Data Mining

ENSIA 2025-2026

Lab sheet N°10: Association Rules

Exercise 1:

In market basket analysis, describe whether the following rules are subjectively interesting:

- A rule that has high support and high confidence (e.g. Milk → Bread)
- A rule that has reasonably high support but low confidence (e.g. Milk → Tuna)
- A rule that has low support and low confidence (e.g. Cooking oil → Laundry detergent)
- A rule that has low support and high confidence (e.g. Running Shoes → Performance Socks)

Exercise 2:

Consider the data set shown in the following Table:

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}

1. Compute the support for itemsets {e}, {b,d}, and {b,d,e} by treating each transaction ID as a market basket.
2. Use the results in question 1 to compute the confidence for the association rules {b,d} → {e} and {e} → {b,d}. Is confidence a symmetric measure?

3. Repeat question 1 by treating each customer ID as a market basket. Each item should be treated as a binary variable (1 if an item appears in at least one transaction bought by the customer, and 0 otherwise).
4. Use the results in question 3 to compute the confidence for the association rules {b,d} → {e} and {e} → {b,d}.
5. Suppose ***s1*** and ***c1*** are the support and confidence values of an association rule ***r*** when treating each transaction ID as a market basket. Also, let ***s2*** and ***c2*** be the support and confidence values of ***r*** when treating each customer ID as a market basket. Discuss whether there are any relationships between ***s1*** and ***s2*** or ***c1*** and ***c2***.

Exercise 3:

Consider the following set of frequent 3-itemsets:

{1, 2, 3}, {1, 2, 4}, {1, 2, 5}, {1, 3, 4}, {1, 3, 5}, {2, 3, 4}, {2, 3, 5}, {3, 4, 5}.

Assume that there are only five items in the dataset.

1. List all candidate 4-itemsets obtained by a candidate generation procedure using the $F_{K-1} \times F_1$ merging strategy.
2. List all candidate 4-itemsets obtained by the candidate generation procedure in the Apriori algorithm ($F_{K-1} \times F_{K-1}$).
3. List all candidate 4-itemsets that survive the candidate pruning step of the Apriori algorithm.

Exercise 4:

Consider the market basket transactions shown in the following table.

1. What is the maximum number of association rules that can be extracted from this data (including rules that have zero support)?
2. What is the maximum size of frequent itemsets that can be extracted (assuming ***minsup*** > 0)

Transaction ID	Items Bought
1	Milk, Juice, Diapers
2	Bread, Butter, Milk

3	Milk, Diapers, Cookies
4	Bread, Butter, Cookies
5	Juice, Cookies, Diapers
6	Milk, Diapers, Bread, Butter
7	Bread, Butter, Diapers
8	Juice, Diapers
9	Milk, Diapers, Bread, Butter
10	Juice, Cookies

3. Write an expression for the maximum number of size-3 itemsets that can be derived from this dataset.
4. Find an itemset (of size 2 or larger) that has the largest support.
5. Find a pair of items, ***a*** and ***b***, such that the rules { ***a*** } → { ***b*** } and { ***b*** } → { ***a*** } have the same confidence.

Exercise 5:

1. What is the confidence for the rules ∅ → A and A → ∅ ?
2. Let ***c1***, ***c2***, and ***c3*** be the confidence values of the rules { p } → { q } , { p } → { q, r }, and { p, r } → { q }, respectively. If we assume that ***c1***, ***c2***, and ***c3*** have different values, what are the possible relationships that may exist among ***c1***, ***c2***, and ***c3***? Which rule has the lowest confidence?
3. Repeat the analysis in question 2, assuming the rules have identical support. Which rule has the highest confidence?
4. Transitivity: Suppose the confidence of the rules A → B and B → C are larger than some threshold, ***minconf***. Is it possible that A → C has a confidence less than ***minconf***?