Norwegian University of Science and Technology

TDT4300 — **Assignment 2**

ASSOCIATION ANALYSIS

tdt4300-undass@idi.ntnu.no

Spring 2023

1 Apriori Algorithm

Given the data in Table 1 (market basket transaction), your task is to describe the purchasing behavior of customers in the form of association rules. First, you need to generate the frequent itemsets and second, you need to generate the association rules. Apply the Apriori algorithm for following tasks and **describe thoroughly the process and the outcome of each step.**

(a) Generate frequent 2-, 3- and 4-itemsets using the $F_{k-1} \times F_{k-1}$ method. Consider the support threshold $minsup = 0.5$ and use the data presented in Table 1.

(b) Using the frequent 4-itemsets from the previous task, generate association rules. Consider the confidence threshold $minconf = 0.8$.

TID	**Items**
110	A, C, F, G, H
111	B, C, D, E, G
112	B, C, E, F, H
113	A, B, C, G
114	C, D, E, H
115	A, B, C, G, H
116	A, B, C, D, G, H
117	B, C, E, G
118	A, B, C, F, G, H
119	A, B, C, D, E, G, H

Table 1: Market basket transactions.

2 FP-Growth Algorithm

Use the Frequent Pattern Growth algorithm to discover the frequent itemsets in the given transaction dataset (Table 1). Consider the support threshold $minsup = 0.5$. Construct an FP-tree and mine the frequent itemsets by creating conditional (sub-)pattern bases. Use the table notation with columns: item, conditional pattern base, conditional FP-tree, frequent patterns generated. The recursive steps of the FP-Growth algorithm must be clearly captured using the aforementioned table notation. Sort items alphabetically in case of ties in the item support. **Describe thoroughly the process and the outcome of each step.**

3 KNIME

For this task you will need to install and use the KNIME[1] data analytics platform. You are given a file *market_basket_transactions.arff* which contains the very same transaction as in Table 1. Your task is to implement two simple workflows for mining association rules, one implementing Apriori algorithm and second implementing FP-Growth algorithm. Use the WEKA nodes both for Apriori and FP-Growth. Use the same parameters as in the previous tasks, e.i. $minsup = 0.5$ and $minconf = 0.8$. **Present pictures of your workflows, and the outputs from both Apriori and FP-Growth nodes. Deliver also the exported KNIME workflows.**

Note: In Knime 4.X Weka nodes are not automatically included. Go to this link[2] and follow the directions under *Add to KNIME Analytics Platform* (drag the extension to your Workflow board in KNIME to install).

4 Compact Representation of Frequent Itemsets

Given the compact representation of frequent itemsets in Table 2, use the appropriate algorithm to generate all frequent itemsets including the support counts. **Describe thoroughly each step of the algorithm and present the final result.**

[1] `http://www.knime.org/downloads/overview`

[2] `https://hub.knime.com/knime/extensions/org.knime.features.ext.weka_3.7/latest`

Closed Frequent Itemsets	**Support Count**
{b}	10
{d}	13
{a, d}	11
{b, d}	7
{b, e}	8
{d, e}	6
{a, b, e}	7
{a, c, d}	6
{b, d, e}	4
{a, c, d, e}	5

Table 2: Closed frequent itemsets.

Submission Requirements

In this assignment we expect you to submit following artifacts:

- A PDF file with the report.
  - Text must not be handwritten.
  - Make sure that the document follows the usual conventions (**names**, assignment/task number, etc.).
- A ZIP file with KNIME workflows.

All assignment artifacts are to be delivered using ***BlackBoard***. You are allowed to **work in pairs**, however, the identical artifacts must be delivered individually.