BUSINESS DATA MINING (IDS 572)

HOMEWORK 10
DUE DATE: WEDNESDAY, NOVEMBER 30 AT 3:00 PM

- Please provide succinct answers to the questions below.
- You should submit an electronic pdf or word file in blackboard.
- Please include the names of all team-members in your write up and in the name of the file.

Problem 1. Subota Automobiles is trying to determine how to combine the options on their cars. For one particular model, the only options are (M) a manual transmission, (S) a sunroof, and (A) an improved audio system. Over the course of a month, the following sales (in thousands) occurred.

None	22
sunroof	0
manual	5
audio	3
S & M	8
S & A	4
M & A	14
S & M & A	18

Each line above describes all options included. For example, there were 0 units sold with sunroof only but there were 4 thousand units sold with both sunroof and improved audio. For each of the following, give the required value, or state that there is not enough information given.

(a) Fraction of all customers who have a sunroof (perhaps with other options).

(b) Confidence, Lift and Support of the rule sunroof $\Rightarrow$ improved audio

(c) Confidence, Lift and Support of the rule (M and A) $\Rightarrow$ S

(d) Confidence, Lift and Support of the rule M $\Rightarrow$ S

(e) Confidence, Lift and Support of the rule (not M) $\Rightarrow$ S

Problem 2. Consider the following lines from a Market Basket analysis:

Rule	Confidence	Lift	Support
Apples $\Rightarrow$ Pears	0.5	1.1111	0.3
Apples $\Rightarrow$ Benanas	0.6666	1.3333	0.4

For each of the following, find the required fraction, or state that there is not enough information given.

(a) Fraction of all your customers who bought both Apples and Bananas

(b) Of those customers who bought Apples, the fraction who did not buy Pears.

(c) Fraction of your customers who did not buy Bananas.

(d) Fraction of your customers who bought Apples.

Problem 3. This is a computational exercise on Association Rules mining. Do this by hand. The following data refers to 800 students at a university:

MAJOR	STATUS	AGE	NATIONALITY	NO.
Arts	Graduate	Old	U.S.	38
Arts	Graduate	Old	International	18
Arts	Undergraduate	Old	U.S.	29
Arts	Undergraduate	Young	U.S.	108
Engineering	Graduate	Old	International	78
Engineering	Undergraduate	Young	International	100
Engineering	Undergraduate	Young	U.S.	242
Science	Graduate	Old	International	25
Science	Undergraduate	Young	U.S.	162

Suppose we wish to mine this database for all rules with a support (coverage) of at least 20% and confidence (accuracy) of at least 90%.

(a) Use the A Priori algorithm to come up with all frequent 1, 2, 3 and 4 itemsets (L_1, L_2, L_3, L_4).

(b) Pick the 4-itemset and the 3-itemset with the maximum support from L_4 and L_3 respectively, and generate all rules satisfying the specified confidence level from each of these frequent itemsets.