Big Data Analytics

ESSEC

Home work 6: Frequent Itemsets

1. (**Exercise 6.1.1 MMDS book)** Suppose there are 100 items, numbered 1 to 100, and also 100 baskets, also numbered 1 to 100. Item i is in basket b if and only if i divides b with no remainder. Thus, item 1 is in all the baskets, item 2 is in all fifty of the even-numbered baskets, and so on. Basket 12 consists of items $\{1, 2, 3, 4, 6, 12\}$, since these are all the integers that divide 12. Answer the following questions:

   (a) If the support threshold is 5, which items are frequent?

   (b) what is the confidence of the following association rules?

      i. $\{5, 7\} \rightarrow 2$

      ii. $\{2, 3, 4\} \rightarrow 5$

2. (**Exercise 6.1.3 MMDS book)** Suppose there are 100 items, numbered 1 to 100, and also 100 baskets, also numbered 1 to 100. Item i is in basket b if and only if b divides i with no remainder. For example, basket 12 consists of items

$$\{12, 24, 36, 48, 60, 72, 84, 96\}.$$

   Answer the following questions:

   (a) If the support threshold is 5, which items are frequent?

   (b) what is the confidence of the following association rule?

      i. $\{24, 60\} \rightarrow 8$

3. (**Apriori algorithm)** Apply the Apriori algorithm on the grocery store example with support threshold $s = 1/3$ and confidence threshold $c = 60\%$. Indicate the association rules that are generated and highlight the strong ones, sort them by confidence.

Transaction ID	Items
1	Milk, Bread, Juice
2	Milk, Bread
3	Milk, Coke, Chips
4	Chips, Coke
5	Chips, Juice
6	Milk, Coke, Chips