COMPSCI 361 – Assignment 3
Association Rule Mining

This is worth 5% of your final grade. Due Date: May 1, 2020, 23:59 NZT

Problem

Your task for this assignment is to investigate a dataset and perform a association rule mining task. You are given one dataset. You can download the data file from canvas under (A3). Each line within the data file represents a transaction. There are two files on Canvas supermarket.arff and supermarket.csv. They are both similar files but in different formats. You may select one of the files to use.

In the supermarket.csv, each transaction has multiple items (separated by a comma). Note also the transactions are of unequal lengths.

The assignment involves:

- Preparing and preprocessing the data
- Selecting an appropriate tool/algorithm to use
- Mining of the dataset, including appropriate parameter setting
- Determining which of the results are interesting based on “interestingness" measure(s).

You will need to describe your findings in a report. It should address the following points: (1) Description: What is in the data, and what pre-processing was done to help with task at hand (if any)? If any choices were made describe your reasoning behind any choices you made.

(2) Selection of tool: There are various implementations of rule mining techniques which can be used. These are normally available in implementations within WEKA, R, Python. In your report note the choice of tool selected.

(3) Mining: In-depth discussion of the parameter settings, and the time required. You must show that you have systematically chosen the parameter setting (e.g. evaluated a range of parameters, or developed an auxiliary function to determine the correct thresholds). As examples: you may discuss the different minimum values used for support and confidence through empirical and systematic parameter selection, i.e., trying the all minimum support levels at 0.1 intervals. (Please note this is just an example).

(4) Describing results: A summary of results (number of rules, general description), and a selection of rules that you deem is interesting based on “interestingness" measure(s), i.e., confidence, lift. Please specify the “interestingness" measure(s) you used. Note there are various number

of interesting rules that may be produced, please restrict the number of rules you discuss to a maximum number of 10 rules.

What to submit?

The final report has to be deposited to Canvas. Please name your report file **"Your_UPI.pdf"**. Your report should be no more than one page long. As a rough guide of page length, you may use font Times New Roman with size 12pt and single spacing. This includes any images or references you may choose to show or use.

Reproducible machine learning is one of the criteria of this assignment. So you need to report processing and parameters for recreating your results. You should include a description of the different runs (if you had carried out multiple run), and why you needed to make changes from your initial choices, in the report. One simple question that will indicate whether you have fulfilled minimum requirement of reproduciblity is "Can someone reproduce your results based on your explanation?".

Grading rubric

- Reproducible – 1 mark
- Discussion on how you selected your final parameters. – 1 mark
- Discussion on the results – 2 mark
- Overall quality of report, including readability/clarity – 1 mark

Copyright © University of Auckland. This material is provided to you for your own use. You may not copy or distribute any part of this material to any other person. Failure to comply with this warning may expose you to legal action for copyright infringement and/or disciplinary action by the University.