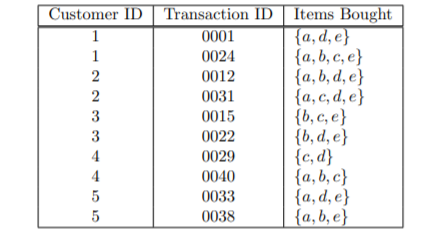
Data Mining Assignment 3

# Read Chapter 6 (only sections 6.1 and 6.7).

# 2) Do Chapter 6 textbook problem #2 (parts a,b,c,d only) on page 404.



# (a) Compute the support for itemsets {e}, {b,d}, and {b,d,e} by treating

# each transaction ID as a market basket.

**s({e}) = 8 /10 = 0.8**

**s({b, d}) = 2/ 10 = 0.2**

**s({b, d, e}) = 2/ 10 = 0.2**

# (b) Use the results in part (a) to compute the confidence for the associa-

# tion rules {b,d} ---, {e} and {"} ------+ {b,d}. Is confidence a symmetric

# measure?

**c(bd −→ e) = 0.2/ 0.2 = 100%**

**c(e −→ bd) = 0.2/ 0.8 = 25%**

# ( c ) Repeat part (a) by treating each customer ID as a market basket. Each

# item should be treated as a binary variable (1 if an item appears in atIeast one transaction bought by the customer, and 0 otherwise.)

**s({e}) = 4 /5 = 0.8**

**s({b, d}) = 5/ 5 = 1**

**s({b, d, e}) = 4 /5 = 0.8**

# (d) Use the results in part (c) to compute the confidence for the association

# rules {b, d} - {e} and {“} ------ {b,d,}.

**c(bd −→ e) = 0.8 /1 = 80%**

**c(e −→ bd) = 0.8/ 0.8 = 100%**

# 3) Do Chapter 6 textbook problem #6 (parts d,e only) on page 406.

# 

# (d) Find an itemset (of size 2 or larger) that has the largest support.

**{Bread, Butter}**

# (e) Find a pair of items, a and b, such that the rules {o} -, {b} and {b} -----+

# {a} have the same confidence.

# **(Beer, Cookies) or (Bread, Butter)** 4) Using the data at [www.stats202.com/more\_stats202\_logs.txt](http://www.stats202.com/more_stats202_logs.txt) and treating each row as a "market basket" compute the support and confidence for the rule ip=65.57.245.11 → "Mozilla/5.0 (X11; U; Linux i686 (x86\_64); en-US; rv:1.8.1.3) Gecko/20070309 Firefox/2.0.0.3".

# State what the support and confidence values mean in plain English in this context.