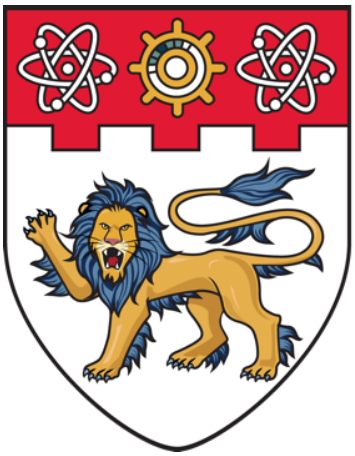

NANYANG TECHNOLOGICAL UNIVERSITY

SINGAPORE

EE4483
DATA-MINING ASSIGNMENT REPORT

Hu Xiaoxiang
U1521319A
EEE

13 Nov, 2017

Contents

1 Question 1: **1**

2 Question 2: **1**

3 Question 3: **1**

4 Question 4: **1**

5 Question 5: **1**

1 Question 1:

Minimum Support (%)	No. of Frequent Itemsets
20	20
10	68
5	268
3	659

2 Question 2:

Minimum Support (%)	No. of Frequent Itemsets	No. of Frequent 3-Itemsets	No. of Frequent 2-Itemsets
3	659	424	190

Percentage of frequent 3-itemset = 424 / 659 = 64.34%
Percentage of frequent 2-itemset = 190 / 659 = 28.83%

3 Question 3:

Minimum Support (%)	Minimum Confidence (%)	No. of Association Rules
5	50	117
10	50	0

Explanation: The smaller the minimum support is, the more the number of strong rules generate.

4 Question 4:

Rule1: Ice Cream, Olive, Tea –> Banana (minsup=3.3%)
Rule2: Banana, Ham, Salad –> Apple (minsup=3.3%)
Rule3: Ham, Diaper, Coffee –> Ice Cream (minsup=2.8%)

5 Question 5:

Interesting Rules:
Coffee, Salad, Lemon –> Egg, Apple (Minsup=2%, Minconf=100%, Lift = 10.588)
Nuts, Coffee, Salad –> Apple, Ketchup (Minsup=2%, Minconf=100%, Lift = 9)

These are the two most interesting rules I found when minimum support is 2% and minimum confidence is 100%. The measure used to identify the interestingness of the rule is lift. Lift shows the correlaton between the two itemsets. If the lift of the rules is high, it means the probability of occurrence of the antecedent is low. The higher the value of lift is, the more positively correlated these two datasets are.