

## Part 4 – Association rules

In this part, we use the file "supermarket.arff", which can be found in the Weka installation directory "\$WEKAHOME\data". The data set has 4627 transactions from a small New Zealander supermarket. Attributes from 1 to 216 are the departments of the supermarket, and the value t indicates that the customer has bought at least one product from the department in question. The last attribute is total, which can be either low or high.

What to report?

Answer the following questions and report the answers as well as the required Weka outputs in your exercise report.

1. What are association rules?
2. In association rule mining, what does support mean? How is confidence defined?
3. Give a description of the Apriori algorithm.
4. Analyze supermarket data for association rules between the products. Open "supermarket.arff" in Weka explorer. Select tab Associate and Apriori. In Apriori associator, set the parameter "car" to "True". It enables class association rules, i.e. you can decide the attribute on the right hand side of the association rules. The index of the chosen attribute is set to parameter "classIndex". Generate rules for three of the attributes: bread and cake ("classIndex" 13), milk-cream ("classIndex" 61), and tea ("classIndex" 17). The default value of the parameter "minMetric" is quite high so you can lower the value in order to mine less confident rules too. Study the mined rules, their support and confidence levels.

Report found rules and confidence values. Discuss the support and confidence values of the rules.

5. Discuss your findings. What kind of knowledge of the shopping habits are you able to mine from the data? If you were the shop owner, are there any actions you would take based on the mined knowledge?

Last modified: 10.4.2017 by Risto Vehmas