Assignment 8

Due: 4/1

Note: Show all your work.

Problem 1 (20 points). Consider the following transactional database.

TID	Items
100	1, 2, 3, 4, 5, 7
200	1, 3, 5, 6
300	1, 4, 5, 7, 8
400	1, 2, 3, 4, 5
500	2, 3, 4, 5, 7, 8

(1) Mine all frequent itemsets using Apriori. Show all candidate itemsets and frequent itemsets. You should follow the process described in the book and lecture (i.e., C1 → L1 → C2 → L2 → …). Minimum support = 60% (or 3 or more transactions). To save your time, L1 is given below:

L1:

Itemset	1	2	3	4	5	7
Count	4	3	4	4	5	3

(2) Sort all frequent 4-itemsets by their item number. Then, select the first frequent 4-itemset form the sorted list of frequent 4-itemsets and mine all strong rules from this itemset that have the format {W, X} => {Y, Z}, where W, X, Y, and Z are individual items. Assume that minimum confidence = 80%.

Problem 2 (10 points). Consider the following transactional database for sequential pattern mining.

CID	Day	Items
1	1	B, D, H
	14	A, C, D
	24	B, D, F
	31	E, F, G
2	4	A, B, G, H
	9	C, D, E, G
	14	C, D, H
3	1	B, G, H
	24	A, C, D, E
	51	A, D, G, H
4	2	B, G
	12	A, B, C, H
	25	B, C, D, E. G

Determine the supports of the following sequences:

<{H}, {B}>, <{A, C}, {E}>, <{C}, {D, G}>

Problem 3 (20 points). Consider the following contingency table.

	C (buys coffee = Yes)	$\overline{C}$ (buys coffee = No)
T (buys tea = Yes)	473	64
$\overline{T}$ (buys tea = No)	29	753

(1). Compute the *lift*, *all-confidence*, *cosine*, *Kulczynski* and *imbalance ratio* measure, and determine whether buying coffee and buying tea are positively correlated, negatively correlated, or not correlated.
(2). Perform the chi-square test with 5% significance level and determine whether they are correlated or not.

Problem 4 (20 points). You will perform association analysis using JMP Pro. There is a section in *Predictive and Specialized Modeling.pdf* documentation that shows how to perform association analysis. You may want to read this section before starting the assignment. Follow the instructions in *JMP-association-analysis-assignment.pdf* file.

Submission

Include all answers in a single file and name it *lastName_firstName*_HW8.EXT. Here, "EXT" is an appropriate file extension (e.g., docx or pdf). If you have multiple files, then combine all files into a single archive file. Name the archive file as *lastName_firstName*_HW8.EXT. Here, "EXT" is an appropriate archive file extension (e.g., zip or rar). Upload the file to Blackboard.