

# Assignment A3 41403

- Title: Apply apriori algorithm to find frequently occurring items from given data.
- problem statement: Apply apriori algorithm to find frequently occurring items from given data and generate strong association rules using support and confidence threshold.  
e.g. market basket Analysis
- Objectives: To understand apriori algorithm & its application
- Outcomes: Students will be able to understand apriori algorithm & predict association rules.
- SW & HW req: Jupyter, Python
- Theory:  
Association rules: These rules help discover relationships b/w seemingly independent relational datasets or other data repositories.
  - support count: frequency of occurrence of itemset.
  - Association rule: An implication expression of the form  $x \Rightarrow y$  where  $x$  &  $y$  are  $\neq$  any itemset.
  - frequency itemset: An itemset whose values are greater than or equal to minsup threshold



## Rules Evaluation metrics:

- Support:  $(x+y) \div \text{total}$   
interpreted as fraction of transaction that contains both  $x$  &  $y$
- confidence:  
 $\text{supp}(x \cup y) \div \text{supp}(x)$   
measures how often each item in  $y$  appear in transaction that contain item in  $x$  also
- Lift (L):  
 $(x \Rightarrow y) = \text{conf}(x \Rightarrow y) \div \text{supp}(y)$   
Lift value near 1 indicates  $x$  &  $y$  almost often appear together as expected greater than 1 means they appear together more than expected & less than 1 means they appear less than expected. Greater lift indicates stronger association

## Apriori Algorithm:

- used for finding frequent itemset in a dataset for boolean association rule
- Name of the algorithm is apriori because it uses prior knowledge of frequent itemset properties



- An iterative approach or level wise search where frequent itemset are used to find  $k+1$  itemset

### Limitation of Apriori Algorithm

- slow
- Space required to hold large number of candidate sets
- Dataset used : Market basket
- Test case / Algorithms / Example

TID	Items
T <sub>1</sub>	I <sub>1</sub> , I <sub>2</sub> , I <sub>5</sub>
T <sub>2</sub>	I <sub>2</sub> , I <sub>4</sub>
T <sub>3</sub>	I <sub>1</sub> , I <sub>3</sub>
T <sub>4</sub>	I <sub>1</sub> , I <sub>2</sub> , I <sub>4</sub>
T <sub>5</sub>	I <sub>1</sub> , I <sub>3</sub>
T <sub>6</sub>	I <sub>2</sub> , I <sub>3</sub>
T <sub>7</sub>	I <sub>1</sub> , I <sub>3</sub>
T <sub>8</sub>	I <sub>1</sub> , I <sub>2</sub> , I <sub>3</sub> , I <sub>5</sub>
T <sub>9</sub>	I <sub>1</sub> , I <sub>2</sub> , I <sub>3</sub>

Support - 32%.      confidence 60%



41403

I	item	freq.	support %
	$I_1$	6	$6/9 = 0.67 = 67\%$
	$I_2$	7	$7/9 = 0.78 = 78\%$
	$I_3$	6	$6/9 = 0.67 = 67\%$
	$I_4$	2	$2/9 = 0.22 = 22\%$
	$I_5$	2	$2/9 = 0.22 = 22\%$

Select items on support  $\geq 32\%$ .  $I_1, I_2, I_3$

II	item	freq	support %
	$I_1 I_2$	4	$4/9 = 0.44 = 44\%$
	$I_1 I_3$	4	$4/9 = 0.44 = 44\%$
	$I_2 I_3$	4	$4/9 = 0.44 = 44\%$

Rule	Support	confidence	confidence
$I_1 \rightarrow I_2$	4	$4/6 = 0.67$	67
$I_2 \rightarrow I_3$	4	$4/7 = 0.57$	57
$I_1 \rightarrow I_3$	4	$4/6 = 0.67$	67
$I_3 \rightarrow I_1$	4	$4/6 = 0.67$	67
$I_2 \rightarrow I_3$	4	$4/7 = 0.57$	57
$I_3 \rightarrow I_2$	4	$4/6 = 0.67$	67

Association rules accepted

$I_1 \rightarrow I_2$ ,  $I_1 \rightarrow I_3$ ,  $I_3 \rightarrow I_1$ ,  $I_2 \rightarrow I_3$

- conclusion: Thus we have understood and implement apriori algorithm to find the association rule in dataset.