====

Minimum support: 0.45 (2082 instances)
Minimum metric <confidence>: 0.1
Number of cycles performed: 11

Generated sets of large itemsets:

Size of set of large itemsets L(1): 13
Size of set of large itemsets L(2): 7
```

Best Rules Found :

Best rules found:

```
1. biscuits=t 2605 ==> bread and cake=t 2083    conf:(0.8)
2. milk-cream=t 2939 ==> bread and cake=t 2337    conf:(0.8)
3. fruit=t 2962 ==> bread and cake=t 2325    conf:(0.78)
4. baking needs=t 2795 ==> bread and cake=t 2191    conf:(0.78)
5. frozen foods=t 2717 ==> bread and cake=t 2129    conf:(0.78)
6. vegetables=t 2961 ==> bread and cake=t 2298    conf:(0.78)
7. vegetables=t 2961 ==> fruit=t 2207    conf:(0.75)
8. fruit=t 2962 ==> vegetables=t 2207    conf:(0.75)
9. bread and cake=t 3330 ==> milk-cream=t 2337    conf:(0.7)
10. bread and cake=t 3330 ==> fruit=t 2325    conf:(0.7)
```

Observations :

- Both support and confidence value were taken as 0.1, a low value.
- The number of cycles dropped to 11.
- The number of generated set of large item sets dropped to 2.
- Taking the best rule the value for both the left hand side and the right hand side have increased with a very large margin.
- The top rule has the LHS value of 2605 and RHS value of 2083.
- Also the value of confidence is 0.8.

For,

- Support : 0.1
- Confidence : 0.5

Associator Model :

```
Apriori
=====

Minimum support: 0.4 (1851 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 12

Generated sets of large itemsets:

Size of set of large itemsets L(1): 15
Size of set of large itemsets L(2): 9
```

Best Rules Found :

Best rules found:

```
1. milk-cream=t 2939 ==> bread and cake=t 2337    conf:(0.8)
2. baking needs=t 2795 ==> bread and cake=t 2191    conf:(0.78)
3. frozen foods=t 2717 ==> bread and cake=t 2129    conf:(0.78)
4. vegetables=t 2961 ==> bread and cake=t 2298    conf:(0.78)
5. juice-sat-cord-ms=t 2463 ==> bread and cake=t 1869    conf:(0.76)
6. bread and cake=t 3330 ==> milk-cream=t 2337    conf:(0.7)
7. baking needs=t 2795 ==> vegetables=t 1949    conf:(0.7)
8. frozen foods=t 2717 ==> vegetables=t 1882    conf:(0.69)
9. bread and cake=t 3330 ==> vegetables=t 2298    conf:(0.69)
10. milk-cream=t 2939 ==> vegetables=t 2025    conf:(0.69)
```

Observations :

- Here, the number of cycles is 12 which is one more than the above case when support and confidence was equal to 0.1.
- Also, the number of generated set of large item sets is 2.
- Taking the best rule the value for both the left hand side and the right hand side have increased with a very large margin.
- The top rule has the LHS value of 2939 and RHS value of 2337.
- Also the value of confidence is 0.8.

For,

- Support : 0.1
- Confidence : 0.9

Associator Model :


```
Apriori
=====
```

```
Minimum support: 0.15 (694 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 17
```

```
Generated sets of large itemsets:
```

```
Size of set of large itemsets L(1): 44
```

```
Size of set of large itemsets L(2): 380
```

```
Size of set of large itemsets L(3): 910
```

```
Size of set of large itemsets L(4): 633
```

```
Size of set of large itemsets L(5): 105
```

```
Size of set of large itemsets L(6): 1
```

Best Rules Found :

```
Best rules found:
```

```
1. biscuits=t frozen foods=t fruit=t total=high 788 ==> bread and cake=t 723    conf:(0.92)
2. baking needs=t biscuits=t fruit=t total=high 760 ==> bread and cake=t 696    conf:(0.92)
3. baking needs=t frozen foods=t fruit=t total=high 770 ==> bread and cake=t 705    conf:(0.92)
4. biscuits=t fruit=t vegetables=t total=high 815 ==> bread and cake=t 746    conf:(0.92)
5. party snack foods=t fruit=t total=high 854 ==> bread and cake=t 779    conf:(0.91)
6. biscuits=t frozen foods=t vegetables=t total=high 797 ==> bread and cake=t 725    conf:(0.91)
7. baking needs=t biscuits=t vegetables=t total=high 772 ==> bread and cake=t 701    conf:(0.91)
8. biscuits=t fruit=t total=high 954 ==> bread and cake=t 866    conf:(0.91)
9. frozen foods=t fruit=t vegetables=t total=high 834 ==> bread and cake=t 757    conf:(0.91)
10. frozen foods=t fruit=t total=high 969 ==> bread and cake=t 877    conf:(0.91)
```

Observations :

- Here, the number of cycles is 17.
- The number of generated set of large item sets is 6 which is more than the above two cases.
- The top rule has the LHS value of 788 and RHS value of 723.
- Also the value of confidence is 0.92.

➤ Considering Medium Support Value :

- Support : 0.5
- Confidence : 0.1

Associator Model :

```
Apriori
=====
```

```
Minimum support: 0.5 (2314 instances)
Minimum metric <confidence>: 0.1
Number of cycles performed: 10
```

```
Generated sets of large itemsets:
```

```
Size of set of large itemsets L(1): 10
```

```
Size of set of large itemsets L(2): 2
```

Best Rules Found :

```
Best rules found:
```

1. milk-cream=t 2939 ==> bread and cake=t 2337 conf:(0.8)
2. fruit=t 2962 ==> bread and cake=t 2325 conf:(0.78)
3. bread and cake=t 3330 ==> milk-cream=t 2337 conf:(0.7)
4. bread and cake=t 3330 ==> fruit=t 2325 conf:(0.7)

Observations :

- Here, the number of cycles is 10 .
- Also, the number of generated set of large item sets is 2.
- The top rule has the LHS value of 2939 and RHS value of 2337.
- Also the value of confidence is 0.8.

For,

- Support : 0.5
- Confidence : 0.5

Associator Model :

```
Apriori
=====
```

```
Minimum support: 0.5 (2314 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 10
```

```
Generated sets of large itemsets:
```

```
Size of set of large itemsets L(1): 7
```

```
Size of set of large itemsets L(2): 1
```

Best Rules Found :

Best rules found:

1. milk-cream=t 2939 ==> bread and cake=t 2337 conf:(0.8)
2. bread and cake=t 3330 ==> milk-cream=t 2337 conf:(0.7)

Observations :

- Here, the number of cycles is 10 .
- Also, the number of generated set of large item sets is 2.
- In this case, the number of cycles and the number of large item sets is same as the above one.
- The top rule has the LHS value of 2939 and RHS value of 2337.
- Also the value of confidence is 0.8.

For,

- Support : 0.5
- Confidence : 0.9

Associator Model :

```
Apriori
=====

Minimum support: 0.5 (2314 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 10

Generated sets of large itemsets:

Size of set of large itemsets L(1): 7
Size of set of large itemsets L(2): 1
```

Best Rules Found :

None

Observations :

- Here, the number of cycles is 10 .
- Also, the number of generated set of large item sets is 2.
- But in this case no best rules were produced.

➤ **Considering High Support Value :**

- Support : 0.9
- Confidence : 0.1/0.5/0.9

No Large itemsets and rules found.

➤ **After removing attributes :**

The attributes total, fruit and biscuits has been removed as shown in the below screen shots.

213	department213
214	department214
215	department215
216	department216
217	total

81	department81
82	produce misc
83	fruit
84	plants
85	potatoes

16	juice-sat-cord-ms
17	tea
18	biscuits
19	canned fish-meat
20	canned fruit

Associator Model :

```
Apriori
=====

Minimum support: 0.1 (463 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 48
Size of set of large itemsets L(2): 485
Size of set of large itemsets L(3): 1505
Size of set of large itemsets L(4): 1562
Size of set of large itemsets L(5): 495
Size of set of large itemsets L(6): 31
```

Best Rules Found :

Best rules found:

1. tissues-paper prd=t milk-cream=t beef=t vegetables=t 549 ==> bread and cake=t 498 conf:(0.91)
2. canned fruit=t tissues-paper prd=t milk-cream=t 519 ==> bread and cake=t 470 conf:(0.91)
3. baking needs=t pet foods=t milk-cream=t margarine=t 592 ==> bread and cake=t 534 conf:(0.9)
4. baking needs=t tissues-paper prd=t milk-cream=t beef=t 520 ==> bread and cake=t 469 conf:(0.9)
5. baking needs=t laundry needs=t tissues-paper prd=t milk-cream=t 547 ==> bread and cake=t 493 conf:(0.9)
6. pet foods=t milk-cream=t margarine=t vegetables=t 584 ==> bread and cake=t 526 conf:(0.9)
7. baking needs=t frozen foods=t milk-cream=t beef=t vegetables=t 521 ==> bread and cake=t 469 conf:(0.9)

Observations :

- After removing 3 attributes the minimum support decreased to 0.1 that is 463 instances but the number of cycles performed increased to 18.
- Also 6 large itemsets were generated.
- For the best rule generated left hand side went down to 549 for the same confidence value of 0.9.
- In the first best rule LHS consists of attributes tissues-paper prd, milk-cream, beef, vegetables and have a value of 549. RHS consists of attributes bread and cake and a value of 498. In this case the confidence is 0.91.
- In the second best rule LHS consists of attributes tissues-paper prd, milk-cream, canned fruit and have a value of 519. RHS consists of attributes bread and cake and a value of 470. In this case the confidence is 0.91.
- In the third best rule LHS consists of attributes baking needs, pet foods, milk-cream, margarine and have a value of 592. RHS consists of attributes bread and cake and a value of 534. In this case the confidence is 0.9.
- In the fourth best rule LHS consists of attributes baking needs, tissue-paper prd, milk-cream, beef and have a value of 520. RHS consists of attributes bread and cake and a value of 469. In this case the confidence is 0.9.

➤ **Dataset :Vote**

➤ **Using Apriori algorithm**

Associator model :

```
Apriori
=====

Minimum support: 0.45 (196 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 11

Generated sets of large itemsets:

Size of set of large itemsets L(1): 20
Size of set of large itemsets L(2): 17
Size of set of large itemsets L(3): 6
Size of set of large itemsets L(4): 1
```

Best Rules Found :

```
Best rules found:

1. adoption-of-the-budget-resolution=y physician-fee-freeze=n 219 ==> Class=democrat 219    conf:(1)
2. adoption-of-the-budget-resolution=y physician-fee-freeze=n aid-to-nicaraguan-contras=y 198 ==> Class=democrat 198    conf:(1)
3. physician-fee-freeze=n aid-to-nicaraguan-contras=y 211 ==> Class=democrat 210    conf:(1)
4. physician-fee-freeze=n education-spending=n 202 ==> Class=democrat 201    conf:(1)
5. physician-fee-freeze=n 247 ==> Class=democrat 245    conf:(0.99)
6. el-salvador-aid=n Class=democrat 200 ==> aid-to-nicaraguan-contras=y 197    conf:(0.99)
7. el-salvador-aid=n 208 ==> aid-to-nicaraguan-contras=y 204    conf:(0.98)
8. adoption-of-the-budget-resolution=y aid-to-nicaraguan-contras=y Class=democrat 203 ==> physician-fee-freeze=n 198    conf:(0.98)
9. el-salvador-aid=n aid-to-nicaraguan-contras=y 204 ==> Class=democrat 197    conf:(0.97)
10. aid-to-nicaraguan-contras=y Class=democrat 218 ==> physician-fee-freeze=n 210    conf:(0.96)
```

Observations :

- We can observe that minimum support is given as 0.45 that is 196 instances.
- In the first best rule LHS consists of attributes, adoption of the budget resolution, physician-fee-freeze and have a value of 219. RHS consists of attributes class and a value of 219. In this case the confidence is 1.
- In the second best rule LHS consists of adoption of the budget resolution, physician-fee-freeze , aid-to-nicaraguan-contras and have a value of 198. RHS consists of attributes class and a value of 198. In this case the confidence is 1.
- In the third best rule LHS consists of attributes physician-fee-freeze , aid-to-nicaraguan-contras and have a value of 211. RHS consists of attributes class and a value of 210. In this case the confidence is 1.
- In the fourth best rule LHS consists of attributes physician-fee-freeze, education-spending and have a value of 202. RHS consists of attributes class and a value of 201. In this case the confidence is 1.

- **After modifying support and confidence value :**
- **Considering Low Support Value :**
 - Support : 0.1
 - Confidence : 0.1

Associator model :

```
Apriori
=====
Minimum support: 0.5 (217 instances)
Minimum metric <confidence>: 0.1
Number of cycles performed: 10

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12
Size of set of large itemsets L(2): 4
Size of set of large itemsets L(3): 1
```

Best Rules Found :

```
Best rules found:

1. adoption-of-the-budget-resolution=y physician-fee-freeze=n 219 ==> Class=democrat 219    conf:(1)
2. physician-fee-freeze=n 247 ==> Class=democrat 245    conf:(0.99)
3. adoption-of-the-budget-resolution=y Class=democrat 231 ==> physician-fee-freeze=n 219    conf:(0.95)
4. Class=democrat 267 ==> physician-fee-freeze=n 245    conf:(0.92)
5. adoption-of-the-budget-resolution=y 253 ==> Class=democrat 231    conf:(0.91)
6. aid-to-nicaraguan-contras=y 242 ==> Class=democrat 218    conf:(0.9)
7. physician-fee-freeze=n Class=democrat 245 ==> adoption-of-the-budget-resolution=y 219    conf:(0.89)
8. physician-fee-freeze=n 247 ==> adoption-of-the-budget-resolution=y 219    conf:(0.89)
9. physician-fee-freeze=n 247 ==> adoption-of-the-budget-resolution=y Class=democrat 219    conf:(0.89)
10. adoption-of-the-budget-resolution=y 253 ==> physician-fee-freeze=n 219    conf:(0.87)
```

Observations :

- Both support and confidence value were taken as 0.1, a low value.
- The number of cycles dropped to 10.
- The number of generated set of large item sets dropped to 3.
- The top rule has the LHS value of 219 and RHS value of 219.
- Also the value of confidence is 1.

For,

- Support : 0.1
- Confidence : 0.5

Associator model :

```

Apriori
=====
Minimum support: 0.5 (217 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 10

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12
Size of set of large itemsets L(2): 4
Size of set of large itemsets L(3): 1

```

Best Rules Found :

```

Best rules found:

1. adoption-of-the-budget-resolution=y physician-fee-freeze=n 219 ==> Class=democrat 219    conf:(1)
2. physician-fee-freeze=n 247 ==> Class=democrat 245    conf:(0.99)
3. adoption-of-the-budget-resolution=y Class=democrat 231 ==> physician-fee-freeze=n 219    conf:(0.95)
4. Class=democrat 267 ==> physician-fee-freeze=n 245    conf:(0.92)
5. adoption-of-the-budget-resolution=y 253 ==> Class=democrat 231    conf:(0.91)
6. aid-to-nicaraguan-contras=y 242 ==> Class=democrat 218    conf:(0.9)
7. physician-fee-freeze=n Class=democrat 245 ==> adoption-of-the-budget-resolution=y 219    conf:(0.89)
8. physician-fee-freeze=n 247 ==> adoption-of-the-budget-resolution=y 219    conf:(0.89)
9. physician-fee-freeze=n 247 ==> adoption-of-the-budget-resolution=y Class=democrat 219    conf:(0.89)
10. adoption-of-the-budget-resolution=y 253 ==> physician-fee-freeze=n 219    conf:(0.87)

```

Observations :

- The number of cycles is 10.
- The number of generated set of large item sets dropped is 3.
- The top rule has the LHS value of 219 and RHS value of 219.
- Also the value of confidence is 1.

For,

- Support : 0.1
- Confidence : 0.9

Associator model :


```
Apriori
=====

Minimum support: 0.45 (196 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 11

Generated sets of large itemsets:

Size of set of large itemsets L(1): 20
Size of set of large itemsets L(2): 17
Size of set of large itemsets L(3): 6
Size of set of large itemsets L(4): 1
```

Best Rules Found :

```
Best rules found:

1. adoption-of-the-budget-resolution=y physician-fee-freeze=n 219 ==> Class=democrat 219    conf:(1)
2. adoption-of-the-budget-resolution=y physician-fee-freeze=n aid-to-nicaraguan-contras=y 198 ==> Class=democrat 198    conf:(1)
3. physician-fee-freeze=n aid-to-nicaraguan-contras=y 211 ==> Class=democrat 210    conf:(1)
4. physician-fee-freeze=n education-spending=n 202 ==> Class=democrat 201    conf:(1)
5. physician-fee-freeze=n 247 ==> Class=democrat 245    conf:(0.99)
6. el-salvador-aid=n Class=democrat 200 ==> aid-to-nicaraguan-contras=y 197    conf:(0.99)
7. el-salvador-aid=n 208 ==> aid-to-nicaraguan-contras=y 204    conf:(0.98)
8. adoption-of-the-budget-resolution=y aid-to-nicaraguan-contras=y Class=democrat 203 ==> physician-fee-freeze=n 198    conf:(0.98)
9. el-salvador-aid=n aid-to-nicaraguan-contras=y 204 ==> Class=democrat 197    conf:(0.97)
10. aid-to-nicaraguan-contras=y Class=democrat 218 ==> physician-fee-freeze=n 210    conf:(0.96)
```

Observations :

- The number of cycles is 11.
- The number of generated set of large item sets dropped is 4.
- The top rule has the LHS value of 219 and RHS value of 219.
- Also the value of confidence is 1.

Considering Medium Support Value :

- Support : 0.5
- Confidence : 0.1

Associator model :

```

Apriori
=====

Minimum support: 0.5 (218 instances)
Minimum metric <confidence>: 0.1
Number of cycles performed: 10

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12
Size of set of large itemsets L(2): 4
Size of set of large itemsets L(3): 1

```

Best Rules Found :

```

Best rules found:
1. adoption-of-the-budget-resolution=y physician-fee-freeze=n 219 ==> Class=democrat 219    conf:(1)
2. physician-fee-freeze=n 247 ==> Class=democrat 245    conf:(0.99)
3. adoption-of-the-budget-resolution=y Class=democrat 231 ==> physician-fee-freeze=n 219    conf:(0.95)
4. Class=democrat 267 ==> physician-fee-freeze=n 245    conf:(0.92)
5. adoption-of-the-budget-resolution=y 253 ==> Class=democrat 231    conf:(0.91)
6. aid-to-nicaraguan-contras=y 242 ==> Class=democrat 218    conf:(0.9)
7. physician-fee-freeze=n Class=democrat 245 ==> adoption-of-the-budget-resolution=y 219    conf:(0.89)
8. physician-fee-freeze=n 247 ==> adoption-of-the-budget-resolution=y 219    conf:(0.89)
9. physician-fee-freeze=n 247 ==> adoption-of-the-budget-resolution=y Class=democrat 219    conf:(0.89)
10. adoption-of-the-budget-resolution=y 253 ==> physician-fee-freeze=n 219    conf:(0.87)

```

Observations :

- The number of cycles is 10.
- The number of generated set of large item sets is 3.
- The top rule has the LHS value of 219 and RHS value of 219.
- Also the value of confidence is 1.

For,

- Support : 0.5
- Confidence : 0.5

Associator model :

```

Apriori
=====

Minimum support: 0.5 (218 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 10

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12
Size of set of large itemsets L(2): 4
Size of set of large itemsets L(3): 1

```

Best Rules Found :

```

Best rules found:

1. adoption-of-the-budget-resolution=y physician-fee-freeze=n 219 ==> Class=democrat 219    conf:(1)
2. physician-fee-freeze=n 247 ==> Class=democrat 245    conf:(0.99)
3. adoption-of-the-budget-resolution=y Class=democrat 231 ==> physician-fee-freeze=n 219    conf:(0.95)
4. Class=democrat 267 ==> physician-fee-freeze=n 245    conf:(0.92)
5. adoption-of-the-budget-resolution=y 253 ==> Class=democrat 231    conf:(0.91)
6. aid-to-nicaraguan-contras=y 242 ==> Class=democrat 218    conf:(0.9)
7. physician-fee-freeze=n Class=democrat 245 ==> adoption-of-the-budget-resolution=y 219    conf:(0.89)
8. physician-fee-freeze=n 247 ==> adoption-of-the-budget-resolution=y 219    conf:(0.89)
9. physician-fee-freeze=n 247 ==> adoption-of-the-budget-resolution=y Class=democrat 219    conf:(0.89)
10. adoption-of-the-budget-resolution=y 253 ==> physician-fee-freeze=n 219    conf:(0.87)

```

Observations :

- The number of cycles is 10.
- The number of generated set of large item sets is 3.
- The top rule has the LHS value of 219 and RHS value of 219.
- Also the value of confidence is 1.

For,

- Support : 0.5
- Confidence : 0.9

Associator model :

```
Apriori
=====

Minimum support: 0.5 (218 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 10

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12
Size of set of large itemsets L(2): 4
Size of set of large itemsets L(3): 1
```

Best Rules Found :

```
Best rules found:

1. adoption-of-the-budget-resolution=y physician-fee-freeze=n 219 ==> Class=democrat 219    conf:(1)
2. physician-fee-freeze=n 247 ==> Class=democrat 245    conf:(0.99)
3. adoption-of-the-budget-resolution=y Class=democrat 231 ==> physician-fee-freeze=n 219    conf:(0.95)
4. Class=democrat 267 ==> physician-fee-freeze=n 245    conf:(0.92)
5. adoption-of-the-budget-resolution=y 253 ==> Class=democrat 231    conf:(0.91)
6. aid-to-nicaraguan-contras=y 242 ==> Class=democrat 218    conf:(0.9)
```

Observations :

- The number of cycles is 10.
- The number of generated set of large item sets is 3.
- The top rule has the LHS value of 219 and RHS value of 219.
- Also the value of confidence is 1.

➤ Considering High Support Value :

- Support : 0.9
- Confidence : 0.1/0.5/0.9

No Large itemsets and rules found.

➤ **After Removing Attributes :**

Three attributes, physician-fee-freeze, el-salvador-aid and crime were deleted from the dataset

Associator model

```
Apriori
=====

Minimum support: 0.4 (174 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 12

Generated sets of large itemsets:

Size of set of large itemsets L(1): 22
Size of set of large itemsets L(2): 20
Size of set of large itemsets L(3): 8
Size of set of large itemsets L(4): 1
```

```
Best rules found:

1. aid-to-nicaraguan-contras=y education-spending=n 194 ==> Class=democrat 186    conf:(0.96)
2. adoption-of-the-budget-resolution=y anti-satellite-test-ban=y Class=democrat 184 ==> aid-to-nicaraguan-contras=y 176    conf:(0.96)
3. adoption-of-the-budget-resolution=y education-spending=n 201 ==> Class=democrat 192    conf:(0.96)
4. mx-missile=y Class=democrat 188 ==> aid-to-nicaraguan-contras=y 179    conf:(0.95)
5. anti-satellite-test-ban=y Class=democrat 200 ==> aid-to-nicaraguan-contras=y 189    conf:(0.95)
6. adoption-of-the-budget-resolution=y aid-to-nicaraguan-contras=y 215 ==> Class=democrat 203    conf:(0.94)
7. adoption-of-the-budget-resolution=y anti-satellite-test-ban=y aid-to-nicaraguan-contras=y 187 ==> Class=democrat 176    conf:(0.94)
8. aid-to-nicaraguan-contras=y mx-missile=y 192 ==> Class=democrat 179    conf:(0.93)
9. anti-satellite-test-ban=y aid-to-nicaraguan-contras=y Class=democrat 189 ==> adoption-of-the-budget-resolution=y 176    conf:(0.93)
10. aid-to-nicaraguan-contras=y Class=democrat 218 ==> adoption-of-the-budget-resolution=y 203    conf:(0.93)
```

Observations :

- In the first best rule LHS have a value of 194. RHS value of 186. In this case the confidence is 0.96.
- In the second best rule LHS have a value of 184. RHS consists of attributes class and a value of 176. In this case the confidence is 0.96.
- In the third best rule LHS have a value of 201. RHS consists of attributes class and a value of 192. In this case the confidence is 0.96.
- In the fourth best rule LHS have a value of 188. RHS a value of 179. In this case the confidence is 0.95.

➤ **Using FPGrowth Algorithm :**

➤ **Parameters and their Default Values of FPGrowth Algorithm :**

Parameter	Value
delta	0.05
findAllRulesForSupportLevel	False
lowerBoundMinSupport	0.1
maxNumberOfItems	-1
metricType	Confidence
minMetric	0.9
numRulesToFind	10
positiveIndex	2
rulesMustContain	
transactionsMustContain	
upperBoundMinSupport	1.0
useORForMustContainList	false

Parameters Description :

delta -- Iteratively decrease support by this factor. Reduces support until min support is reached or required number of rules has been generated.

findAllRulesForSupportLevel -- Find all rules that meet the lower bound on minimum support and the minimum metric constraint. Turning this mode on will disable the iterative support reduction procedure to find the specified number of rules.

lowerBoundMinSupport -- Lower bound for minimum support as a fraction or number of instances.

maxNumberOfItems -- The maximum number of items to include in frequent item sets. -1 means no limit.

metricType -- Set the type of metric by which to rank rules. Confidence is the proportion of the examples covered by the premise that are also covered by the consequence(Class association rules can only be mined using confidence). Lift is confidence divided by the proportion of all examples that are covered by the consequence. This is a measure of the importance of the association that is independent of support. Leverage is the proportion of additional examples covered by both the premise and consequence above those expected if the premise and consequence were independent of each other. The total number of examples that this represents is presented in brackets following the leverage. Conviction is another measure of departure from independence.

minMetric -- Minimum metric score. Consider only rules with scores higher than this value.

numRulesToFind -- The number of rules to output

positiveIndex -- Set the index of binary valued attributes that is to be considered the positive index. Has no effect for sparse data (in this case the first index (i.e. non-zero values) is always treated as positive. Also has no effect for unary valued attributes (i.e. when using the Weka Apriori-style format for market basket data, which uses missing value "?" to indicate absence of an item.

rulesMustContain -- Only print rules that contain these items. Provide a comma separated list of attribute names.

transactionsMustContain -- Limit input to FPGrowth to those transactions (instances) that contain these items. Provide a comma separated list of attribute names.

upperBoundMinSupport -- Upper bound for minimum support as a fraction or number of instances. Start iteratively decreasing minimum support from this value.

useORForMustContainList -- Use OR instead of AND for transactions/rules must contain lists.

➤ Dataset : Supermarket

➤ Using FPGrowth Association Algorithm

```
=== Run information ===  
  
Scheme:      weka.associations.FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1  
Relation:     supermarket  
Instances:    4627  
Attributes:   217  
[list of attributes omitted]
```

Best Rules Found :

```
FPGrowth found 16 rules (displaying top 10)  
  
1. [fruit=t, frozen foods=t, biscuits=t, total=high]: 788 ==> [bread and cake=t]: 723 <conf:(0.92)> lift:(1.27) lev:(0.03) conv:(3.35)  
2. [fruit=t, baking needs=t, biscuits=t, total=high]: 760 ==> [bread and cake=t]: 696 <conf:(0.92)> lift:(1.27) lev:(0.03) conv:(3.28)  
3. [fruit=t, baking needs=t, frozen foods=t, total=high]: 770 ==> [bread and cake=t]: 705 <conf:(0.92)> lift:(1.27) lev:(0.03) conv:(3.27)  
4. [fruit=t, vegetables=t, biscuits=t, total=high]: 815 ==> [bread and cake=t]: 746 <conf:(0.92)> lift:(1.27) lev:(0.03) conv:(3.26)  
5. [fruit=t, party snack foods=t, total=high]: 854 ==> [bread and cake=t]: 779 <conf:(0.91)> lift:(1.27) lev:(0.04) conv:(3.15)  
6. [vegetables=t, frozen foods=t, biscuits=t, total=high]: 797 ==> [bread and cake=t]: 725 <conf:(0.91)> lift:(1.26) lev:(0.03) conv:(3.06)  
7. [vegetables=t, baking needs=t, biscuits=t, total=high]: 772 ==> [bread and cake=t]: 701 <conf:(0.91)> lift:(1.26) lev:(0.03) conv:(3.01)  
8. [fruit=t, biscuits=t, total=high]: 954 ==> [bread and cake=t]: 866 <conf:(0.91)> lift:(1.26) lev:(0.04) conv:(3)  
9. [fruit=t, vegetables=t, frozen foods=t, total=high]: 834 ==> [bread and cake=t]: 757 <conf:(0.91)> lift:(1.26) lev:(0.03) conv:(3)  
10. [fruit=t, frozen foods=t, total=high]: 969 ==> [bread and cake=t]: 877 <conf:(0.91)> lift:(1.26) lev:(0.04) conv:(2.92)
```

Observations :

- In FPGrowth algorithm, the number of cycles and the number of large itemsets generated are not considered.
- The only outcome of FPGrowth algorithm is the number of rules and the best rules found.

- Here in this case 16 rules are generated, out of which 10 are displayed.
- In the first best rule LHS consists of attributes fruit, frozen food, biscuits, high total and have a value of 788. RHS consists of attributes bread and cake and a value of 723. In this case the confidence is 0.92, lift is 1.27, leverage is 0.03 and conviction is 3.35.
- In the second best rule LHS consists of attributes fruit, baking needs, biscuits, high total and have a value of 760. RHS consists of attributes bread and cake and a value of 696. In this case the confidence is 0.92, lift is 1.27, leverage is 0.03 and conviction is 3.28.
- In the third best rule LHS consists of attributes fruit, baking needs, frozen food, high total and have a value of 770. RHS consists of attributes bread and cake and a value of 705. In this case the confidence is 0.92, lift is 1.27, leverage is 0.03 and conviction is 3.27.
- In the fourth best rule LHS consists of attributes fruit, vegetables, biscuits, high total and have a value of 815. RHS consists of attributes bread and cake and a value of 746. In this case the confidence is 0.92, lift is 1.27, leverage is 0.03 and conviction is 3.26.

Modifying Support/Confidence Values :

Considering Low Support Value :

- Support : 0.1
- Confidence : 0.1

Best Rules Found :

```

FPGrowth found 14 rules (displaying top 10)

1. [biscuits=t]: 2605 ==> [bread and cake=t]: 2083 <conf:(0.8)> lift:(1.11) lev:(0.04) conv:(1.4)
2. [milk-cream=t]: 2939 ==> [bread and cake=t]: 2337 <conf:(0.8)> lift:(1.1) lev:(0.05) conv:(1.37)
3. [fruit=t]: 2962 ==> [bread and cake=t]: 2325 <conf:(0.78)> lift:(1.09) lev:(0.04) conv:(1.3)
4. [baking needs=t]: 2795 ==> [bread and cake=t]: 2191 <conf:(0.78)> lift:(1.09) lev:(0.04) conv:(1.29)
5. [frozen foods=t]: 2717 ==> [bread and cake=t]: 2129 <conf:(0.78)> lift:(1.09) lev:(0.04) conv:(1.29)
6. [vegetables=t]: 2961 ==> [bread and cake=t]: 2298 <conf:(0.78)> lift:(1.08) lev:(0.04) conv:(1.25)
7. [vegetables=t]: 2961 ==> [fruit=t]: 2207 <conf:(0.75)> lift:(1.16) lev:(0.07) conv:(1.41)
8. [fruit=t]: 2962 ==> [vegetables=t]: 2207 <conf:(0.75)> lift:(1.16) lev:(0.07) conv:(1.41)
9. [bread and cake=t]: 3330 ==> [milk-cream=t]: 2337 <conf:(0.7)> lift:(1.1) lev:(0.05) conv:(1.22)
10. [bread and cake=t]: 3330 ==> [fruit=t]: 2325 <conf:(0.7)> lift:(1.09) lev:(0.04) conv:(1.19)

```

Observation :

- In this case, 14 rules are generated.
- The top rule has the LHS value of 2605 and RHS value of 2083.
- Also the value of confidence is 0.8, lift is 1.11, leverage is 0.04 and conviction is 1.4.

For,

- Support : 0.1

- Confidence : 0.5

Best Rules Found :

```
FPGrowth found 14 rules (displaying top 10)

1. [biscuits=t]: 2605 ==> [bread and cake=t]: 2083 <conf:(0.8)> lift:(1.11) lev:(0.04) conv:(1.4)
2. [milk-cream=t]: 2939 ==> [bread and cake=t]: 2337 <conf:(0.8)> lift:(1.1) lev:(0.05) conv:(1.37)
3. [fruit=t]: 2962 ==> [bread and cake=t]: 2325 <conf:(0.78)> lift:(1.09) lev:(0.04) conv:(1.3)
4. [baking needs=t]: 2795 ==> [bread and cake=t]: 2191 <conf:(0.78)> lift:(1.09) lev:(0.04) conv:(1.29)
5. [frozen foods=t]: 2717 ==> [bread and cake=t]: 2129 <conf:(0.78)> lift:(1.09) lev:(0.04) conv:(1.29)
6. [vegetables=t]: 2961 ==> [bread and cake=t]: 2298 <conf:(0.78)> lift:(1.08) lev:(0.04) conv:(1.25)
7. [vegetables=t]: 2961 ==> [fruit=t]: 2207 <conf:(0.75)> lift:(1.16) lev:(0.07) conv:(1.41)
8. [fruit=t]: 2962 ==> [vegetables=t]: 2207 <conf:(0.75)> lift:(1.16) lev:(0.07) conv:(1.41)
9. [bread and cake=t]: 3330 ==> [milk-cream=t]: 2337 <conf:(0.7)> lift:(1.1) lev:(0.05) conv:(1.22)
10. [bread and cake=t]: 3330 ==> [fruit=t]: 2325 <conf:(0.7)> lift:(1.09) lev:(0.04) conv:(1.19)
```

Observation :

- In this case, 14 rules are generated.
- The top rule has the LHS value of 2605 and RHS value of 2083.
- Also the value of confidence is 0.8, lift is 1.11, leverage is 0.04 and conviction is 1.4.

For,

- Support : 0.1
- Confidence : 0.9

Best Rules Found :

```
FPGrowth found 16 rules (displaying top 10)

1. [fruit=t, frozen foods=t, biscuits=t, total=high]: 788 ==> [bread and cake=t]: 723 <conf:(0.92)> lift:(1.27) lev:(0.03) conv:(3.35)
2. [fruit=t, baking needs=t, biscuits=t, total=high]: 760 ==> [bread and cake=t]: 696 <conf:(0.92)> lift:(1.27) lev:(0.03) conv:(3.28)
3. [fruit=t, baking needs=t, frozen foods=t, total=high]: 770 ==> [bread and cake=t]: 705 <conf:(0.92)> lift:(1.27) lev:(0.03) conv:(3.27)
4. [fruit=t, vegetables=t, biscuits=t, total=high]: 815 ==> [bread and cake=t]: 746 <conf:(0.92)> lift:(1.27) lev:(0.03) conv:(3.26)
5. [fruit=t, party snack foods=t, total=high]: 854 ==> [bread and cake=t]: 779 <conf:(0.91)> lift:(1.27) lev:(0.04) conv:(3.15)
6. [vegetables=t, frozen foods=t, biscuits=t, total=high]: 797 ==> [bread and cake=t]: 725 <conf:(0.91)> lift:(1.26) lev:(0.03) conv:(3.06)
7. [vegetables=t, baking needs=t, biscuits=t, total=high]: 772 ==> [bread and cake=t]: 701 <conf:(0.91)> lift:(1.26) lev:(0.03) conv:(3.01)
8. [fruit=t, biscuits=t, total=high]: 954 ==> [bread and cake=t]: 866 <conf:(0.91)> lift:(1.26) lev:(0.04) conv:(3)
9. [fruit=t, vegetables=t, frozen foods=t, total=high]: 834 ==> [bread and cake=t]: 757 <conf:(0.91)> lift:(1.26) lev:(0.03) conv:(3)
10. [fruit=t, frozen foods=t, total=high]: 969 ==> [bread and cake=t]: 877 <conf:(0.91)> lift:(1.26) lev:(0.04) conv:(2.92)
```

Observation :

- In this case, 16 rules are generated.
- The top rule has the LHS value of 788 and RHS value of 723.
- Also the value of confidence is 0.92, lift is 1.27, leverage is 0.03 and conviction is 3.35.

Considering Medium Support Value :

- Support : 0.5
- Confidence : 0.1

Best Rules Found :

```
FPGrowth found 4 rules (displaying top 4)

1. [milk-cream=t]: 2939 ==> [bread and cake=t]: 2337 <conf:(0.8)> lift:(1.1) lev:(0.05) conv:(1.37)
2. [fruit=t]: 2962 ==> [bread and cake=t]: 2325 <conf:(0.78)> lift:(1.09) lev:(0.04) conv:(1.3)
3. [bread and cake=t]: 3330 ==> [milk-cream=t]: 2337 <conf:(0.7)> lift:(1.1) lev:(0.05) conv:(1.22)
4. [bread and cake=t]: 3330 ==> [fruit=t]: 2325 <conf:(0.7)> lift:(1.09) lev:(0.04) conv:(1.19)
```

Observation :

- In this case, only 4 rules are generated.
- The top rule has the LHS value of 2939 and RHS value of 2337.
- Also the value of confidence is 0.8, lift is 1.1, leverage is 0.05 and conviction is 1.37.

For,

- Support : 0.5
- Confidence : 0.5

Best Rules Found :

```
FPGrowth found 4 rules (displaying top 4)

1. [milk-cream=t]: 2939 ==> [bread and cake=t]: 2337 <conf:(0.8)> lift:(1.1) lev:(0.05) conv:(1.37)
2. [fruit=t]: 2962 ==> [bread and cake=t]: 2325 <conf:(0.78)> lift:(1.09) lev:(0.04) conv:(1.3)
3. [bread and cake=t]: 3330 ==> [milk-cream=t]: 2337 <conf:(0.7)> lift:(1.1) lev:(0.05) conv:(1.22)
4. [bread and cake=t]: 3330 ==> [fruit=t]: 2325 <conf:(0.7)> lift:(1.09) lev:(0.04) conv:(1.19)
```

Observation :

- In this case, only 4 rules are generated.
- The top rule has the LHS value of 2939 and RHS value of 2337.
- Also the value of confidence is 0.8, lift is 1.1, leverage is 0.05 and conviction is 1.37.

For,

- Support : 0.5
- Confidence : 0.9

Best Rules Found :

None

Considering High Support Values :

- Support : 0.9
- Confidence : 0.1/0.5/0.9

No best rules have been found.

➤ After Removing Attributes :

213	department213
214	department214
215	department215
216	department216
217	total

81	department81
82	produce misc
83	fruit
84	plants
85	potatoes

16	juice-sat-cord-ms
17	tea
18	biscuits
19	canned fish-meat
20	canned fruit

Best Rules Found :

FPGrowth found 7 rules (displaying top 7)	
1.	[vegetables=t, milk-cream=t, tissues-paper prd=t, beef=t]: 549 ==> [bread and cake=t]: 498 <conf:(0.91)> lift:(1.26) lev:(0.02) conv:(2.96)
2.	[milk-cream=t, tissues-paper prd=t, canned fruit=t]: 519 ==> [bread and cake=t]: 470 <conf:(0.91)> lift:(1.26) lev:(0.02) conv:(2.91)
3.	[milk-cream=t, baking needs=t, margarine=t, pet foods=t]: 592 ==> [bread and cake=t]: 534 <conf:(0.9)> lift:(1.25) lev:(0.02) conv:(2.81)
4.	[milk-cream=t, baking needs=t, tissues-paper prd=t, beef=t]: 520 ==> [bread and cake=t]: 469 <conf:(0.9)> lift:(1.25) lev:(0.02) conv:(2.8)
5.	[milk-cream=t, baking needs=t, tissues-paper prd=t, laundry needs=t]: 547 ==> [bread and cake=t]: 493 <conf:(0.9)> lift:(1.25) lev:(0.02) conv:(2.79)
6.	[vegetables=t, milk-cream=t, margarine=t, pet foods=t]: 584 ==> [bread and cake=t]: 526 <conf:(0.9)> lift:(1.25) lev:(0.02) conv:(2.77)
7.	[vegetables=t, milk-cream=t, baking needs=t, frozen foods=t, beef=t]: 521 ==> [bread and cake=t]: 469 <conf:(0.9)> lift:(1.25) lev:(0.02) conv:(2.76)

Observations :

- After removing 3 attributes only 7 rules are generated.

- In the first best rule LHS consists of attributes vegetables, milk-cream, tissues-paper prd, beef, and have a value of 549. RHS consists of attributes bread and cake and a value of 498. In this case the confidence is 0.91, lift is 1.26, leverage is 0.02 and conviction is 2.96.
- In the second best rule LHS consists of attributes milk-cream, tissues-paper prd, canned fruit and have a value of 519. RHS consists of attributes bread and cake and a value of 470. In this case the confidence is 0.91, lift is 1.26, leverage is 0.02 and conviction is 2.91.
- In the third best rule LHS consists of attributes milk-cream, baking needs, margarine, pet foods and have a value of 592. RHS consists of attributes bread and cake and a value of 534. In this case the confidence is 0.9, lift is 1.25, leverage is 0.02 and conviction is 2.81.
- In the fourth best rule LHS consists of attributes milk-cream, baking needs, tissues-paper prd, beef and have a value of 520. RHS consists of attributes bread and cake and a value of 469. In this case the confidence is 0.9, lift is 1.25, leverage is 0.02 and conviction is 2.8.

DataSet : Vote

Using FPGrowth Algorithm

Best Rules Found :

FPGrowth found 40 rules (displaying top 10)

```
1. [el-salvador-aid=y, Class=repblican]: 157 ==> [physician-fee-freeze=y]: 156 <conf:(0.99)> lift:(2.44) lev:(0.21) conv:(46.56)
2. [crime=y, Class=repblican]: 158 ==> [physician-fee-freeze=y]: 155 <conf:(0.98)> lift:(2.41) lev:(0.21) conv:(23.43)
3. [religious-groups-in-schools=y, physician-fee-freeze=y]: 160 ==> [el-salvador-aid=y]: 156 <conf:(0.98)> lift:(2) lev:(0.18) conv:(16.4)
4. [Class=repblican]: 168 ==> [physician-fee-freeze=y]: 163 <conf:(0.97)> lift:(2.38) lev:(0.22) conv:(16.61)
5. [adoption-of-the-budget-resolution=y, anti-satellite-test-ban=y, mx-missile=y]: 161 ==> [aid-to-nicaraguan-contras=y]: 155 <conf:(0.96)> lift:(1.73) lev:(0.15) conv:(10.45)
6. [physician-fee-freeze=y, Class=repblican]: 163 ==> [el-salvador-aid=y]: 156 <conf:(0.96)> lift:(1.96) lev:(0.18) conv:(10.45)
7. [religious-groups-in-schools=y, el-salvador-aid=y, superfund-right-to-sue=y]: 160 ==> [crime=y]: 153 <conf:(0.96)> lift:(1.68) lev:(0.14) conv:(8.6)
8. [el-salvador-aid=y, superfund-right-to-sue=y]: 170 ==> [crime=y]: 162 <conf:(0.95)> lift:(1.67) lev:(0.15) conv:(8.12)
9. [crime=y, physician-fee-freeze=y]: 168 ==> [el-salvador-aid=y]: 160 <conf:(0.95)> lift:(1.95) lev:(0.18) conv:(9.57)
10. [el-salvador-aid=y, physician-fee-freeze=y]: 168 ==> [crime=y]: 160 <conf:(0.95)> lift:(1.67) lev:(0.15) conv:(8.02)
```

Observation :

- Here in this case 40 rules are generated.
- In the first best rule LHS consists of attributes el-salvador-aid, class and have a value of 157. RHS consists of attributes physician-fee-freeze and a value of 156. In this case the confidence is 0.99, lift is 2.44, leverage is 0.21 and conviction is 46.56.
- In the second best rule LHS consists of attributes crime, class and have a value of 158. RHS consists of attributes physician-fee-freeze and a value of 155. In this case the confidence is 0.98, lift is 2.41, leverage is 0.21 and conviction is 23.43.
- In the third best rule LHS consists of attributes religious-groups-in-school, physician-fee-freeze and have a value of 160. RHS consists of attributes el-salvador-aid and a value of 156. In this case the confidence is 0.98, lift is 2, leverage is 0.18 and conviction is 16.4.

- In the fourth best rule LHS consists of attributes class and have a value of 168. RHS consists of attributes physician-fee-freeze and a value of 163. In this case the confidence is 0.97, lift is 2.38, leverage is 0.22 and conviction is 16.61.

➤ **After Modifying support and confidence value**

➤ **Considering Low support value :**

- Support : 0.1
- Confidence : 0.1

Best Rules Found :

```

FPGrowth found 10 rules (displaying top 10)
1. [el-salvador-aid=y]: 212 ==> [religious-groups-in-schools=y]: 197 <conf:(0.93)> lift:(1.49) lev:(0.15) conv:(4.96)
2. [aid-to-nicaraguan-contras=y]: 242 ==> [adoption-of-the-budget-resolution=y]: 215 <conf:(0.89)> lift:(1.53) lev:(0.17) conv:(3.62)
3. [anti-satellite-test-ban=y]: 239 ==> [aid-to-nicaraguan-contras=y]: 210 <conf:(0.88)> lift:(1.58) lev:(0.18) conv:(3.53)
4. [aid-to-nicaraguan-contras=y]: 242 ==> [anti-satellite-test-ban=y]: 210 <conf:(0.87)> lift:(1.58) lev:(0.18) conv:(3.3)
5. [crime=y]: 248 ==> [religious-groups-in-schools=y]: 214 <conf:(0.86)> lift:(1.38) lev:(0.14) conv:(2.66)
6. [adoption-of-the-budget-resolution=y]: 253 ==> [aid-to-nicaraguan-contras=y]: 215 <conf:(0.85)> lift:(1.53) lev:(0.17) conv:(2.88)
7. [anti-satellite-test-ban=y]: 239 ==> [adoption-of-the-budget-resolution=y]: 201 <conf:(0.84)> lift:(1.45) lev:(0.14) conv:(2.56)
8. [adoption-of-the-budget-resolution=y]: 253 ==> [anti-satellite-test-ban=y]: 201 <conf:(0.79)> lift:(1.45) lev:(0.14) conv:(2.15)
9. [religious-groups-in-schools=y]: 272 ==> [crime=y]: 214 <conf:(0.79)> lift:(1.38) lev:(0.14) conv:(1.98)
10. [religious-groups-in-schools=y]: 272 ==> [el-salvador-aid=y]: 197 <conf:(0.72)> lift:(1.49) lev:(0.15) conv:(1.83)

```

Observation :

- In this case, 10 rules are generated.
- The top rule has the LHS value of 212 and RHS value of 197.
- Also the value of confidence is 0.93, lift is 1.49, leverage is 0.15 and conviction is 4.96.

For,

- Support : 0.1
- Confidence : 0.5

Best Rules Found :

```

FPGrowth found 10 rules (displaying top 10)
1. [el-salvador-aid=y]: 212 ==> [religious-groups-in-schools=y]: 197 <conf:(0.93)> lift:(1.49) lev:(0.15) conv:(4.96)
2. [aid-to-nicaraguan-contras=y]: 242 ==> [adoption-of-the-budget-resolution=y]: 215 <conf:(0.89)> lift:(1.53) lev:(0.17) conv:(3.62)
3. [anti-satellite-test-ban=y]: 239 ==> [aid-to-nicaraguan-contras=y]: 210 <conf:(0.88)> lift:(1.58) lev:(0.18) conv:(3.53)
4. [aid-to-nicaraguan-contras=y]: 242 ==> [anti-satellite-test-ban=y]: 210 <conf:(0.87)> lift:(1.58) lev:(0.18) conv:(3.3)
5. [crime=y]: 248 ==> [religious-groups-in-schools=y]: 214 <conf:(0.86)> lift:(1.38) lev:(0.14) conv:(2.66)
6. [adoption-of-the-budget-resolution=y]: 253 ==> [aid-to-nicaraguan-contras=y]: 215 <conf:(0.85)> lift:(1.53) lev:(0.17) conv:(2.88)
7. [anti-satellite-test-ban=y]: 239 ==> [adoption-of-the-budget-resolution=y]: 201 <conf:(0.84)> lift:(1.45) lev:(0.14) conv:(2.56)
8. [adoption-of-the-budget-resolution=y]: 253 ==> [anti-satellite-test-ban=y]: 201 <conf:(0.79)> lift:(1.45) lev:(0.14) conv:(2.15)
9. [religious-groups-in-schools=y]: 272 ==> [crime=y]: 214 <conf:(0.79)> lift:(1.38) lev:(0.14) conv:(1.98)
10. [religious-groups-in-schools=y]: 272 ==> [el-salvador-aid=y]: 197 <conf:(0.72)> lift:(1.49) lev:(0.15) conv:(1.83)

```

Observations :

- In this case, 10 rules are generated.

- The top rule has the LHS value of 212 and RHS value of 197.
- Also the value of confidence is 0.93, lift is 1.49, leverage is 0.15 and conviction is 4.96.

For,

- Support : 0.1
- Confidence : 0.9

Best Rules Found :

```
FPGrowth found 40 rules (displaying top 10)

1. [el-salvador-aid=y, Class=republican]: 157 ==> [physician-fee-freeze=y]: 156 <conf:(0.99)> lift:(2.44) lev:(0.21) conv:(46.56)
2. [crime=y, Class=republican]: 158 ==> [physician-fee-freeze=y]: 155 <conf:(0.98)> lift:(2.41) lev:(0.21) conv:(23.43)
3. [religious-groups-in-schools=y, physician-fee-freeze=y]: 160 ==> [el-salvador-aid=y]: 156 <conf:(0.98)> lift:(2) lev:(0.18) conv:(16.4)
4. [Class=republican]: 168 ==> [physician-fee-freeze=y]: 163 <conf:(0.97)> lift:(2.38) lev:(0.22) conv:(16.61)
5. [adoption-of-the-budget-resolution=y, anti-satellite-test-ban=y, mx-missile=y]: 161 ==> [aid-to-nicaraguan-contras=y]: 155 <conf:(0.96)> lift:(1.73) lev:(0.15) conv:(8.6)
6. [physician-fee-freeze=y, Class=republican]: 163 ==> [el-salvador-aid=y]: 156 <conf:(0.96)> lift:(1.96) lev:(0.18) conv:(10.45)
7. [religious-groups-in-schools=y, el-salvador-aid=y, superfund-right-to-sue=y]: 160 ==> [crime=y]: 153 <conf:(0.96)> lift:(1.68) lev:(0.14) conv:(8.6)
8. [el-salvador-aid=y, superfund-right-to-sue=y]: 170 ==> [crime=y]: 162 <conf:(0.95)> lift:(1.67) lev:(0.15) conv:(8.12)
9. [crime=y, physician-fee-freeze=y]: 168 ==> [el-salvador-aid=y]: 160 <conf:(0.95)> lift:(1.95) lev:(0.18) conv:(9.57)
10. [el-salvador-aid=y, physician-fee-freeze=y]: 168 ==> [crime=y]: 160 <conf:(0.95)> lift:(1.67) lev:(0.15) conv:(8.02)
```

Observations:

- In this case, 40 rules are generated.
- The top rule has the LHS value of 157 and RHS value of 156.
- Also the value of confidence is 0.99, lift is 2.44, leverage is 0.21 and conviction is 46.56.

For,

- Support : 0.4
- Confidence : 0.1

Best Rules Found :

```
FPGrowth found 10 rules (displaying top 10)

1. [el-salvador-aid=y]: 212 ==> [religious-groups-in-schools=y]: 197 <conf:(0.93)> lift:(1.49) lev:(0.15) conv:(4.96)
2. [aid-to-nicaraguan-contras=y]: 242 ==> [adoption-of-the-budget-resolution=y]: 215 <conf:(0.89)> lift:(1.53) lev:(0.17) conv:(3.62)
3. [anti-satellite-test-ban=y]: 239 ==> [aid-to-nicaraguan-contras=y]: 210 <conf:(0.88)> lift:(1.58) lev:(0.18) conv:(3.53)
4. [aid-to-nicaraguan-contras=y]: 242 ==> [anti-satellite-test-ban=y]: 210 <conf:(0.87)> lift:(1.58) lev:(0.18) conv:(3.3)
5. [crime=y]: 248 ==> [religious-groups-in-schools=y]: 214 <conf:(0.86)> lift:(1.38) lev:(0.14) conv:(2.66)
6. [adoption-of-the-budget-resolution=y]: 253 ==> [aid-to-nicaraguan-contras=y]: 215 <conf:(0.85)> lift:(1.53) lev:(0.17) conv:(2.88)
7. [anti-satellite-test-ban=y]: 239 ==> [adoption-of-the-budget-resolution=y]: 201 <conf:(0.84)> lift:(1.45) lev:(0.14) conv:(2.56)
8. [adoption-of-the-budget-resolution=y]: 253 ==> [anti-satellite-test-ban=y]: 201 <conf:(0.79)> lift:(1.45) lev:(0.14) conv:(2.15)
9. [religious-groups-in-schools=y]: 272 ==> [crime=y]: 214 <conf:(0.79)> lift:(1.38) lev:(0.14) conv:(1.98)
10. [religious-groups-in-schools=y]: 272 ==> [el-salvador-aid=y]: 197 <conf:(0.72)> lift:(1.49) lev:(0.15) conv:(1.83)
```

Observations :

- In this case, 10 rules are generated.

- The top rule has the LHS value of 212 and RHS value of 197.
- Also the value of confidence is 0.93, lift is 1.49, leverage is 0.15 and conviction is 4.96.

For,

- Support : 0.4
- Confidence : 0.5

Best Rules Found :

```
FPGrowth found 10 rules (displaying top 10)
1. [el-salvador-aid=y]: 212 ==> [religious-groups-in-schools=y]: 197 <conf:(0.93)> lift:(1.49) lev:(0.15) conv:(4.96)
2. [aid-to-nicaraguan-contras=y]: 242 ==> [adoption-of-the-budget-resolution=y]: 215 <conf:(0.89)> lift:(1.53) lev:(0.17) conv:(3.62)
3. [anti-satellite-test-ban=y]: 239 ==> [aid-to-nicaraguan-contras=y]: 210 <conf:(0.88)> lift:(1.58) lev:(0.18) conv:(3.53)
4. [aid-to-nicaraguan-contras=y]: 242 ==> [anti-satellite-test-ban=y]: 210 <conf:(0.87)> lift:(1.58) lev:(0.18) conv:(3.3)
5. [crime=y]: 248 ==> [religious-groups-in-schools=y]: 214 <conf:(0.86)> lift:(1.38) lev:(0.14) conv:(2.66)
6. [adoption-of-the-budget-resolution=y]: 253 ==> [aid-to-nicaraguan-contras=y]: 215 <conf:(0.85)> lift:(1.53) lev:(0.17) conv:(2.88)
7. [anti-satellite-test-ban=y]: 239 ==> [adoption-of-the-budget-resolution=y]: 201 <conf:(0.84)> lift:(1.45) lev:(0.14) conv:(2.56)
8. [adoption-of-the-budget-resolution=y]: 253 ==> [anti-satellite-test-ban=y]: 201 <conf:(0.79)> lift:(1.45) lev:(0.14) conv:(2.15)
9. [religious-groups-in-schools=y]: 272 ==> [crime=y]: 214 <conf:(0.79)> lift:(1.38) lev:(0.14) conv:(1.98)
10. [religious-groups-in-schools=y]: 272 ==> [el-salvador-aid=y]: 197 <conf:(0.72)> lift:(1.49) lev:(0.15) conv:(1.83)
```

Observations :

- In this case, 10 rules are generated.
- The top rule has the LHS value of 212 and RHS value of 197.
- Also the value of confidence is 0.93, lift is 1.49, leverage is 0.15 and conviction is 4.96.

For,

- Support : 0.4
- Confidence : 0.9

Best Rules Found :

```
FPGrowth found 6 rules (displaying top 6)
1. [adoption-of-the-budget-resolution=y, anti-satellite-test-ban=y]: 201 ==> [aid-to-nicaraguan-contras=y]: 187 <conf:(0.93)> lift:(1.67) lev:(0.17) conv:(5.95)
2. [el-salvador-aid=y]: 212 ==> [religious-groups-in-schools=y]: 197 <conf:(0.93)> lift:(1.49) lev:(0.15) conv:(4.96)
3. [crime=y, el-salvador-aid=y]: 194 ==> [religious-groups-in-schools=y]: 180 <conf:(0.93)> lift:(1.48) lev:(0.13) conv:(4.85)
4. [mx-missile=y]: 207 ==> [aid-to-nicaraguan-contras=y]: 192 <conf:(0.93)> lift:(1.67) lev:(0.18) conv:(5.74)
5. [el-salvador-aid=y]: 212 ==> [crime=y]: 194 <conf:(0.92)> lift:(1.61) lev:(0.17) conv:(4.8)
6. [religious-groups-in-schools=y, el-salvador-aid=y]: 197 ==> [crime=y]: 180 <conf:(0.91)> lift:(1.6) lev:(0.16) conv:(4.7)
```

Observations :

- In this case, 6 rules are generated.
- The top rule has the LHS value of 201 and RHS value of 187.
- Also the value of confidence is 0.93, lift is 1.67, leverage is 0.17 and conviction is 5.95.

For,

- Support : 0.5/0.9
- Confidence : 0.1

Best Values for support value of 0.5 and 0.9 for any value of confidence, **result not found**

➤ **After Removing attributes :**

Three attributes, physician-fee-freeze, el-salvador-aid and crime were deleted from the dataset.

Best Found Rules :

1. [adoption-of-the-budget-resolution=y, anti-satellite-test-ban=y, mx-missile=y]: 161 => [aid-to-nicaraguan-contras=y]: 155 <conf:(0.96)> lift:(1.73) lev:(0.15) conv:
2. [anti-satellite-test-ban=y, mx-missile=y]: 182 => [aid-to-nicaraguan-contras=y]: 173 <conf:(0.95)> lift:(1.71) lev:(0.16) conv:(8.07)
3. [adoption-of-the-budget-resolution=y, mx-missile=y]: 180 => [aid-to-nicaraguan-contras=y]: 171 <conf:(0.95)> lift:(1.71) lev:(0.16) conv:(7.99)
4. [superfund-right-to-sue=y, education-spending=y]: 139 => [religious-groups-in-schools=y]: 132 <conf:(0.95)> lift:(1.52) lev:(0.1) conv:(6.51)
5. [aid-to-nicaraguan-contras=y, handicapped-infants=y]: 146 => [adoption-of-the-budget-resolution=y]: 137 <conf:(0.94)> lift:(1.61) lev:(0.12) conv:(6.11)
6. [anti-satellite-test-ban=y, duty-free-exports=y]: 145 => [aid-to-nicaraguan-contras=y]: 135 <conf:(0.93)> lift:(1.67) lev:(0.12) conv:(5.85)
7. [adoption-of-the-budget-resolution=y, anti-satellite-test-ban=y]: 201 => [aid-to-nicaraguan-contras=y]: 187 <conf:(0.93)> lift:(1.67) lev:(0.17) conv:(5.95)
8. [education-spending=y]: 171 => [religious-groups-in-schools=y]: 159 <conf:(0.93)> lift:(1.49) lev:(0.12) conv:(4.93)
9. [mx-missile=y]: 207 => [aid-to-nicaraguan-contras=y]: 192 <conf:(0.93)> lift:(1.67) lev:(0.18) conv:(5.74)
10. [export-administration-act-south-africa=y, mx-missile=y]: 148 => [aid-to-nicaraguan-contras=y]: 137 <conf:(0.93)> lift:(1.66) lev:(0.13) conv:(5.47)

Observations :

- In the first best rule LHS consists of attributes adoption-of-the-budget-resolution, anti-satellite-test-ban, mx-missile and have a value of 161. RHS consists of attributes aid-to-nicaraguan-contras and a value of 155. In this case the confidence is 0.96, lift is 1.73, leverage is 0.15.
- In the second best rule LHS consists of attributes anti-satellite-test-ban, mx-missile and have a value of 182. RHS consists of attributes aid-to-nicaraguan-contras and a value of 173. In this case the confidence is 0.95, lift is 1.71, leverage is 0.16 and conviction is 8.07.
- In the third best rule LHS consists of attributes adoption-of-the-budget-resolution, mx-missile and have a value of 180. RHS consists of attributes aid-to-nicaraguan-contras and a value of 171. In this case the confidence is 0.95, lift is 1.71, leverage is 0.16 and conviction is 7.99.
- In the fourth best rule LHS consists of attributes superfund-right-to-sue, education-spending and have a value of 139. RHS consists of attributes religious-groups-in-school and a value of 132. In this case the confidence is 0.95, lift is 1.52, leverage is 0.1 and conviction is 6.51.

Comparing the association rules to the SimpleCart decision tree For Vote Data set :

Associator Model :


```
Apriori
=====

Minimum support: 0.45 (196 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 11

Generated sets of large itemsets:

Size of set of large itemsets L(1): 20
Size of set of large itemsets L(2): 17
Size of set of large itemsets L(3): 6
Size of set of large itemsets L(4): 1
```

Best Found Rules :

```
Best rules found:

1. adoption-of-the-budget-resolution=y physician-fee-freeze=n 219 ==> Class=democrat 219    conf:(1)
2. adoption-of-the-budget-resolution=y physician-fee-freeze=n aid-to-nicaraguan-contras=y 198 ==> Class=democrat 198    conf:(1)
3. physician-fee-freeze=n aid-to-nicaraguan-contras=y 211 ==> Class=democrat 210    conf:(1)
4. physician-fee-freeze=n education-spending=n 202 ==> Class=democrat 201    conf:(1)
5. physician-fee-freeze=n 247 ==> Class=democrat 245    conf:(0.99)
6. el-salvador-aid=n Class=democrat 200 ==> aid-to-nicaraguan-contras=y 197    conf:(0.99)
7. el-salvador-aid=n 208 ==> aid-to-nicaraguan-contras=y 204    conf:(0.98)
8. adoption-of-the-budget-resolution=y aid-to-nicaraguan-contras=y Class=democrat 203 ==> physician-fee-freeze=n 198    conf:(0.98)
9. el-salvador-aid=n aid-to-nicaraguan-contras=y 204 ==> Class=democrat 197    conf:(0.97)
10. aid-to-nicaraguan-contras=y Class=democrat 218 ==> physician-fee-freeze=n 210    conf:(0.96)
```

Observations :

- We can observe that minimum support is given as 0.45 that is 196 instances.
- In the first best rule LHS consists of attributes, adoption of the budget resolution, physician-fee-freeze and have a value of 219. RHS consists of attributes class and a value of 219. In this case the confidence is 1.
- In the second best rule LHS consists of adoption of the budget resolution, physician-fee-freeze , aid-to-nicaraguan-contras and have a value of 198. RHS consists of attributes class and a value of 198. In this case the confidence is 1.
- In the third best rule LHS consists of attributes physician-fee-freeze , aid-to-nicaraguan-contras and have a value of 211. RHS consists of attributes class and a value of 210. In this case the confidence is 1.
- In the fourth best rule LHS consists of attributes physician-fee-freeze, education-spending and have a value of 202. RHS consists of attributes class and a value of 201. In this case the confidence is 1.

Simple Cart Decision Tree Algorithm :

CART Decision Tree

```
physician-fee-freeze=(y)
| synfuels-corporation-cutback=(n): republican(141.7/4.0)
| synfuels-corporation-cutback!=(n)
| | mx-missile=(n)
| | | adoption-of-the-budget-resolution=(n): republican(19.28/3.31)
| | | adoption-of-the-budget-resolution!=(n)
| | | | anti-satellite-test-ban=(y): republican(2.2/0.0)
| | | | anti-satellite-test-ban!=(y): democrat(5.01/0.02)
| | | mx-missile!=(n): democrat(4.99/1.02)
| physician-fee-freeze!=(y): democrat(249.66/3.74)
```

- Number of leaf nodes : 6
- Size of the tree : 11

➤ Classification Accuracy :

- Total Number of Instances : 435
- Correctly Classified Instances : 415
- Incorrectly Classified Instances : 20
- Percentage of Correctly Classified Instances : 95.4023
- Percentage of Incorrectly Classified Instances : 4.5977

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.955	0.048	0.97	0.955	0.962	0.967	democrat
	0.952	0.045	0.93	0.952	0.941	0.967	republican
Weighted Avg.	0.954	0.047	0.954	0.954	0.954	0.967	

- Here the classification accuracy is around 95% which is remarkable.

➤ Confusion Matrix :

=== Confusion Matrix ===

```
a  b  <-- classified as
255 12 | a = democrat
 8 160 | b = republican
```

➤ Observation :

- For class democrat out of 267 instances, 255 instances are classified as democrat and the rest 12 are classified as republican.

- For class republican, out of 168 instances, 8 has been misclassified as democrat.

Conclusion :

- Best support and confidence for algorithm mainly depend on the dataset and its properties. Form the above observations done for different combination of support and confidence, it can be concluded that for apriori algorithm, number cycles performed decreases as support count increases also is the number of large itemset generate. For too large support the result may not be displayed. In case of confidence, default confidence gives lower LHS for the best rule. For FP growth it is similar to apriori, but from observations it can also be noticed that higher support and confidence decrease the number of rules.
- When some attributes are removed from dataset, the dataset behaves like new set. The Number of attributes and instances will be changed. the behavior of this data depends on dataset and attributes remained. form our observation it can be noticed that number of cycles performed increased for apriori but number of rules generated decreased for both and this depends on attributes.